BRIDGEROUTER: GAIN-AWARE INSTANCE ROUTING OVER CROSS-MODAL BRIDGE PORTFOLIOS

Anonymous authors
Paper under double-blind review

ABSTRACT

Lightweight representation bridges make pretrained, task-adapted EEG foundation models executable on other biosignal modalities without retraining their backbones. However, selecting a single bridge pathway globally assigns every instance to the same transfer route and ignores complementary errors across routes. We therefore formulate multi-bridge transfer as an instance-level algorithm-selection problem and introduce BRIDGEROUTER, a gain-aware router over a portfolio of frozen cross-modal pathways. A validation-selected Single Best Solver (SBS) serves as a protected default route. During training, task labels are used to compute the reduction in negative log-likelihood provided by each alternative route relative to the SBS, and a lightweight MLP predicts these route-specific gains from a target-modality representation extracted by a frozen encoder. At inference, BRIDGEROUTER observes only the target modality, requires neither EEG, task labels, nor predictions from all candidate routes, and executes exactly one bridge pathway: the route with the largest predicted gain when that gain exceeds a validation-calibrated threshold, and the SBS otherwise. We evaluate seven EEG foundation-model bridge pathways for EMG-based freezing-of-gait detection on FOG and ECG-based sleep classification on ISRUC. Across five seeds, BRIDGEROUTER achieves $81.38 \pm 1.11\%$ balanced accuracy on FOG, improving over the BIOT SBS by 11.49 percentage points, and $62.92 \pm 2.17\%$ on ISRUC, improving over the CBraMod SBS by 3.83 points. Per-instance bridge oracles reach 99.96% and 78.61%, respectively, revealing substantial remaining portfolio headroom. These results show that anchor-relative gain prediction can convert complementary cross-modal bridge behavior into deployable instance-level routing improvements while retaining single-route inference.

1 INTRODUCTION

What if the best model for transfer is not the best model for every instance?

Cross-modal transfer provides a natural way to reuse knowledge from pretrained biosignal foundation models when the source modality is unavailable at deployment. For example, a pretrained EEG model can be connected to EMG or ECG through a lightweight representation bridge learned from paired recordings (Li et al., 2026). When multiple pretrained EEG models are available, a natural strategy is to build a bridge for each model and simply choose the one with the best average performance.

However, average performance can hide an important property of a transfer portfolio: *different pathways can fail on different instances*. A pathway that is weaker overall may still correctly predict examples that a stronger pathway misses. In such a case, its value does not come from being the best standalone model, but from the errors it can correct. This suggests a different question:

> *Given the prediction we already have, can another transfer pathway do better on this instance?*

Answering this question is challenging because the information needed to determine the correct pathway is unavailable at inference. During training, paired EEG–target recordings and task labels reveal which pathways succeed or fail on each instance. At deployment, however, only the 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.

modality is observed. A useful router must therefore learn pathway complementarity from training-time transfer outcomes while making its decisions entirely from target-modality information.

We introduce **BridgeRouter**, a gain-aware framework for instance-level selection among complementary cross-modal pathways. We construct a portfolio of target-side executable pathways, each combining a frozen pretrained EEG foundation model with a lightweight bridge. One pathway is selected on the validation set as the Single Best Solver and serves as the default route. During training, the router learns the task-label-derived reduction in negative log-likelihood provided by each alternative pathway relative to this anchor. At inference, the router predicts these anchor-relative gains using only the target-modality representation. It selects the most promising alternative when its predicted gain exceeds a validation-calibrated protection threshold and selects the anchor otherwise. The routing decision is made before pathway execution, so exactly one selected bridge pathway is executed for each instance. BridgeRouter therefore treats pathway selection as a conditional decision rather than a global model-selection problem.

We evaluate BridgeRouter on two cross-modal biosignal transfer tasks: adapting pretrained EEG foundation models to EMG inputs for freezing-of-gait detection and to ECG inputs for sleep classification. We examine whether different bridge pathways exhibit complementary errors, whether their anchor-relative gains can be predicted without access to EEG at inference, and whether adaptive routing improves over the validation-selected SBS. Fixed-route performance, per-instance bridge oracles, and rescue-harm diagnostics are used to characterize the available portfolio headroom and the behavior of the learned routing policy.

Our contributions are:

- We formulate cross-modal transfer with multiple pretrained source models as an instance-level pathway selection problem, emphasizing the complementary errors that distinguish portfolio value from standalone performance.
- We propose **BridgeRouter**, which learns task-label-derived, anchor-relative negative-log-likelihood gains from target-modality representations and performs target-modality-only, single-route selection at inference.
- We evaluate gain-aware routing across EEG-to-EMG and EEG-to-ECG transfer tasks and quantify bridge complementarity through fixed-route performance, per-instance oracle headroom, and rescue-harm analyses.

2 Related Work

2.1 EEG Foundation Models and Cross-Modal Representation Transfer

Recent EEG foundation models (FMs) leverage self-supervised pretraining on large-scale unlabeled EEG recordings to learn transferable representations for downstream neural decoding tasks. Representative models, including BENDR Kostas et al. (2021), BIOT Yang et al. (2023), EEGPT Wang et al. (2024), LaBraM Jiang et al. (2024), and CBraMod Wang et al. (2025), adopt different architectures, pretraining objectives, and data sources. Recent benchmark studies have further shown substantial variation in their downstream performance across datasets and tasks, suggesting that the choice of a pretrained EEG representation can depend strongly on the target setting Xiong et al. (2026).

Beyond within-modality adaptation, recent work has explored transferring pretrained biosignal representations across modalities. Lightweight representation bridges provide a practical way to connect a frozen source foundation model to a target biosignal modality without retraining the source model. Such approaches make pretrained representations reusable when the target modality has limited labeled data or when a model pretrained on another modality provides useful prior knowledge. However, when multiple source foundation models are available, each can induce a different cross-modal transfer pathway through its corresponding bridge. Existing cross-modal transfer methods typically evaluate these pathways independently and select a fixed model or bridge based on aggregate performance. They therefore do not explicitly consider whether different pathways make complementary predictions on individual target instances.

2.2 Model Routing and Per-Instance Model Selection

A complementary line of work studies conditional model selection, where different inputs can be processed by different models or computational pathways. Mixture-of-experts methods dynamically route inputs to specialized experts, while model cascades and conditional computation adapt the amount or type of computation according to the input Shazeer et al. (2017); Wang et al. (2018). More generally, per-instance model selection aims to exploit heterogeneity in model behavior rather than relying on a single model for all inputs.

These approaches motivate adaptive selection in a transfer portfolio, but cross-modal transfer introduces an additional constraint. At training time, paired source–target recordings and task labels can reveal how different pathways behave on the same instance. At deployment, however, the source modality may be unavailable, and executing every source pathway may defeat the purpose of lightweight transfer. Thus, a routing policy must infer the usefulness of alternative pathways from information available on the target side alone. BridgeRouter addresses this setting by predicting anchor-relative pathway gains solely from the target-side representation and selecting a single route before bridge execution, without requiring access to the source modality or executing the entire pathway portfolio at inference.

2.3 Algorithm Selection and Complementary Model Portfolios

Automated Algorithm Selection (AS) studies how to select an algorithm from a portfolio according to the characteristics of an individual problem instance Rice (1976); Smith-Miles (2009). AS has been applied to domains including Boolean satisfiability, combinatorial optimization, and machine learning. A central motivation is that algorithms with different strengths can perform differently across instances, making instance-specific selection potentially more effective than committing to a single globally strongest algorithm. Traditional AS methods often learn performance predictors from handcrafted instance features, while more recent approaches employ learned representations, including convolutional and graph-based encoders, to automatically characterize instances Loreggia et al. (2016); Cao et al. (2025).

Recent work has extended this perspective to foundation-model selection. EEG-AS Zhang et al. (2026) formulates EEG foundation model selection as an instance-level AS problem and learns to estimate foundation-model behavior from inference-available EEG representations and an anchor model. Its formulation demonstrates that foundation-model portfolios can benefit from instance-dependent selection rather than relying exclusively on a Single Best Solver. However, EEG-AS considers selection among pretrained EEG models within the same modality. In contrast, BridgeRouter considers a portfolio of *cross-modal transfer pathways*, where each candidate pathway combines a pretrained source model with a target-side bridge.

This distinction also changes the selection objective. Rather than predicting which pathway is globally best or reconstructing the absolute performance of every candidate, BridgeRouter focuses on the *relative benefit of switching* from an anchor pathway to an alternative pathway. This formulation explicitly exploits complementary errors: an alternative pathway is valuable when it succeeds on instances where the anchor fails, even if its overall performance is lower. BridgeRouter therefore connects automated algorithm selection with cross-modal transfer by treating pathway complementarity as a source of instance-level adaptive performance.

3 Methods

3.1 Problem Setup and Notation

We consider a target biosignal task on modality m, such as EMG or ECG, where knowledge from pretrained EEG foundation models is available. During training, paired samples (x^e, x^m, y) are given, with x^e and x^m denoting synchronized EEG and target-modality observations and y the task label. This paired setting provides the information needed to establish cross-modal transfer. At inference, however, EEG is unavailable, so the final prediction must be obtained from x^m alone.

Our goal is not to transfer one EEG model into the target modality and use it as a fixed predictor. Instead, we construct multiple transfer pathways, each based on a different EEG foundation model

and its corresponding cross-modal bridge. Let F_k^e denote the k-th frozen EEG foundation model and let E^m denote the frozen pretrained target-modality encoder. The target-modality representation available to the router is denoted by

$$\phi(x^m) = E^m(x^m).$$

A lightweight bridge B_k connects a representation produced by the target-modality encoder to the representation space of F_k^e, defining a pathway π_k. The resulting pathways form a portfolio $\{\pi_k\}_{k=1}^K$, with pathway output $\hat{y}_{ik}$ and class probabilities p_{ik} for instance i.

We designate one member of the portfolio as the anchor pathway π_a. The anchor is the Single Best Solver (SBS), selected exclusively according to validation balanced accuracy:

$$a = \arg \max_{k \in \{1,\dots,K\}} \mathrm{BAcc}_{D_{\mathrm{val}}}(\pi_k).$$

It serves as the default route in the routing decision and is frozen before test evaluation.

This portfolio introduces a different question from conventional cross-modal transfer. Rather than asking which EEG foundation model transfers best on average, we ask whether the best pathway can vary from one target-modality instance to another. This distinction is important because the router must make its decision without access to the privileged EEG observation. Given only the target representation $\phi(x^m)$, the router G_θ estimates whether switching from the anchor to another pathway is beneficial. Thus, the routing decision is learned from training-time transfer outcomes but, at inference, depends only on information available from x^m.

3.2 Cross-modal Pathways via Representation Bridging

The key to constructing such a portfolio is to make each EEG foundation model executable from the target modality alone. We therefore learn a lightweight bridge that maps target-modality representations into the representation space of each frozen EEG foundation model. Following the representation-bridging formulation of BioX-Bridge (Li et al., 2026), the bridge is learned from paired EEG-target observations, while the foundation model itself remains frozen.

For the k-th pathway, the target-modality representation is mapped as

$$z_{ik}^e = B_k(\phi_k(x_i^m)), \tag{1}$$

where $\phi_k(x_i^m)$ denotes the representation extracted from the selected source layer of the frozen target-modality encoder and z_{ik}^e lies in the representation space of F_k^e. During training, the representation extracted from the corresponding EEG observation provides the reference for cross-modal alignment. This alignment stage uses the paired EEG observation only as privileged training information.

Once the bridge is learned, the pathway can be executed without observing EEG. The bridged representation is passed directly to the remaining layers of the frozen EEG foundation model:

$$p_{ik} = \mathrm{softmax}\left[\mathrm{Head}_k\left(F_{k,>d_k}^e\left(B_k(\phi_k(x_i^m))\right)\right)\right], \qquad \hat{y}_{ik} = \arg\max_c p_{ik,c}. \tag{2}$$

The resulting computation is therefore driven by x^m even though the frozen backbone originates from EEG. Different choices of F_k^e and B_k consequently give different target-side executable pathways, allowing the pretrained models to contribute complementary transfer behavior rather than being reduced to a single transferred predictor.

3.3 Portfolio View of Multi-Bridge Transfer

Because the pathways are built from different foundation models and bridges, they need not make the same prediction on a given target-modality sample. A pathway may capture a particular pattern well while another pathway fails on the same instance. Consequently, their average accuracies do not fully characterize the value of the portfolio: even a weaker pathway on average may provide the correct prediction on samples where the anchor fails.

We therefore evaluate transfer quality at the instance level. For a sample i, let

$$c_{ik} = \mathbf{1}[\hat{y}_{ik} = y_i]. \tag{3}$$

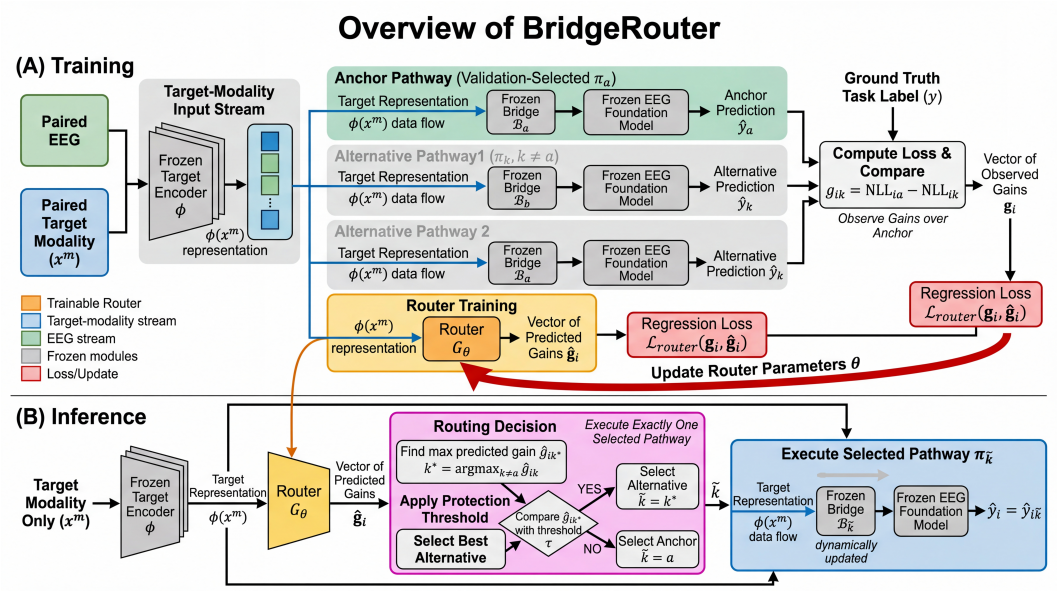

Figure 1: Overview of BridgeRouter. **(A) Training.** Paired EEG–target-modality data are used to train cross-modal bridges and the router, with the validation-selected pathway π_a serving as the anchor. The router is supervised by the anchor-relative gain $g_{ik} = \mathrm{NLL}_{ia} - \mathrm{NLL}_{ik}$. **(B) Inference.** Given only the target modality, the router predicts the gains of alternative pathways and selects the best alternative when its predicted gain exceeds the protection threshold τ; otherwise, it retains the anchor.

This variable indicates whether pathway k predicts the label correctly. The corresponding portfolio oracle is

$$c_i^* = \max_k c_{ik}, \tag{4}$$

which indicates whether at least one available pathway can solve the instance correctly.

The gap between a fixed pathway and this oracle exposes the opportunity for routing: the portfolio contains useful predictions that a single globally selected pathway cannot recover. The problem is therefore not to determine one universally preferred bridge, but to identify, for each target-modality instance, whether another pathway is more suitable than the anchor. BridgeRouter addresses this problem by learning this instance-level pathway preference from training-time transfer outcomes, while keeping the routing decision entirely target-modality based at inference.

3.4 BridgeRouter: Training and Inference

Figure 1 summarizes the proposed BridgeRouter framework. As discussed above, different pathways can be preferable for different target-modality instances, making a fixed choice of pathway insufficient. At inference time, however, the ground-truth label is unavailable, so we cannot directly determine which pathway will provide the largest improvement. This limitation raises a practical question: given a particular target-modality instance, which pathway should we use? BridgeRouter addresses this question by predicting the benefit of each alternative route from the target-modality representation and selecting either the anchor or one alternative before executing a bridge pathway.

Training. The key question is not whether an alternative pathway is accurate in isolation, but whether it can improve upon the pathway we would use by default. We therefore train the router to predict each alternative pathway's loss reduction relative to the anchor. During router training, all pathways can be evaluated on the training data, allowing us to measure how much each alternative improves upon the anchor for each instance.

For sample i, let

$$\mathrm{NLL}_{ik} = -\log p_{ik,y_i} \tag{5}$$

denote the negative log-likelihood of the prediction produced by pathway k. Using the anchor as the reference, we define the observed benefit of switching to an alternative pathway $k \neq a$ as

$$g_{ik} = \mathrm{NLL}_{ia} - \mathrm{NLL}_{ik}\,. \tag{6}$$

A positive g_{ik} indicates that pathway k improves upon the anchor prediction for that instance, while a negative value indicates that switching to pathway k increases the prediction loss.

The router learns to predict the switching benefits from information available at deployment,

$$\hat{\mathbf{g}}_i = G_\theta(\phi(x_i^m))\,, \tag{7}$$

where $\hat{\mathbf{g}}_i$ contains the predicted gains for all alternative pathways $k \neq a$. We train the router using a regression loss between $\hat{\mathbf{g}}_i$ and the observed gains $\mathbf{g}_i$. The paired EEG observation is used to establish the cross-modal bridges, while the task label is used to compute the route-specific negative log-likelihood gains. Neither the EEG observation nor the task label is provided as an input to the router. At inference, the router requires only the target-modality representation.

Inference. At deployment, the router computes the target-modality representation $\phi(x_i^m)$ and predicts the relative gain of each alternative pathway:

$$\hat{\mathbf{g}}_i = G_\theta(\phi(x_i^m))\,. \tag{8}$$

The router then identifies the most promising alternative:

$$k_i^* = \arg\max_{k \neq a} \hat{g}_{ik}. \tag{9}$$

A large predicted gain alone, however, does not necessarily justify switching away from the validation-selected SBS. We therefore use a protection threshold τ, selected exclusively on the validation split, to distinguish a potentially useful correction from a marginal predicted improvement. The selected route is

$$\tilde{k}_i = \begin{cases} k_i^*, & \hat{g}_{ik_i^*} > \tau, \\ a, & \hat{g}_{ik_i^*} \leq \tau. \end{cases} \tag{10}$$

After this routing decision, exactly one pathway, $\pi_{\tilde{k}_i}$, is executed using the target-modality input. The final prediction is

$$\hat{y}_i = \hat{y}_{i\tilde{k}_i}. \tag{11}$$

Thus, the anchor serves as the default route: BridgeRouter switches to an alternative only when its predicted gain exceeds the validation-selected threshold. The anchor does not need to be executed before routing because the router conditions only on the target-modality representation. BridgeRouter therefore retains SBS protection while requiring exactly one bridge-pathway execution per instance.

4 Experiments and Discussion

4.1 Experimental Setup

Datasets. We evaluate on two multimodal physiological-signal datasets: the freezing-of-gait (**FOG**) dataset Zhang et al. (2022) and the **ISRUC** sleep dataset Khalighi et al. (2016). FOG contains synchronized EEG and EMG recordings from 12 participants. We use three-second windows and obtain 37,200 paired EEG–EMG windows after preprocessing for binary FOG classification. For ISRUC, we use Subgroup I with synchronized EEG and ECG recordings. Following BioX-Bridge Li et al. (2026), the original five sleep stages are grouped into Wake versus Sleep, with each 30-second epoch treated as one sample. We use six EEG derivations (F3–A2, C3–A2, O1–A2, F4–A1, C4–A1, and O2–A1). Thus, EEG is the old modality, while EMG and ECG are the new modalities for FOG and ISRUC, respectively. Further preprocessing details are given in Appendix A.1.

Data Partitioning. We follow the partitioning protocol of BioX-Bridge Li et al. (2026). Each dataset is divided into D_{old}, D_{pair}, D_{val}, and D_{new}, with approximate proportions of 33%, 33%, 11%, and 22%. D_{old} is used to train task-specific heads for the frozen EEG models, D_{pair} is used for bridge and routing training, D_{val} is used for model selection, and D_{new} is used only for final evaluation. ISRUC is split by subject, so no subject appears in more than one subset. FOG follows the sample-wise BioX protocol and therefore does not evaluate unseen-subject generalization. The detailed FOG split is reported in Appendix A.1.

Foundation Models and Bridge Routes. We evaluate seven pretrained EEG foundation models: LaBraM Jiang et al. (2024), EEGPT Wang et al. (2024), REVE El Ouahidi et al. (2026), BIOT Yang et al. (2023), CBraMod Wang et al. (2025), CSBrain Zhou et al. (2026), and BENDR Kostas et al. (2021). We freeze all EEG backbones and train a linear task head on D_{old}. For the new modalities, we use pretrained NormWear for EMG Luo et al. (2025) and HuBERT-ECG for ECG Coppola et al. (2024), which are also frozen. Each EEG model defines one cross-modal bridge, giving seven candidate routes per dataset. Bridge layer positions are selected following BioX-Bridge.

Baselines and Comparators. We first evaluate each bridge pathway independently. The single best solver (SBS) is selected exclusively by validation balanced accuracy and frozen before test evaluation. This procedure selects BIOT on FOG and CBraMod on ISRUC. The SBS is never reselected according to test performance.

Our primary method is BRIDGEROUTER, a gain-aware instance-level routing framework over cross-modal bridge pathways. It predicts the instance-specific reduction in negative log-likelihood of each candidate route relative to the validation-selected SBS. The router retains the SBS unless the largest predicted gain exceeds a threshold selected using validation data. It uses task-label-derived gain targets during training but requires neither labels nor bridge predictions at inference. It observes only the target-modality representation and executes one selected bridge pathway.

We additionally report fixed-route performance and per-instance bridge oracles. The fixed routes establish the performance of each available pathway without routing. The oracles quantify portfolio complementarity by selecting a correct pathway using the unknown ground-truth label. They are non-deployable and are not used for training, model selection, or calibration.

Implementation Details. BRIDGEROUTER uses the pooled NormWear block-6 representation on FOG and the 512-dimensional HuBERT-ECG block-4 representation on ISRUC. Its MLP contains hidden dimensions 256 and 128, GELU activations, and dropout of 0.2. The model predicts the route-specific negative log-likelihood gain relative to the validation-selected SBS and is trained using class-balanced supervision.

For each instance, the route with the largest predicted gain is identified. The SBS is retained unless this gain exceeds a protection threshold selected on the validation split. At inference, the router observes only the target-modality representation and executes one bridge pathway. It does not require the test label or the predictions of all candidate bridges. We evaluate seeds 42-46 while keeping all bridge checkpoints and data partitions fixed.

Evaluation Metrics. Balanced accuracy is the primary metric because both tasks are class-imbalanced. We additionally report accuracy and macro F1. For routing diagnostics, a rescue occurs when the SBS is incorrect and the selected alternative is correct; a harm occurs when the SBS is correct and the selected alternative is incorrect. All stochastic routing results report mean and sample standard deviation over five seeds.

4.2 FROZEN BRIDGE PERFORMANCE

Table 1 reports the performance of every frozen cross-modal pathway on the two final test sets. Bridge quality varies substantially across receivers and does not directly follow native EEG performance.

On FOG, BIOT is both the validation-selected SBS and the strongest test pathway, reaching 69.89% balanced accuracy. BENDR and CSBrain follow at 65.96% and 64.05%, whereas LaBraM, EEGPT, and REVE exhibit near-chance and strongly class-biased bridge predictions.

Table 1: Frozen cross-modal bridge performance on the final test sets. † denotes the validation-selected SBS. All values are percentages.

(a) FOG

Receiver	Accuracy	BAcc	Macro-F1
LaBraM	57.63	49.98	36.59
EEGPT	58.39	51.09	39.83
REVE	42.91	50.41	31.14
BIOT†	**69.07**	**69.89**	**68.98**
CBraMod	57.30	58.02	57.23
CSBrain	65.07	64.05	64.10
BENDR	65.72	65.96	65.47

(b) ISRUC

Receiver	Accuracy	BAcc	Macro-F1
LaBraM	78.087	60.486	61.772
CBraMod†	78.087	59.090	60.229
REVE	79.693	61.732	63.445
EEGPT	75.920	**65.171**	65.000
CSBrain	79.096	62.905	64.438
BENDR	**81.506**	64.445	**66.755**
BIOT	80.518	64.051	66.005

Table 2: Bridge-portfolio headroom on the final test sets. Oracles are non-deployable.

Dataset	SBS	SBS BAcc	K7 Oracle BAcc	Headroom
FOG	BIOT	69.89	99.96	+30.07
ISRUC	CBraMod	59.09	78.61	+19.52

On ISRUC, CBraMod is selected as the SBS because it obtains the highest validation balanced accuracy. Its test balanced accuracy, however, drops to 59.09%. EEGPT and BENDR reach 65.17% and 64.45% on test, respectively. These results do not change the SBS retrospectively: selecting EEGPT after observing test labels would constitute test-set model selection. Instead, the ranking reversal demonstrates that a globally selected pathway may fail to generalize across unseen subjects.

The pathway errors remain complementary on both datasets. The seven-route oracle reaches 99.96% balanced accuracy on FOG and 78.61% on ISRUC, leaving 30.07 and 19.52 percentage points of headroom over the respective validation-selected SBSs. The FOG oracle is partly inflated by complementary class-biased routes, but the validation-filtered compact-four oracle still reaches 93.69%. Therefore, substantial complementarity remains even after removing collapsed pathways.

4.3 Gain-Aware Bridge Routing

Table 3 presents the primary cross-dataset result. On FOG, the BridgeRouter reaches $81.38 \pm 1.11\%$ balanced accuracy, improving over the BIOT SBS by 11.49 percentage points. On ISRUC, it reaches $62.92 \pm 2.17\%$, improving over the validation-selected CBraMod SBS by 3.83 points. Every ISRUC seed improves over CBraMod in balanced accuracy.

The ISRUC improvement is smaller and more variable than the FOG improvement. This is consistent with its subject-disjoint protocol and the pronounced validation-to-test pathway-ranking shift. The result shows that relative-gain supervision transfers across the two tasks, but it does not imply dominance over every fixed test pathway. In particular, EEGPT and BENDR obtain higher ISRUC test balanced accuracy than the five-seed router mean, although neither was selected as the SBS using validation balanced accuracy.

Relative to the corresponding SBS-to-oracle gaps, the router recovers approximately 38.2% of the available FOG headroom and 19.6% of the ISRUC headroom. At inference, the router observes only the target-modality representation, requires no test label, and executes one bridge pathway. Thus, its computational cost is substantially lower than evaluating all seven routes.

Table 3: Primary BRIDGEROUTER results on the final test sets. Router values are mean ± sample standard deviation over five seeds. Δ is measured relative to the validation-selected SBS. Oracles are non-deployable.

Dataset	Method	Accuracy	BAcc	Macro-F1	ΔBAcc
FOG	BIOT SBS	69.07	69.89	68.98	–
FOG	BRIDGEROUTER	$\mathbf{81.49 \pm 1.20}$	$\mathbf{81.38 \pm 1.11}$	$\mathbf{81.17 \pm 1.18}$	$\mathbf{+11.49}$
FOG	K7 Oracle	99.96	99.96	–	+30.07
ISRUC	CBraMod SBS	78.087	59.090	60.229	–
ISRUC	BRIDGEROUTER	$\mathbf{78.658 \pm 0.747}$	$\mathbf{62.924 \pm 2.172}$	$\mathbf{64.252 \pm 2.252}$	$\mathbf{+3.834}$
ISRUC	K7 Oracle	90.127	78.612	82.976	+19.522

The ISRUC result must nevertheless be interpreted conservatively. Its five-seed switch rate is highly variable, largely because one seed switches substantially more often than the remaining four. Moreover, the router mean remains below the strongest fixed test pathways. The consistent improvement over the validation-selected SBS demonstrates useful instance-level gain prediction, while the remaining oracle gap shows that reliable cross-subject route selection remains unresolved.

4.4 CROSS-DATASET FINDINGS

The two datasets support three common conclusions.

First, frozen bridge pathways exhibit substantial but heterogeneous complementarity. Neither native EEG performance nor a single global bridge ranking fully describes instance-level pathway behavior. The seven-route oracle reaches 99.96% balanced accuracy on FOG and 78.61% on ISRUC, leaving substantial headroom over the validation-selected SBS on both datasets.

Second, predicting relative task loss provides useful routing supervision across both clinical tasks. The BRIDGEROUTER improves over the validation-selected SBS by 11.49 balanced-accuracy points on FOG and by 3.83 points on ISRUC. Because only one selected bridge is executed, this improvement does not require evaluating the complete portfolio at inference.

Third, protocol differences materially affect routing difficulty. FOG uses a sample-wise split containing the same participants across subsets, whereas ISRUC evaluates unseen subjects. ISRUC also exhibits a substantial validation-to-test pathway-ranking reversal: CBraMod is the validation-selected SBS but is not the strongest individual test pathway. Consequently, improvement over the correctly selected SBS can coexist with lower performance than pathways identified only after observing test results.

These findings establish BRIDGEROUTER as the primary cross-dataset method while showing that robust instance-level route selection under cross-subject distribution shift remains an open problem.

5 CONCLUSION

We presented BRIDGEROUTER, an instance-level routing framework for cross-modal transfer from pretrained EEG foundation models. Rather than relying on a single globally selected bridge, our approach learns to predict the relative task-loss gain of candidate bridges and selects an alternative route only when the predicted gain is sufficiently large. This enables instance-adaptive bridge selection while executing only one bridge pathway at inference time.

Across the FOG and ISRUC datasets, we found substantial complementarity among frozen EEG-to-target-modality bridges. The proposed router improves over the validation-selected single best solver by 11.49 and 3.83 percentage points in balanced accuracy on FOG and ISRUC, respectively. At the same time, the remaining gap to the oracle and the smaller and more variable routing gains under the subject-disjoint ISRUC setting highlight the difficulty of reliable instance-level routing under distribution shift.

These results suggest that cross-modal transfer can benefit from treating pretrained foundation models as a portfolio of complementary bridge routes rather than committing to a single global choice. Future work will investigate more robust routing under cross-subject and cross-domain shifts, as well as richer signals for estimating when a bridge is likely to provide a meaningful transfer gain.

AI USE STATEMENT

Generative AI tools were used to assist with language editing and polishing of the manuscript, including improving clarity, grammar, and organization. They were not used to generate research hypotheses, determine the experimental methodology, conduct experiments, or make scientific conclusions. All AI-assisted text was reviewed and revised by the authors, who take full responsibility for the content of the manuscript.

ETHICS STATEMENT

This work uses existing physiological datasets for methodological research on EEG foundation models. We did not collect new data or interact directly with human participants. All experiments were conducted using the datasets under their respective access and usage conditions. The study focuses on methodological evaluation and does not make clinical decisions or provide medical recommendations. We report both positive and negative experimental findings and describe the evaluation protocol to reduce the risk of misleading conclusions arising from model selection or test-set reuse.

REPRODUCIBILITY STATEMENT

We describe the main experimental settings, model configurations, and evaluation procedures used in this study in the main paper and appendix. Where applicable, we report the foundation-model checkpoints, bridge configurations, and random seeds used in our experiments. We also provide implementation details necessary to reproduce the proposed method and the reported evaluation protocol.

REFERENCES

Siyuan Cao, Hongxuan Wu, Jiabao Brad Wang, Yiliang Yuan, and Mustafa Misir. MC-GNNAS-Dock: Multi-criteria gnn-based algorithm selection for molecular docking. In *PRICAI 2025: Trends in Artificial Intelligence*, pp. 686–696, 2025. doi: 10.1007/978-981-95-7075-1_46.

Edoardo Coppola, Mattia Savardi, Mauro Massussi, Marianna Adamo, Marco Metra, and Alberto Signoroni. HuBERT-ECG: A self-supervised foundation model for broad and scalable cardiac applications. *medRxiv*, 2024.

Yassine El Ouahidi, Jonathan Lys, Philipp Thölke, Nicolas Farrugia, Bastien Pasdeloup, Vincent Gripon, Karim Jerbi, and Giulia Lioi. Reve: A foundation model for eeg-adapting to any setup with large-scale pretraining on 25,000 subjects. *Advances in Neural Information Processing Systems (NeurIPS)*, 38:22541–22577, 2026.

Wei-Bang Jiang, Li-Ming Zhao, and Bao-Liang Lu. Large brain model for learning generic representations with tremendous eeg data in bci. In *International Conference on Learning Representations (ICLR)*, 2024.

S. Khalighi, T. Sousa, J. M. Santos, and U. Nunes. ISRUC-Sleep: A comprehensive public dataset for sleep researchers. *Computer Methods and Programs in Biomedicine*, 124:180–192, 2016. doi: 10.1016/j.cmpb.2015.10.013.

Demetres Kostas, Stephane Aroca-Ouellette, and Frank Rudzicz. Bendr: Using transformers and large scale self-supervised learning for generalizable neural decoding. *Frontiers in Human Neuroscience*, 15:653659, 2021.

Chenqi Li, Yu Liu, Timothy Denison, and Tingting Zhu. BioX-Bridge: Model bridging for unsupervised cross-modal knowledge transfer across biosignals. In *International Conference on Learning Representations*, pp. 111389–111415, 2026. URL https://proceedings.iclr.cc/paper_files/paper/2026/file/b4ec8a0ae0373d694dc8529ca523e9a9-Paper-Conference.pdf.

Andrea Loreggia, Yuri Malitsky, Horst Samulowitz, and Vijay Saraswat. Deep learning for algorithm portfolios. In *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence*, volume 30, 2016. doi: 10.1609/aaai.v30i1.10170.

Yunfei Luo, Yuliang Chen, Asif Salekin, and Tauhidur Rahman. Toward foundation model for multivariate wearable sensing of physiological signals, 2025. URL https://arxiv.org/abs/2412.09758.

John R. Rice. The algorithm selection problem. *Advances in Computers*, 15:65–118, 1976. doi: 10.1016/S0065-2458(08)60520-3.

Noam Shazeer, Azalia Mirhoseini, Krzysztof Maziarz, Andy Davis, Quoc V. Le, Geoffrey Hinton, and Jeff Dean. Outrageously large neural networks: The sparsely-gated mixture-of-experts layer. In *International Conference on Learning Representations*, 2017.

Kate Smith-Miles. Cross-disciplinary perspectives on meta-learning for algorithm selection. *ACM Computing Surveys*, 41(1), 2009. doi: 10.1145/1456650.1456656.

Guangyu Wang, Wenchao Liu, Yuhong He, Cong Xu, Lin Ma, and Haifeng Li. Eegpt: Pretrained transformer for universal and reliable representation of eeg signals. *Advances in Neural Information Processing Systems*, 37:39249–39280, 2024.

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*, volume 2025, pp. 75310–75346, 2025.

Xin Wang, Fisher Yu, Zi-Yi Dou, Trevor Darrell, and Joseph E. Gonzalez. SkipNet: Learning dynamic routing in convolutional networks. In *European Conference on Computer Vision*, pp. 409–424, 2018.

Wei Xiong, Jiangtong Li, Jie Li, Kun Zhu, and Changjun Jiang. EEG-FM-Bench: A comprehensive benchmark for the systematic evaluation and diagnostic analyses of EEG foundation models. In *Forty-third International Conference on Machine Learning*, 2026. URL https://openreview.net/forum?id=vGeNaFHdET.

Chaoqi Yang, M Westover, and Jimeng Sun. Biot: Biosignal transformer for cross-data learning in the wild. *Advances in Neural Information Processing Systems (NeurIPS)*, 36:78240–78260, 2023.

W. Zhang, Z. Yang, H. Li, D. Huang, L. Wang, Y. Wei, L. Zhang, L. Ma, H. Feng, J. Pan, Y. Guo, and P. Chan. Multimodal data for the detection of freezing of gait in parkinson's disease. *Scientific Data*, 9(1):606, 2022. doi: 10.1038/s41597-022-01713-8.

Yunzhen Zhang, Ruoxi Piao, Hasan Onur Keles, and Mustafa Misir. EEG-AS: Instance-level foundation model selection for EEG foundation models via behavior reconstruction, 2026. URL https://arxiv.org/abs/2609.00653.

Yuchen Zhou, Jiamin Wu, Zichen Ren, Zhouheng Yao, Weiheng Lu, Kunyu Peng, Qihao Zheng, Chunfeng Song, Wanli Ouyang, and Chao Gou. Csbrain: A cross-scale spatiotemporal brain foundation model for eeg decoding. *Advances in Neural Information Processing Systems (NeurIPS)*, 38:87150–87195, 2026.

A Appendix

A.1 Experimental Setup

Datasets. We evaluate our method on two multimodal physiological-signal benchmarks: the freezing-of-gait (**FOG**) dataset Zhang et al. (2022) and the **ISRUC** sleep dataset Khalighi et al. (2016). These datasets represent two distinct clinical prediction settings and provide synchronized recordings from multiple physiological modalities.

The FOG dataset contains synchronized EEG and EMG recordings from 12 participants performing gait-related tasks. We segment each recording into three-second windows. The EEG signals are resampled to 200 Hz, producing six-channel windows with 600 temporal samples. The synchronized EMG windows are processed using the preprocessing pipeline associated with the pretrained

Table 4: FOG12 data partitioning. All four subsets contain samples from the same 12 participants.

SPLIT	SAMPLES	CLASS 0	CLASS 1	POS.
D_{old}	12,400	7,154	5,246	42.31%
D_{pair}	12,400	7,025	5,375	43.35%
D_{val}	4,133	2,398	1,735	41.98%
D_{new}	8,267	4,766	3,501	42.35%

NormWear encoder. After preprocessing and quality filtering, the FOG dataset contains 37,200 paired EEG–EMG windows for binary FOG classification.

ISRUC is a polysomnographic sleep-staging dataset containing synchronized EEG and ECG recordings from multiple participants. We use Subgroup I of the ISRUC-Sleep dataset, which contains overnight polysomnographic recordings with synchronized EEG and ECG signals. Sleep stages are annotated at 30-second resolution. We use the six EEG derivations F3–A2, C3–A2, O1–A2, F4–A1, C4–A1, and O2–A1, together with the synchronized ECG signal. Each annotated 30-second epoch is treated as one sample for the downstream binary classification task. Following BioX-Bridge Li et al. (2026), we collapse the original five sleep stages into a binary Wake-versus-Sleep task, where Wake is retained as one class and all sleep stages are grouped into the other.

In both datasets, we consider EEG as the old modality, and EMG and ECG as the new modalities for FOG and ISRUC, respectively.

Data Partitioning Protocol. For direct comparison with BioX-Bridge Li et al. (2026), we follow its dataset-specific partitioning protocol. For each dataset, the available recordings are divided into four subsets: D_{old}, D_{pair}, D_{val}, and D_{new}, with approximate proportions of 33%, 33%, 11%, and 22%, respectively. Here, D_{old} provides labeled observations from the old modality for training the task-specific heads of the frozen foundation models; D_{pair} contains synchronized old- and new-modality observations for learning the cross-modal bridges and transfer signals; D_{val} is used for model selection and calibration; and D_{new} is reserved for final evaluation. For ISRUC, partitioning is performed at the subject level, such that all epochs from a given participant are assigned to a single subset and no participant is shared across D_{old}, D_{pair}, D_{val}, and D_{new}.

The resulting FOG split is summarized in Table 4. The FOG evaluation follows the BioX-style sample-wise protocol and is therefore not an unseen-subject evaluation.

EEG Foundation Models and Bridge Routes. We evaluate seven EEG foundation models: LaBraM Jiang et al. (2024), EEGPT Wang et al. (2024), REVE El Ouahidi et al. (2026), BIOT Yang et al. (2023), CBraMod Wang et al. (2025), CSBrain Zhou et al. (2026), and BENDR Kostas et al. (2021). Each model is initialized using its publicly released pretrained weights. For both datasets, we freeze the EEG backbone and train only a task-specific linear classification head on D_{old}. Both FOG and ISRUC are formulated as binary classification tasks.

For the new-modality encoders, we use pretrained NormWear Luo et al. (2025) to encode EMG in FOG and pretrained HuBERT-ECG Coppola et al. (2024) to encode ECG in ISRUC. Both encoders remain frozen throughout bridge training and evaluation. We construct one bridge for each EEG foundation model, yielding seven candidate routes per dataset. For every route, we follow the layer-position selection procedure of BioX-Bridge to determine the source layer in the new-modality encoder and the destination layer in the corresponding EEG foundation model.

A.2 Task-Label-Free Transfer-Aware Routing

We investigate whether bridge-pathway selection can be learned without task labels or label-derived correctness and gain targets. For each pathway, the task-label-free router predicts three signals from the target-modality representation: native-teacher competence, normalized teacher-to-bridge Jensen-Shannon divergence, and teacher-bridge prediction-flip probability.

The prespecified transfer-aware score is

$$\hat{s}_{ik} = \hat{c}_{ik}(1 - \hat{f}_{ik}) - \hat{d}_{ik},$$

Table 5: Task-label-free routing over five seeds. FOG is evaluated on TEST using the compact-four portfolio. ISRUC is evaluated on VAL using the canonical seven-route portfolio. Δ is measured relative to the dataset-specific validation-selected SBS.

Dataset	Split	Method	Accuracy	BAcc	ΔBAcc
FOG	TEST	BIOT SBS	69.07	69.89	–
FOG	TEST	Competence only	69.55 ± 0.13	70.22 ± 0.10	$+0.33$
FOG	TEST	Transfer only	71.60 ± 0.36	69.90 ± 0.48	$+0.01$
FOG	TEST	Transfer aware	$\mathbf{72.37 \pm 0.32}$	$\mathbf{71.91 \pm 0.43}$	$\mathbf{+2.02}$
ISRUC	VAL	CBraMod SBS	79.889	65.919	–
ISRUC	VAL	Competence only	78.602 ± 0.089	61.580 ± 0.053	-4.339
ISRUC	VAL	Transfer only	81.179 ± 0.159	64.544 ± 0.075	-1.375
ISRUC	VAL	Transfer aware	78.627 ± 0.204	61.314 ± 0.247	-4.605

where $\hat{c}_{ik}$ is predicted teacher competence, $\hat{f}_{ik}$ is predicted flip probability, and $\hat{d}_{ik}$ is predicted normalized Jensen-Shannon divergence. We additionally evaluate competence-only and transfer-only variants to isolate the contribution of the two signal families.

The FOG router uses the validation-filtered compact-four portfolio consisting of BIOT, CBraMod, CSBrain, and BENDR. The ISRUC router uses the complete seven-route portfolio. Router checkpoint selection is based on held-out transfer-signal prediction loss rather than downstream classification performance.

On FOG, transfer-aware routing improves over the BIOT SBS by 2.02 ± 0.43 balanced-accuracy points. Transfer-only routing rescues 1035.8 ± 38.9 BIOT errors but harms 826.6 ± 47.6 BIOT successes. Combining competence and transfer risk reduces the number of harmful switches to 307.4 ± 80.1 while retaining 580.4 ± 69.9 rescues, producing a net gain of 273.0 ± 26.5 correct predictions.

The result does not transfer to ISRUC. None of the three label-free variants exceeds the validation-selected CBraMod SBS. Transfer-only is the strongest label-free variant at $64.54 \pm 0.07\%$ balanced accuracy, but remains 1.38 points below CBraMod. The prespecified transfer-aware combination reaches only $61.31 \pm 0.25\%$.

These results suggest that transfer-derived signals can be informative within an individual pathway without being calibrated across heterogeneous foundation models. Native-teacher confidence, teacher-bridge divergence, and prediction disagreement may have different numerical meanings for different receivers. Consequently, the same fixed transfer-aware score does not provide a universally reliable estimate of downstream pathway utility.

Because the ISRUC label-free formulation did not satisfy its frozen validation criterion, it was not promoted to final test evaluation. The ISRUC values are therefore reported on VAL, whereas the previously frozen FOG experiment is reported on TEST. This experiment is included as an exploratory extension and negative cross-dataset result rather than as the primary method.

A.3 FOG Direct-EEG Performance

Table 6 compares each frozen EEG model evaluated on the paired EEG signal with the corresponding NormWear-to-EEG bridge evaluated on EMG. Direct-EEG performance does not determine cross-modal bridge performance: CSBrain and BENDR improve after bridging, whereas LaBraM, EEGPT, and REVE undergo substantial degradation and produce strongly class-biased predictions. This ranking mismatch motivates treating a bridge pathway as the unit of selection rather than selecting a receiver solely from its native-EEG performance.

A.4 FOG Output-Level Transfer Loss and Representation Distance

We compare, for every paired instance and receiver, the class distribution p^e_{ik} produced by direct EEG with the class distribution p^b_{ik} produced after bridge injection. We report Jensen–Shannon (JS) divergence, prediction-flip rate, and the AUC obtained when JS divergence is used to detect a bridge failure conditional on a correct EEG teacher. As shown in Table 7, output discrepancy is highly receiver-dependent. JS divergence is especially diagnostic for EEGPT and REVE, but sub-

Table 6: FOG direct-EEG and bridged-EMG balanced accuracy on the final test split. Values are percentages.

Receiver	Direct EEG	Bridged EMG	Bridge − direct
LaBraM	66.02	49.98	−16.04
EEGPT	72.89	51.09	−21.80
REVE	72.59	50.41	−22.18
BIOT	76.89	69.89	−7.00
CBraMod	60.54	58.02	−2.52
CSBrain	58.25	64.05	+5.80
BENDR	58.50	65.96	+7.46

Table 7: FOG output-level transfer diagnostics on the final test split.

Receiver	EEG BAcc	Bridge BAcc	Mean JS	Flip rate	JS AUC_{fail}
LaBraM	66.02	49.98	0.0719	37.32%	0.9449
EEGPT	72.89	51.09	0.1190	44.26%	0.9665
REVE	72.59	50.41	0.1287	57.48%	0.9904
BIOT	76.89	69.89	0.0554	32.39%	0.8260
CBraMod	60.54	58.02	0.1759	48.65%	0.8311
CSBrain	58.25	64.05	0.0149	38.16%	0.6365
BENDR	58.50	65.96	0.0025	40.87%	0.6755

stantially less discriminative for CSBrain and BENDR. A single uncalibrated discrepancy threshold is therefore unlikely to transfer across heterogeneous receivers.

For BIOT, we additionally compare the bridged representation with the paired direct-EEG representation at bridge destination layer 2 and after propagation through the remaining receiver layers. Table 8 shows nearly identical destination-layer discrepancy across the paired, validation, and test splits. The remaining BIOT layers contract rather than amplify this discrepancy: test cosine distance decreases from 0.226 at the destination to 0.134 at the final representation. Nevertheless, a nontrivial output discrepancy remains, with mean test JS divergence 0.055.

A.5 Collapsed Routes and Compact-Portfolio Headroom

We define a route as collapsed using validation data only when one predicted class accounts for at least 95% of its outputs and validation balanced accuracy is below 55%. This criterion removes LaBraM, EEGPT, and REVE and retains BIOT, CBraMod, CSBrain, and BENDR. Table 9 reports the validation-defined decision together with test behavior.

The full seven-route oracle reaches 99.96% test balanced accuracy and misses only three instances. Removing the three collapsed routes reduces the oracle to 93.69% and increases the number of jointly missed instances to 514. Thus, collapsed routes inflate the full oracle, but they do not explain away portfolio complementarity: the compact-four oracle remains 23.80 points above the BIOT SBS. When the target-only MLP is added as the default predictor, the compact-four residual oracle reaches $97.07 \pm 0.08\%$ balanced accuracy across the five MLP seeds, indicating that the retained bridges can correct 82–83% of the target-only MLP errors.

A.6 FOG Supervised Controls and Higher-Capacity Routing

Table 10 compares the proposed gain-aware router with target-only classifiers and a higher-capacity strict EEG-AS adaptation. All stochastic results use seeds 42–46 and report sample standard deviation. The target-only MLP is the strongest deployable method in this diagnostic. This result shows that the frozen target representation itself contains substantial task information and limits any claim that bridge routing is universally necessary. BridgeRouter instead establishes the narrower result that anchor-relative pathway supervision consistently improves the predefined SBS while retaining single-route bridge execution.

Strict EEG-AS jointly predicts missing route tokens and ranks the seven pathways. It improves over the simpler gain router by 0.81 balanced-accuracy points on average, but it also increases mean route execution from 1.00 to 1.75. Because its corresponding ISRUC evaluation was not completed under

Table 8: BIOT representation-distance diagnostics. "Pair" is the bridge training domain; validation and test are held out from bridge fitting.

Split	Dest. NMSE	Dest. cosine	Final NMSE	Final cosine	Output JS
Pair	0.4915	0.2249	0.2428	0.1177	0.0520
Validation	0.4947	0.2263	0.2768	0.1341	0.0565
Test	0.4969	0.2262	0.2766	0.1341	0.0554

Table 9: Validation-defined FOG route-collapse audit.

Receiver	Val. dominant	Val. BAcc	Collapsed	Test dominant	Test BAcc
LaBraM	99.95%	50.06	Yes	99.95%	49.98
EEGPT	97.63%	51.38	Yes	97.88%	51.09
REVE	99.03%	50.59	Yes	98.98%	50.41
BIOT	52.79%	70.56	No	52.30%	69.89
CBraMod	53.69%	58.67	No	53.48%	58.02
CSBrain	59.11%	63.24	No	58.76%	64.05
BENDR	50.67%	64.75	No	50.85%	65.96

the corrected CBraMod-SBS protocol, we treat it as a FOG-only higher-capacity diagnostic rather than the primary cross-dataset method.

A.7 ISRUC Development-Stage Sensitivity Analyses

To assess sensitivity to the routing target, portfolio size, and source representation, we report additional TRAIN-to-VAL analyses from the development stage using a fixed anchor and seeds 42–46. Thresholds and checkpoints are selected on VAL, and TEST is not accessed in any of these analyses.

Routing target and SBS protection. Table 11 compares controlled changes to the training target, router capacity, and protection rule while holding the earlier H4/K7 development configuration fixed. Within this configuration, the NLL-gain MLP with validation-selected protection gives the highest mean VAL balanced accuracy. Correctness-gain routing selects thresholds that suppress switching, whereas forcing alternative routing with the same NLL-gain scores slightly reduces performance relative to retaining the anchor prediction. Overall, these analyses support the use of validation-selected protection and show that gains depend on the routing objective rather than switching alone.

Portfolio size. The predecessor portfolio study used a nested sequence constructed without TEST access. Starting from a fixed anchor route, each additional route was selected by the largest TRAIN subset-oracle balanced accuracy under a fixed tie rule. Table 12 shows that oracle balanced accuracy increases monotonically as routes are added, whereas deployable router performance is non-monotonic. The strongest exploratory router result occurs at K4, while further increases in oracle headroom do not translate directly into higher deployable performance. These development-stage comparisons are sensitivity analyses only; neither the K4 result nor the exploratory nested ordering replaces the final canonical K7 portfolio.

Source representation. Table 13 evaluates the source representation while holding the predecessor K7 configuration fixed. Performance increases gradually from H0 to a broad H3–H5 plateau and then declines. The difference between H4 and H5 is only 0.0491 balanced-accuracy points, smaller than the observed seed dispersion. Concatenating all eight layers increases the source representation dimension from 512 to 4096 without improving on the single-layer plateau, indicating limited sensitivity to the exact layer choice within this range.

A.8 Rescue-Harm Diagnostics

The FOG router recovers many more BIOT errors than it destroys BIOT successes (Table 14). ISRUC presents a less stable tradeoff. For seeds 42–45, the router switches conservatively and achieves rescue precision between 75.6% and 94.0%. Seed 46 switches much more aggressively and its rescue precision falls to 42.2%, explaining much of the variance in the aggregate ISRUC result.

Table 10: FOG test controls. "Routes" counts executed bridge pathways and does not include the frozen target encoder.

Method	BAcc	Δ vs. BIOT	Routes
BIOT SBS	69.89	–	1.00
Target-only linear	79.74 ± 0.17	$+9.85$	0
BridgeRouter	81.38 ± 1.11	$+11.49$	1.00
Strict EEG-AS adaptation	82.19 ± 0.47	$+12.30$	1.75
Target-only MLP	83.40 ± 0.34	$+13.51$	0
K7 oracle	99.96	$+30.07$	7.00

Table 11: ISRUC routing-target and protection sensitivity under the earlier H4/K7 development configuration. Results are VAL percentages over seeds 42–46; TEST was not accessed. ΔBAcc is measured relative to the fixed anchor baseline (64.9376 VAL balanced accuracy).

Condition	Accuracy	BAcc	ΔBAcc
NLL gain, MLP, protected	78.6018 ± 0.3586	$\mathbf{65.5515 \pm 0.2820}$	$+0.6140$
Correctness gain, MLP, protected	81.5599 ± 0.0000	64.9376 ± 0.0000	$+0.0000$
True-class-probability gain, MLP, protected	81.2875 ± 0.1775	65.0438 ± 0.1083	$+0.1062$
NLL gain, linear, protected	78.7423 ± 1.5152	65.1418 ± 0.0914	$+0.2042$
NLL gain, MLP, forced alternative	77.4569 ± 0.6811	64.9125 ± 0.2242	-0.0250

A.9 Router Optimization and Reproducibility Details

Router architecture and optimization. The router input is the pooled NormWear block-6 representation on FOG (768 dimensions) and the pooled HuBERT-ECG block-4 representation on ISRUC (512 dimensions). The MLP uses hidden widths 256 and 128, GELU activations, dropout 0.2, and one output for each non-anchor route. We optimize with AdamW (learning rate 3×10^{-4}, weight decay 10^{-4}), batch size 256, a maximum of 100 epochs, gradient-norm clipping at 5, and patience 15. All reported stochastic results use seeds 42–46.

For each route, NLL gains are clipped to their training-set 1st and 99th percentiles and standardized before smooth-L_1 regression. Instances are weighted inversely to training class frequency. At each epoch, the protection threshold is selected on validation data from zero, 101 quantiles of the nonnegative maximum predicted gain, and a value immediately above the observed maximum. Balanced accuracy determines the checkpoint; when balanced accuracy ties, the larger threshold is preferred, yielding the more conservative policy. No threshold or checkpoint is reselected on the test set.

Bridge insertion locations. Table 16 lists the NormWear-to-EEG layer pairs used for FOG. Each target encoder and EEG backbone is frozen while the bridge is fit on paired data. The same validation-selected bridge checkpoint is then used for all routing experiments.

Protocol safeguards. The SBS is selected by validation balanced accuracy and frozen before test evaluation. This gives BIOT on FOG and CBraMod on ISRUC. In particular, EEGPT and BENDR are not retrospectively promoted to the ISRUC SBS despite their stronger test performance. All route probabilities and target features are aligned by persistent sample identifier, and train, validation, and test sample identifiers are checked for overlap. The ISRUC split is subject disjoint; the FOG BioX split is sample-wise and contains the same 12 participants in all four subsets. The portfolio oracle uses the ground-truth label and is reported only as a non-deployable upper bound.

A.10 Inference Requirements and Computational Cost

Table 17 distinguishes information used to train the router from information required at deployment. The gain targets used to train BridgeRouter are task-label-derived, but the trained router requires neither task labels nor the paired EEG modality at inference. It also does not execute candidate routes to make the routing decision. Instead, the frozen target encoder is evaluated once, the router predicts all anchor-relative gains from that representation, and exactly one selected bridge pathway is executed. The K7 oracle evaluates every route and uses the ground-truth label, and is therefore reported only as an upper bound.

Table 12: ISRUC portfolio-size sensitivity under the earlier development configuration. Nested subsets were constructed using TRAIN only and evaluated on VAL; TEST was not accessed. Router values are mean ± sample standard deviation over seeds 42–46. Route identities are shown only to document the deterministic nested construction.

K	Nested portfolio	Oracle BAcc	Router BAcc	ΔBAcc
1	BENDR	64.9376	64.9376	+0.0000
2	+ EEGPT	71.0984	64.9541 ± 0.0369	+0.0165
3	+ LaBraM	72.7167	64.9376 ± 0.0000	+0.0000
4	+ CBraMod	75.9408	$\mathbf{65.7651 \pm 0.2580}$	+0.8276
5	+ CSBrain	76.4753	65.6125 ± 0.1450	+0.6749
6	+ REVE	77.7191	65.6101 ± 0.3183	+0.6725
7	+ BIOT	78.1662	65.5515 ± 0.2820	+0.6140

Table 13: ISRUC HuBERT-ECG source-representation sensitivity under the earlier K7 development configuration. Results are VAL balanced accuracy over seeds 42–46; TEST was not accessed.

Representation	Dimension	BAcc
H0	512	65.2908 ± 0.0726
H1	512	65.3383 ± 0.1490
H2	512	65.4294 ± 0.0512
H3	512	65.5254 ± 0.2802
H4	512	65.5515 ± 0.2820
H5	512	$\mathbf{65.6006 \pm 0.2484}$
H6	512	65.3949 ± 0.2973
H7	512	65.3177 ± 0.2331
H0–H7 concatenated	4096	65.5234 ± 0.1001

For the seven-route portfolio, the router predicts six gains because the SBS is represented implicitly by the fallback decision. With hidden widths 256 and 128, the router contains approximately 0.231M parameters for the 768-dimensional FOG input and 0.165M parameters for the 512-dimensional ISRUC input. This overhead is small relative to the frozen encoders and EEG foundation models. Most importantly, adaptive routing does not multiply bridge-pathway inference cost by the portfolio size.

Table 14: FOG per-seed rescue–harm counts relative to the BIOT SBS.

Seed	Rescued errors	Harmed successes	Net correct	Switch rate
42	1430	385	+1045	68.73%
43	1314	357	+957	66.58%
44	1336	440	+896	67.40%
45	1511	384	+1127	73.93%
46	1497	387	+1110	72.99%

Table 15: ISRUC switching diagnostics relative to the CBraMod SBS.

Seed	False switch|SBS correct	Attempted rescue|SBS wrong	Rescue precision	Net correct
42	2.42%	9.24%	91.64%	+86
43	3.63%	12.18%	85.74%	+154
44	4.05%	14.01%	75.56%	+139
45	1.69%	2.80%	93.97%	−113
46	47.29%	53.40%	42.19%	+274

Table 16: FOG bridge source and destination layers.

Receiver	NormWear source block	EEG destination layer
LaBraM	4	1
EEGPT	5	5
REVE	7	21
BIOT	6	2
CBraMod	4	6
CSBrain	9	6
BENDR	9	6

Table 17: Information and pathway execution required at inference. The target encoder is shared by SBS and BridgeRouter and is not counted as a bridge pathway.

Method	Target signal	Paired EEG	Task label	Bridge routes
Validation-selected SBS	Yes	No	No	1
BridgeRouter	Yes	No	No	1
K7 oracle	Yes	No	Yes	7