WHICH DATA, NOT JUST HOW MUCH: A CONTROLLED AUDIT OF EEG FOUNDATION MODEL PRETRAINING

Anonymous authors
Paper under double-blind review

ABSTRACT

Electroencephalography (EEG) foundation models have achieved strong performance across diverse tasks. Despite differences in architecture and training objectives, most share a common strategy: pretrain on increasingly large data pools to learn better general-purpose representations. This strategy is well established in natural language processing and computer vision, but whether more pretraining data consistently improves EEG transfer remains an open question. We examine 16 candidate datasets, 13 downstream tasks, and 202 configurations across LaBraM, CBraMod, and BIOT, keeping the backbone and pretraining epoch count fixed within each architecture. Adding sources can lower average performance: 121 configurations achieve higher mean test balanced accuracy than full-pool pretraining with fewer training windows. All 39 architecture–task combinations have a smaller, higher-scoring configuration; even adding sources from the target's task family or its own training partition need not help. Similarly sized mixtures yield different benchmark averages and opposing responses on related targets. Capacity experiments show recovery under further expansion at fixed Base capacity; increasing LaBraM to Large narrows the full-pool deficit but leaves a six-source subset ahead by 2.84 percentage points. We compare four practical rankings—Progressive, pairwise dataset valuation (PDV), Size, and Coverage—that identify useful mixtures at different scoring costs, with Size and Coverage sharing their rankings across LaBraM and CBraMod. The findings show that source composition matters alongside data volume and support choosing pretraining datasets using appropriate data-selection methods.

1 INTRODUCTION

Electroencephalography (EEG) foundation models (EEG-FMs) have become the dominant approach to general-purpose representation learning on EEG recordings. Representative systems—LaBraM (Jiang et al., 2024b) (masked vector-quantized token prediction), CBraMod (Wang et al., 2024b) (masked reconstruction), BIOT (Yang et al., 2023) (contrastive), and a growing list of follow-ups (Kostas et al., 2021; Chen et al., 2024; Cui et al., 2024; Wang et al., 2023; Zhang et al., 2023; Wang et al., 2024a; Yuan et al., 2024; Cai et al., 2023; Ouahidi et al., 2025)—differ in backbone and objective but share a common procedure inherited from natural language processing (NLP) and computer vision (CV): pool as many public EEG datasets as possible, pretrain on the union, then fine-tune. Each new model reports gains together with a larger pool, implicitly suggesting that adding more sources is the natural way forward.

This recipe is well-grounded in language and vision, where pretraining on ever-larger heterogeneous corpora yields predictable, often log-linear returns. EEG sits in a different regime: public corpora amount to only tens of thousands of hours (Ouahidi et al., 2025; Xiong et al., 2025), orders of magnitude below the trillion-token / billion-image scale of NLP and CV (Grattafiori et al., 2024), while individual datasets differ sharply in channel count (3–64), sampling rate, paradigm (clinical, sleep, motor, emotion, cognitive), and population, so whether the "more is better" transfers to EEG-FMs remains unsettled. Recent evidence pulls both ways: Banville et al. (2025) report log-linear gains in image decoding as brain-recording hours grow, while EEG-FM-Bench (Xiong et al., 2025) finds model scaling deviates from typical laws (compact, domain-biased architectures beat much

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.

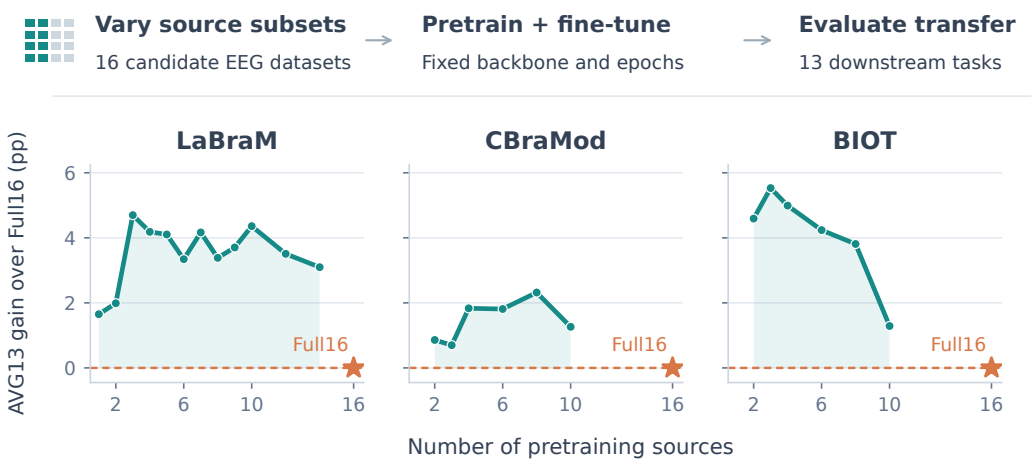


Figure 1: **Expanding the pretraining pool does not guarantee better transfer.** We vary source subsets within a fixed backbone and epoch schedule, then evaluate transfer on 13 tasks. Curves add higher-ranked sources first and show the change in average test balanced accuracy from the full 16-source pool (Full16), in percentage points. Each peaks before the full pool. Stars and dashed lines mark Full16; segments connect evaluated prefixes.

larger ones; gradient conflicts bottleneck pretraining) and AdaBrain-Bench (Wu et al., 2025) reports strongly task-dependent downstream gains; BIOT (Yang et al., 2023) was originally pretrained on only four clinical/sleep datasets, and PRISM (Lahiri et al., 2026) finds targeted diversity substitutes for indiscriminate scale. Yet across these threads, a systematic study of the *pretraining-data* axis itself—which EEG datasets to include, and how their interactions shape downstream performance under fixed backbone and schedule—is still missing, leaving the field adding datasets without controlled evidence on whether doing so helps.

We study this question with 16 candidate pretraining datasets and 13 downstream tasks across LaBraM, CBraMod, and BIOT. Within each architecture, we vary the source pool while keeping the backbone and pretraining epoch count fixed. Figure 1 follows pools expanded by adding higher-ranked datasets first under a dataset-level score. Average performance shows an overall rise followed by a decline, peaking at three, eight, and three sources, respectively. These downturns suggest that some additions offer no further average benefit.

The broader comparison covers 202 configurations spanning different source counts and compositions (Figure 2). Of these, 126 achieve higher average test balanced accuracy than full-pool pretraining, and 121 of those also use fewer training windows. The best evaluated subsets exceed the full pool by 5.58 percentage points for LaBraM, 3.30 for CBraMod, and 5.53 for BIOT. BIOT even performs better from scratch, reaching 54.74% compared with 50.45% after full-pool pretraining. These findings extend to individual tasks: all 39 architecture–task combinations have a smaller configuration with a higher mean. Adding sources from the target's own task family, or even its own training partition, does not guarantee improvement on that target.

Because expansion changes both volume and composition, we also compare mixtures with similar training-window counts. Two LaBraM pools differ by less than 9% in window count but by 7.54 points in average performance; related targets absent from both pools favor different mixtures. Composition therefore matters for both overall and task-level performance.

The losses from pool expansion could reflect insufficient model capacity. At fixed LaBraM Base capacity, one addition lowers performance, but another retaining all earlier sources raises it above the starting pool. Increasing LaBraM to Large improves the full pool more than a fixed six-source subset, narrowing their gap, but the subset retains a 2.84-point advantage. Capacity affects the comparison, while the importance of source composition persists in the larger model.

To inform dataset choice in practice, we compare four source-ranking strategies: Progressive ordering, pairwise dataset valuation (PDV), Size ordering, and Coverage ordering. Progressive ranks sources

by validation performance after single-source pretraining, identifying higher-average subsets for all three architectures. PDV assesses joint pretraining through validation gains on partner tasks over partner-only pretraining. Its observed peaks exceed Progressive's by 0.36 and 0.80 points on LaBraM and CBraMod, while the number of scoring models grows from linear to quadratic in the number of sources. Size ranks whole-corpus window counts; Coverage measures the fraction of a fixed LaBraM tokenizer's codes activated at least twice by source probes. Both construct rankings without downstream labels and reuse the same order across LaBraM and CBraMod, finding subsets with higher averages than the full pool. These comparisons offer practical guidance on selection cost and transfer performance.

2 Related Work

EEG foundation models learn transferable representations through objectives such as token prediction, reconstruction, and contrastive learning (Jiang et al., 2024b; Wang et al., 2024b; Yang et al., 2023). EEG-FM-Bench and AdaBrain-Bench show that downstream gains depend on the task and evaluation protocol (Xiong et al., 2025; Wu et al., 2025), while PRISM examines how population composition affects performance under a fixed architecture (Lahiri et al., 2026). Evidence from heterogeneous mouse spiking recordings shows that selected pretraining sessions can outperform a larger pool (Jiang et al., 2025), whereas OmniMouse reports gains from increasing neural data (Willeke et al., 2026). Data curation and mixture design have also been studied extensively in language and vision (Gadre et al., 2023; Abbas et al., 2023; Tirumala et al., 2023). These studies motivate examining what enters a pretraining pool alongside how large it is. We vary EEG dataset membership within each backbone, examining how pool expansion and composition affect benchmark averages and individual tasks, then comparing source rankings as practical responses. The contribution is the resulting evidence about EEG source composition across architectures and tasks, with the ranking comparisons showing how that evidence can inform mixture construction.

3 Experimental Setup

The 16-source pool spans five categories: clinical EEG—TUAB (Obeid & Picone, 2016), TUEV (Harati et al., 2015), TUSL (von Weltin et al., 2017), and Siena (Detti et al., 2020); sleep staging—HMC (Alvarez-Estevez & Rijsman, 2021); motor imagery—PhysioNet (Schalk et al., 2004), SHU-MI (Ma et al., 2022), BCIC-IV-2a (Tangermann et al., 2012), and BCIC-2020/3 (Jeong et al., 2020); emotion recognition—SEED (Zheng & Lu, 2015), SEED-IV (Zheng et al., 2018), SEED-V (Liu et al., 2021), SEED-VII (Jiang et al., 2024a), and FACED (Chen et al., 2023); and cognitive tasks—Workload (Wilson & Russell, 2003) and in-house SCD.

Full16 uses all 16 sources; Scratch uses no pretraining. AVG13 averages test balanced accuracy (%) over three fine-tuning seeds per task, then equally across 13 tasks, excluding SEED-V, SCD, and BCIC-IV-2a. Appendix B gives the shared pretraining and fine-tuning protocols.

4 Data Scale and Source Composition

4.1 More data does not guarantee better overall performance

Expanding the pretraining pool is a common way to seek better transfer across downstream tasks. To test whether this strategy consistently helps, we track AVG13 for each architecture as sources are added in descending order of their single-source validation performance (Figure 1). The highest observed AVG13 occurs at three, eight, and three sources for LaBraM, CBraMod, and BIOT, respectively; every larger evaluated prefix in each sweep scores below that peak. These downturns suggest that some source additions may provide no further benefit to average downstream performance, or may even reduce it.

To examine such additions directly, we compare nested pools that differ by one source. For LaBraM, adding SEED-IV to TUAB, Siena, and SEED-VII increases the pool from 371,004 to 387,356 windows, while AVG13 falls from 61.72% to 60.84%. For CBraMod, adding Siena to PhysioNet, HMC, TUAB, and TUEV increases the pool from 472,988 to 506,593 windows, while AVG13 falls from 58.18% to 56.04%. Both expanded pools retain every previous source and use the same

backbone, objective, and epoch count as their smaller counterparts. In these mixtures, the additional dataset is accompanied by a lower benchmark mean despite the increase in available training data.

The broader configuration inventory shows that these examples are part of a recurring pattern. Relative to Full16, 121 of the 202 evaluated configurations have higher AVG13 with fewer training windows (Figure 2). Of these, 19 use less than one quarter of Full16's windows: three for LaBraM, four for CBraMod, and twelve for BIOT, including Scratch. The highest observed subset averages reach 61.72%, 58.36%, and 55.98%, exceeding Full16 by 5.58, 3.30, and 5.53 points, respectively. The evidence also extends beyond differences in means: among the reported subset–Full16 tests, 20 of 23 LaBraM configurations and five of six BIOT configurations show significant positive differences after their respective within-family Holm corrections; none of the 19 tested CBraMod configurations passes correction (Appendix D.1). These counts include each reused run only once. For BIOT, even Scratch reaches 54.74%, above Full16's 50.45%. Thus, the deficit of the full pool appears across many configurations, including some that use only a small fraction of its data, rather than depending on one favorable subset.

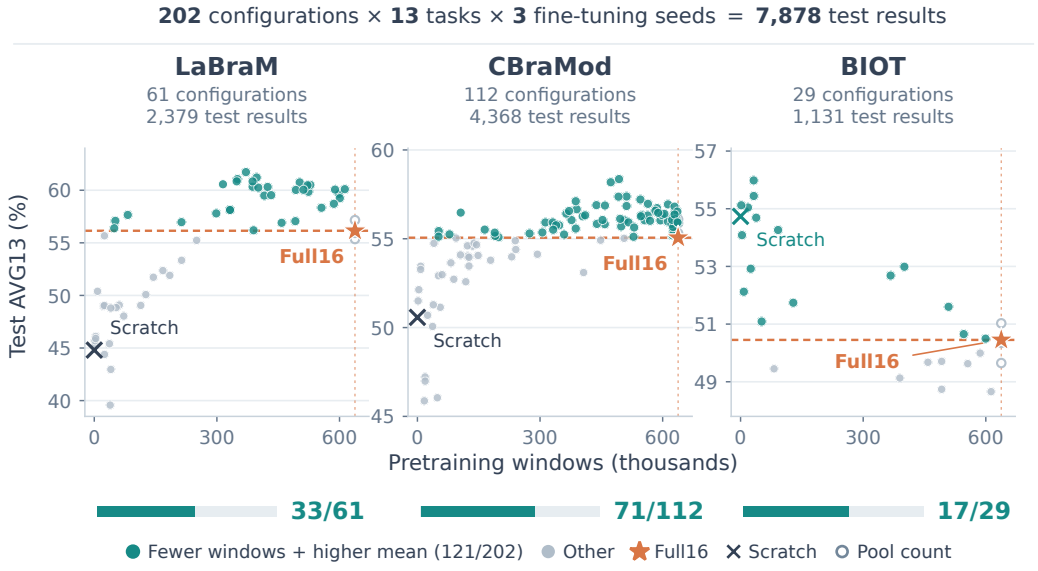


Figure 2: **More pretraining data is not always better.** The 202 configurations summarize 7,878 downstream test results: 13 tasks and three fine-tuning seeds per configuration. Teal marks fewer windows and a higher mean than Full16; bars count these configurations among all those evaluated. Stars and dashed lines mark Full16; crosses mark Scratch. Open circles and the two reused pair-pretraining runs use candidate-pool counts, which need not equal retained training windows. Each panel has its own labeled vertical scale. Colors compare means and do not indicate significance.

4.2 More data does not guarantee better task-level performance

Changes in overall performance do not necessarily reflect changes on individual downstream tasks. We therefore examine each target separately across the complete configuration inventory. Full16 never achieves the highest observed test mean on any of the 13 tasks for any of the three architectures: all 39 architecture–task combinations have a higher-scoring configuration with fewer pretraining windows (Figure 3a). Such alternatives are numerous on several targets. They account for 48 of 61 configurations on LaBraM's Siena task, 106 of 112 on CBraMod's SEED task, and 23 of 29 on BIOT's TUSL task. The counts vary across targets—CBraMod has ten such configurations on TUAB—but every target has examples. More pretraining data therefore does not guarantee better performance at the individual-task level either.

Adding data from the same task family can improve one target while hurting another (Figure 3b). In LaBraM, adding SEED-VII to TUAB and Siena improves all four emotion-recognition tasks by 5.24–8.46 points. Adding SEED-IV next lowers FACED by 7.58 points, while SEED and SEED-VII

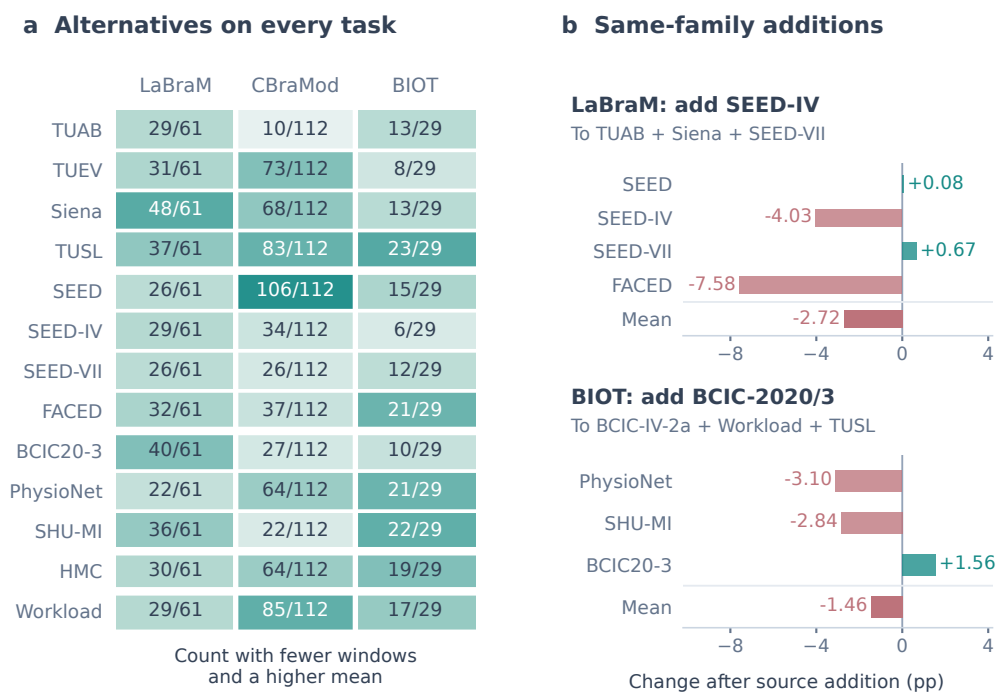


Figure 3: **More data need not improve individual tasks, even when the added source belongs to the same task family.** (a) Configurations with fewer windows and a higher task mean than Full16, divided by all configurations including Full16 and Scratch. Darker cells indicate larger fractions. (b) Changes after adding an emotion or motor-imagery source to a pool already containing that family. Some targets improve, but the family's mean falls; adding SEED-IV also lowers its own downstream score. Differences use the smaller pool as reference. All scores are three-seed means; numerical gains do not denote statistical significance.

improve by only 0.08 and 0.67 points; the four-task average declines by 2.72 points. Motor-imagery additions also produce mixed responses. Adding BCIC-2020/3 to BIOT's BCIC-IV-2a, Workload, and TUSL pool raises the BCIC-2020/3 score by 1.56 points but lowers PhysioNet and SHU-MI by 3.10 and 2.84 points. A shared task category thus does not ensure consistent gains across related targets.

Even adding training-partition data from the downstream target itself does not guarantee a benefit on that target. For SEED-IV emotion recognition, LaBraM pretrained on TUAB, Siena, and SEED-VII achieves 75.79% test balanced accuracy after fine-tuning on SEED-IV. Adding the SEED-IV training partition to this three-source pretraining pool lowers the SEED-IV test mean to 71.76%, a decrease of 4.03 points. For SHU-MI motor imagery, the initial LaBraM pretraining pool contains TUAB, Siena, SEED-IV, SEED-VII, SEED, PhysioNet, and FACED. After fine-tuning on SHU-MI, this model achieves 61.58% test balanced accuracy. Adding the SHU-MI training partition to those seven pretraining sources lowers the SHU-MI test mean to 61.15%, a decrease of 0.43 points. In each comparison, the downstream task and its data splits remain fixed; only the target's training partition is added to pretraining. Including the target source is therefore not sufficient to improve its downstream test performance. Table 22 reports the source additions and task changes; Appendices C and D give detailed task results and the complete configuration distributions.

4.3 Similar data volumes can yield different downstream performance

The preceding additions change both the amount and the composition of pretraining data. To examine composition at a comparable scale, we compare different source combinations with similar training-window counts within each architecture. LaBraM reaches 48.84% AVG13 with SEED, PhysioNet,

and the in-house cognitive dataset SCD (53,552 windows), versus 56.38% with FACED, TUSL, Workload, BCIC-IV-2a, SCD, and SEED-V (48,783 windows). The second pool uses 8.9% fewer windows but scores 7.54 points higher on average. This advantage does not carry over uniformly to other emotion datasets. The mixtures draw their emotion data from SEED and from FACED and SEED-V, respectively, leaving SEED-IV and SEED-VII outside both pools. On these two targets, the higher-average mixture lowers SEED-IV test balanced accuracy from 67.04% to 63.99%, but raises SEED-VII from 21.51% to 30.39%.

TUEV and TUSL favor different CBraMod mixtures even when the pretraining volumes are almost identical. A pool of HMC, SHU-MI, SEED-VII, TUAB, and BCIC-IV-2a contains 439,126 windows; PhysioNet, TUAB, Siena, and HMC contain 437,937, a difference of only 0.27%. The two pools share the clinical source TUAB, with Siena also present in the second, but include neither TUEV nor TUSL. Moving to the second mixture raises TUEV test balanced accuracy from 51.37% to 55.30%, while TUSL falls from 78.79% to 75.23%. These opposing task-level shifts are larger than the change in AVG13, from 55.84% to 56.89%.

An overall performance gap also appears in BIOT when source count is held fixed. BIOT scores 49.13% AVG13 with TUAB and HMC (389,716 windows), but 52.68% with TUAB and TUEV (367,124 windows). Replacing HMC with TUEV accompanies a 3.55-point increase with 5.8% fewer windows.

These comparisons show that similar amounts of pretraining data can yield substantially different benchmark averages and individual-task scores. Even among targets from the same task family that are absent from both pretraining pools, the higher-average combination can score better on one task and worse on another.

5 Source Composition and Model Capacity

The preceding section shows that increasing pretraining data volume does not consistently improve downstream performance. One possible explanation is that Base models have insufficient capacity to benefit from the additional data. We first examine whether further source expansion can reverse a decline at fixed Base capacity. We then compare Base and Large using two fixed source pools and the same training protocol.

5.1 Recovery under further source expansion at Base capacity

A decline after adding data may suggest that the Base model has reached a limit on the data it can use. We examine what happens when the pretraining pool grows further while retaining every source already included. The comparison follows three nested pools, each pretrained separately with the same LaBraM Base architecture, objective, and training protocol.

Pretraining on SCD, SEED-V, BCIC-IV-2a, and Workload gives 48.80% AVG13 from 40,403 windows. Adding TUSL increases the pool to 40,551 windows, but AVG13 falls to 42.96%. We then examine a six-source pool that retains all five sources and also includes FACED. With 48,783 windows, it reaches 56.38%, exceeding both smaller pools (Figure 4a). Appendix D.3 gives the complete recipes and further nested examples across the three architectures.

The first addition lowers AVG13 by 5.84 points; the next raises it by 13.42 points, leaving the largest pool 7.58 points above the starting pool. This recovery occurs without increasing model capacity or removing TUSL: both the number of sources and the number of training windows continue to grow. The initial decline therefore does not mark a point beyond which this Base model can no longer benefit from additional data. Whether a larger model also makes better use of the full pool remains a separate question, which we examine next.

5.2 Increasing capacity from Base to Large

We next compare LaBraM Base and Large on Full16 and a fixed six-source pool containing TUAB, Siena, SEED-VII, SEED-IV, BCIC-IV-2a, and FACED. The subset is fixed from the Base experiments. Base and Large use the same preprocessing, pretraining objective, 50 pretraining epochs, effective batch size of 64, downstream splits, and fine-tuning and model-selection procedures. This comparison

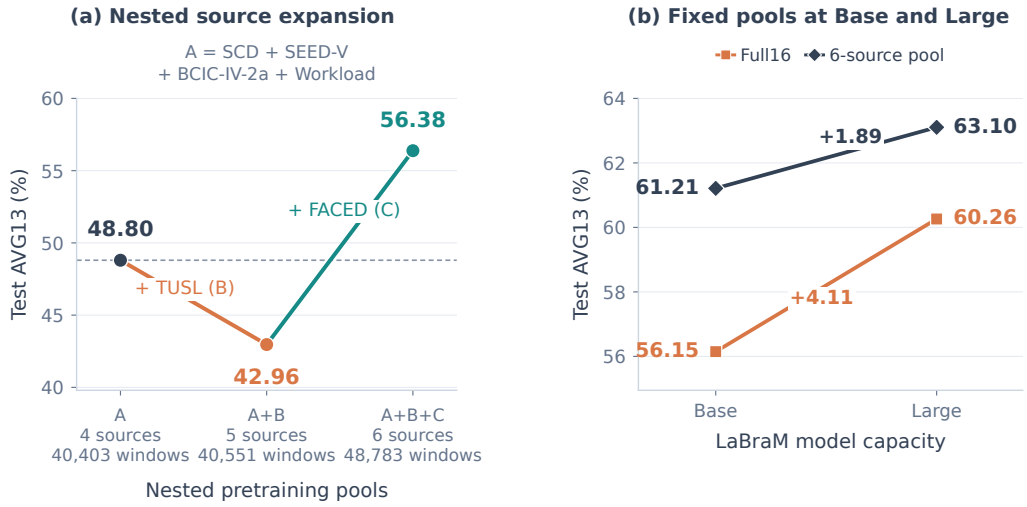


Figure 4: **Further expansion can recover a Base decline; increasing capacity narrows the full-pool deficit.** (a) LaBraM Base on three nested pools. A contains SCD, SEED-V, BCIC-IV-2a, and Workload; B is TUSL and C is FACED. Adding B lowers AVG13, but retaining B and adding C raises it above A. The horizontal positions denote expansion stages, with window counts shown below. (b) A fixed six-source pool and Full16 under matched Base and Large protocols (Section 5.2). The panels have separate vertical scales. Lines compare separately pretrained models; points show AVG13 over three fine-tuning seeds per target and do not encode significance.

changes model capacity without changing the recipe or training protocol; Appendix C.2 reports the complete task results.

Full16 rises from 56.15% AVG13 in Base to 60.26% in Large, a gain of 4.11 points (Figure 4b). The six-source pool improves from 61.21% to 63.10%, a gain of 1.89 points. Full16 therefore benefits more from the capacity increase, and its gap to the subset shrinks from 5.06 to 2.84 points.

The six-source pool improves on Large Full16 in eleven of the thirteen task means, and the subset–Full16 contrast passes Holm correction ($p_H = 0.00513$; Appendix C.2). Full16's larger gain indicates that capacity affects the comparison, while the remaining subset advantage shows that its disadvantage is not confined to Base. These two model sizes do not establish whether further scaling would close the gap. Together with the recovery under further source expansion at fixed capacity, the results motivate evaluating which sources to combine at the intended model size.

6 Selecting Pretraining Sources

The preceding experiments show that more pretraining data does not necessarily lead to better downstream performance, and that the composition of the pretraining pool also has a substantial influence. We compare Progressive ordering, pairwise dataset valuation (PDV), Size ordering, and Coverage ordering, examining both transfer under pool expansion (Figure 5) and the cost of constructing each ranking.

6.1 Progressive ordering

A direct starting point is to combine sources that transfer well individually. Progressive ranks sources by mean validation balanced accuracy after single-source pretraining, evaluated over 14 tasks: AVG13 plus BCIC-IV-2a. The evaluated P-Hk mixtures jointly pretrain on the first k sources in the architecture-specific sequences specified in Appendix E.

Progressive peaks at 60.85% (LaBraM P-H3), 57.38% (CBraMod P-H8), and 55.98% (BIOT P-H3), respectively 4.70, 2.32, and 5.53 points above Full16. Higher means on 13, nine, and twelve of the

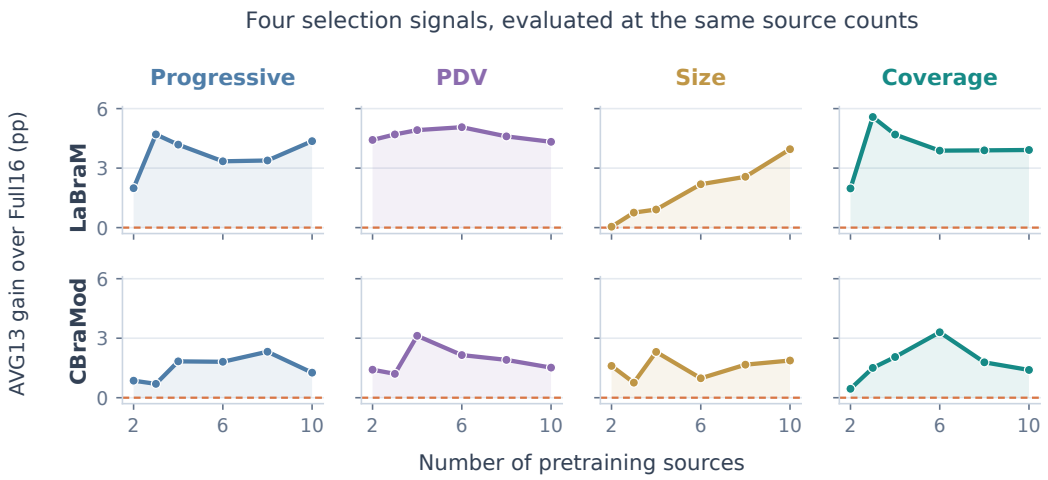


Figure 5: **Rankings guide useful mixtures, but expansion need not improve transfer.** Columns compare Progressive, PDV, Size, and Coverage; rows show LaBraM and CBraMod. Values are AVG13 differences from Full16, used as the common reference (dashed zero line), in percentage points. Lines connect observations at $k = 2, 3, 4, 6, 8, 10$. Appendix E.5 includes additional prefixes and low-ranked controls; Appendix C gives key configurations' task scores.

13 tasks (Appendix F) support standalone transfer as a useful selection signal. Yet peak source counts differ across architectures, and further expansion can lower AVG13 (Figure 5). We next assess what a source contributes when pretrained with a partner.

6.2 Pairwise dataset valuation

PDV ranks 14 sources (excluding SCD and SEED-V) by their mean validation gain on partner tasks from pair versus partner-only pretraining. We test the top-k mixtures on AVG13 (Appendix E).

PDV peaks at 61.21% (LaBraM PDV6) and 58.18% (CBraMod PDV4), respectively 5.06 and 3.12 points above Full16 (Table 27). These peaks exceed Progressive's by 0.36 and 0.80 points; most of the observed improvement over Full16 is already obtained with singleton-based ranking. At matched source counts, no PDV–Progressive contrast survives Holm correction (Table 34).

6.3 Size ordering

Size ranks sources by whole-corpus window counts at each source's own window length, rounded to thousands. Size-Hk/Size-Lk select the k largest/smallest sources; Size-H4 contains TUAB, HMC, TUEV, and Siena. These counts differ from staged training-window exposure (Appendix E).

Size-H10 reaches 60.10% for LaBraM and Size-H4 reaches 57.37% for CBraMod, respectively 3.95 and 2.31 points above Full16. Size nearly matches Progressive's peak mean on CBraMod without downstream source scoring. Under the same ranking, LaBraM improves across the tested prefixes while CBraMod peaks at four sources (Figure 5).

6.4 Coverage ordering

Coverage measures another source property: how broadly a probe activates a common learned codebook. It ranks sources by the fraction of a fixed LaBraM tokenizer's 8,192 codes activated at least twice. The tokenizer is fitted once on the full 16-source training pool. Cov-Hk takes the k highest-ranked sources, and LaBraM and CBraMod reuse the same order even though CBraMod does not use this tokenizer in its pretraining objective. Probe allocation and its dependence on source size are detailed in Appendix E.4.

Cov-H reaches observed peaks of 61.72% for LaBraM Cov-H3 (TUAB, Siena, and SEED-VII) and 58.36% for CBraMod Cov-H6. These pools have higher means than Full16 on 12 and 10 of the 13 tasks, respectively, so their average improvements reflect gains across most targets (Appendix E.4).

The improvements recur across subset sizes: all six tested Cov-H prefixes have higher mean AVG13 than Full16 for both architectures. Expanding the peak configurations to ten sources lowers AVG13 to 60.06% and 56.46%, respectively (Figure 5). Thus, the gains extend beyond a single favorable cutoff, but they do not grow monotonically as more sources are included.

6.5 Practical guidance for source selection

The four methods have different strengths and weaknesses; current evidence does not establish a single best selection strategy. Table 1 reports costs for N candidate sources and ranks against all 61 LaBraM and 112 CBraMod configurations in Figure 2.

Table 1: **Source-scoring cost and observed performance.** Both rank columns use each model's full configuration inventory in Figure 2, including Full16, Scratch, and other controls. Mean best task rank averages each method's best rank among all configurations over 13 tasks; configurations may differ across tasks. Ranks use unrounded three-seed test means and midranks. Sweeps include low-ranked controls (LaBraM/CBraMod: Size/Coverage 12/12 each, Progressive 23/12, PDV 6/7). Costs cover ranking only: $N = 16$ for Progressive and 14 for PDV; the tokenizer is shared.

Method	Ranking cost	LaBraM (61 configurations)			CBraMod (112 configurations)		
		Peak AVG13 (%)	AVG13 rank / all	Mean best task rank ↓	Peak AVG13 (%)	AVG13 rank / all	Mean best task rank ↓
Size	Window counts	60.10 (Size-H10)	13/61	6.73	57.37 (Size-H4)	4/112	11.54
Coverage	1 tokenizer fit + forward passes	61.72 (Cov-H3)	1/61	2.92	58.36 (Cov-H6)	1/112	5.15
Progressive	N models + validation	60.85 (P-H3)	4/61	2.31	57.38 (P-H8)	3/112	13.12
PDV	$N + \binom{N}{2}$ models + validation	61.21 (PDV6)	2/61	2.15	58.18 (PDV4)	2/112	6.62

Size requires corpus counts and no trained scoring model. Coverage needs a fitted tokenizer and probe forward passes, after which its source ranking can be reused across architectures. Both construct rankings without repeated downstream evaluation.

Progressive assesses transfer through single-source pretraining and validation on each architecture. PDV evaluates contributions to partners, increasing scoring models from $O(N)$ to $O(N^2)$.

7 Conclusion

Across the tested EEG architectures and source pool, more pretraining data does not consistently improve benchmark averages or individual-task performance. Substantial differences between similarly sized mixtures show that transfer also depends on which datasets are combined. The capacity experiments reinforce the need to consider composition: further source additions can reverse a decline at fixed Base capacity, and increasing LaBraM to Large narrows but does not eliminate the tested six-source pool's advantage over Full16.

Progressive, PDV, Size, and Coverage identify useful mixtures with different source-scoring requirements and costs. PDV increases the source-scoring model count from linear to quadratic without establishing an additional average benefit over Progressive at matched source counts. Size and Coverage construct rankings without downstream labels and reuse the same order across LaBraM and CBraMod. The evidence does not establish a consistently best strategy. In practice, pretraining datasets for EEG foundation models should be chosen using appropriate data-selection methods.

AI USE STATEMENT

We used generative artificial intelligence (AI) tools for manuscript editing, interpretation of results, figure-generation code, and LaTeX formatting. The authors reviewed the AI-assisted text and interpretations, checked the plotting code and figures against the experimental records, and verified the final manuscript. The authors take full responsibility for the final content of this work, including all text, claims, code, and figures.

ETHICS STATEMENT

We analyze existing, de-identified EEG recordings from 15 public source datasets, the in-house SCD dataset, and six public external evaluation datasets. We recruited no participants and collected no new recordings. SCD was collected for an earlier study with institutional review board (IRB) approval and informed consent; the institution is withheld for anonymity. Source rankings describe performance for the tested models and tasks, not the value of the populations represented. We release no new dataset or pretrained model.

REPRODUCIBILITY STATEMENT

Section 3 describes evaluation. The appendices give preprocessing (Appendix B.1), training settings (Appendix B.2), source ordering (Appendix E), and statistical procedures (Appendix F).

REFERENCES

Amro Abbas, Kushal Tirumala, Dániel Simig, Surya Ganguli, and Ari S Morcos. Semdedup: Data-efficient learning at web-scale through semantic deduplication. *arXiv preprint arXiv:2303.09540*, 2023.

Diego Alvarez-Estevez and Roselyne M Rijsman. Inter-database validation of a deep learning approach for automatic sleep scoring. *PloS one*, 16(8):e0256111, 2021.

Hubert Banville, Yohann Benchetrit, Stéphane d'Ascoli, Jérémy Rapin, and Jean-Rémi King. Scaling laws for decoding images from brain activity. *arXiv preprint arXiv:2501.15322*, 2025.

Donghong Cai, Junru Chen, Yang Yang, Teng Liu, and Yafeng Li. Mbrain: A multi-channel self-supervised learning framework for brain signals. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pp. 130–141, 2023.

Jingjing Chen, Xiaobin Wang, Chen Huang, Xin Hu, Xinke Shen, and Dan Zhang. A large finer-grained affective computing eeg dataset. *Scientific Data*, 10(1):740, 2023.

Yuqi Chen, Kan Ren, Kaitao Song, Yansen Wang, Yifan Wang, Dongsheng Li, and Lili Qiu. Eegformer: Towards transferable and interpretable large-scale eeg foundation model. *arXiv preprint arXiv:2401.10278*, 2024.

Wenhui Cui, Woojae Jeong, Philipp Thölke, Takfarinas Medani, Karim Jerbi, Anand A Joshi, and Richard M Leahy. Neuro-gpt: Towards a foundation model for eeg. In *2024 IEEE International Symposium on Biomedical Imaging (ISBI)*, pp. 1–5. IEEE, 2024.

Paolo Detti, Giampaolo Vatti, and Garazi Zabalo Manrique de Lara. Eeg synchronization analysis for seizure prediction: A study on data of noninvasive recordings. *Processes*, 8(7):846, 2020.

Samir Yitzhak Gadre, Gabriel Ilharco, Alex Fang, Jonathan Hayase, Georgios Smyrnis, Thao Nguyen, Ryan Marten, Mitchell Wortsman, Dhruba Ghosh, Jieyu Zhang, et al. Datacomp: In search of the next generation of multimodal datasets. *Advances in Neural Information Processing Systems*, 36: 27092–27112, 2023.

Amirata Ghorbani and James Zou. Data shapley: Equitable valuation of data for machine learning. In *International conference on machine learning*, pp. 2242–2251. PMLR, 2019.

Aaron Grattafiori, Abhimanyu Dubey, Abhinav Jauhri, Abhinav Pandey, Abhishek Kadian, Ahmad Al-Dahle, Aiesha Letman, Akhil Mathur, Alan Schelten, Alex Vaughan, et al. The llama 3 herd of models. *arXiv preprint arXiv:2407.21783*, 2024.

Amir Harati, Meysam Golmohammadi, Silvia Lopez, Iyad Obeid, and Joseph Picone. Improved eeg event classification using differential energy. In *2015 IEEE Signal Processing in Medicine and Biology Symposium (SPMB)*, pp. 1–4. IEEE, 2015.

Ji-Hoon Jeong, Jeong-Hyun Cho, Kyung-Hwan Shim, Byoung-Hee Kwon, Byeong-Hoo Lee, Do-Yeun Lee, Dae-Hyeok Lee, and Seong-Whan Lee. Multimodal signal dataset for 11 intuitive movement tasks from single upper extremity during multiple recording sessions. *GigaScience*, 9 (10):giaa098, 2020.

Linxing Preston Jiang, Shirui Chen, Emmanuel Tanumihardja, Xiaochuang Han, Weijia Shi, Eric Shea-Brown, and Rajesh PN Rao. Data heterogeneity limits the scaling effect of pretraining neural data transformers. *bioRxiv*, pp. 2025–05, 2025.

Wei-Bang Jiang, Xuan-Hao Liu, Wei-Long Zheng, and Bao-Liang Lu. Seed-vii: A multimodal dataset of six basic emotions with continuous labels for emotion recognition. *IEEE Transactions on Affective Computing*, 16(2):969–985, 2024a.

Wei-Bang Jiang, Li-Ming Zhao, and Bao-Liang Lu. Large brain model for learning generic representations with tremendous eeg data in bci. *arXiv preprint arXiv:2405.18765*, 2024b.

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

Yongchan Kwon and James Zou. Beta shapley: a unified and noise-reduced data valuation framework for machine learning. *arXiv preprint arXiv:2110.14049*, 2021.

Jeet Bandhu Lahiri, Parshva Runwal, Arvasu Kulkarni, Mahir Jain, Aditya Ray Mishra, Siddharth Panwar, and Sandeep Singh. Prism: Exploring heterogeneous pretrained eeg foundation model transfer to clinical differential diagnosis. *arXiv preprint arXiv:2603.02268*, 2026.

Wei Liu, Jie-Lin Qiu, Wei-Long Zheng, and Bao-Liang Lu. Comparing recognition performance and robustness of multimodal deep learning models for multimodal emotion recognition. *IEEE Transactions on Cognitive and Developmental Systems*, (2):715–729, 2021.

Jun Ma, Banghua Yang, Wenzheng Qiu, Yunzhe Li, Shouwei Gao, and Xinxing Xia. A large eeg dataset for studying cross-session variability in motor imagery brain-computer interface. *Scientific Data*, 9(1):531, 2022.

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

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. *arXiv preprint arXiv:2510.21585*, 2025.

Gerwin Schalk, Dennis J McFarland, Thilo Hinterberger, Niels Birbaumer, and Jonathan R Wolpaw. Bci2000: a general-purpose brain-computer interface (bci) system. *IEEE Transactions on biomedical engineering*, 51(6):1034–1043, 2004.

Michael Tangermann, Klaus-Robert Müller, Ad Aertsen, Niels Birbaumer, Christoph Braun, Clemens Brunner, Robert Leeb, Carsten Mehring, Kai J Miller, Gernot R Müller-Putz, et al. Review of the bci competition iv. *Frontiers in neuroscience*, 6:55, 2012.

Kushal Tirumala, Daniel Simig, Armen Aghajanyan, and Ari Morcos. D4: Improving llm pretraining via document de-duplication and diversification. *Advances in Neural Information Processing Systems*, 36:53983–53995, 2023.

Eva von Weltin, Tameem Ahsan, Vinit Shah, Dawer Jamshed, Meysam Golmohammadi, Iyad Obeid, and Joseph Picone. Electroencephalographic slowing: A primary source of error in automatic seizure detection. In *2017 IEEE signal processing in medicine and biology symposium (SPMB)*, pp. 1–5. IEEE, 2017.

Christopher Wang, Vighnesh Subramaniam, Adam Uri Yaari, Gabriel Kreiman, Boris Katz, Ignacio Cases, and Andrei Barbu. Brainbert: Self-supervised representation learning for intracranial recordings. *arXiv preprint arXiv:2302.14367*, 2023.

Guagnyu 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, 2024a.

Jiachen T Wang and Ruoxi Jia. Data banzhaf: A robust data valuation framework for machine learning. In *International conference on artificial intelligence and statistics*, pp. 6388–6421. PMLR, 2023.

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. *arXiv preprint arXiv:2412.07236*, 2024b.

Konstantin F Willeke, Polina Turishcheva, Alex Gilbert, Goirik Chakrabarty, Hasan Bedel, Paul Fahey, Yongrong Qiu, Marissa Weis, Michaela Vystrčilová, Taliah Muhammad, et al. Omnimouse: Scaling properties of multi-modal, multi-task brain models on 150b neural tokens. In *International Conference on Learning Representations*, volume 2026, pp. 134382–134411, 2026.

Glenn F Wilson and Christopher A Russell. Real-time assessment of mental workload using psychophysiological measures and artificial neural networks. *Human factors*, 45(4):635–644, 2003.

Jiamin Wu, Zichen Ren, Junyu Wang, Pengyu Zhu, Yonghao Song, Mianxin Liu, Qihao Zheng, Lei Bai, Wanli Ouyang, and Chunfeng Song. Adabrain-bench: Benchmarking brain foundation models for brain-computer interface applications. *arXiv preprint arXiv:2507.09882*, 2025.

Wei Xiong, Jiangtong Li, Jie Li, Kun Zhu, and Changjun Jiang. Eeg-fm-bench: A comprehensive benchmark for the systematic evaluation of eeg foundation models. *arXiv preprint arXiv:2508.17742*, 2025.

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

Zhizhang Yuan, Fanqi Shen, Meng Li, Yuguo Yu, Chenhao Tan, and Yang Yang. Brainwave: A brain signal foundation model for clinical applications. *arXiv preprint arXiv:2402.10251*, 2024.

Daoze Zhang, Zhizhang Yuan, Yang Yang, Junru Chen, Jingjing Wang, and Yafeng Li. Brant: Foundation model for intracranial neural signal. *Advances in Neural Information Processing Systems*, 36:26304–26321, 2023.

Wei-Long Zheng and Bao-Liang Lu. Investigating critical frequency bands and channels for eeg-based emotion recognition with deep neural networks. *IEEE Transactions on autonomous mental development*, 7(3):162–175, 2015.

Wei-Long Zheng, Wei Liu, Yifei Lu, Bao-Liang Lu, and Andrzej Cichocki. Emotionmeter: A multimodal framework for recognizing human emotions. *IEEE transactions on cybernetics*, 49(3): 1110–1122, 2018.

A LIMITATIONS

This paper is a systematic audit of pretraining data scale and composition for EEG foundation models. We cannot yet precisely classify datasets as "good" or "bad" pretraining sources. A source can have different effects depending on the other sources in the pool, the model, and the downstream task. The rankings studied here summarize performance under particular evaluation settings or measure source properties. They do not provide a general prediction of whether adding a dataset will help, or by how much.

We also cannot identify a globally optimal dataset combination. The 202 evaluated configurations across three architectures sample only part of the possible combinations from the 16-source candidate pool. The best observed configuration is the best among those tested for that architecture; an untested mixture may perform better. A pool that maximizes AVG13 may also differ from one that maximizes performance on a particular task. Our results establish the importance of composition without locating an optimum over the full search space.

The comparison of Progressive, PDV, Size, and Coverage likewise does not establish an optimal selection method. The four strategies offer different trade-offs between scoring cost and downstream performance, and their relative performance varies with architecture and subset size. They provide practical reference points for dataset selection, but the evidence is insufficient to identify a strategy that is best across source pools, models, and tasks. The contribution of this paper is the systematic empirical audit; precise source valuation and optimal dataset selection remain open problems.

B DATASETS AND EXPERIMENTAL DETAILS

B.1 DATASETS AND PREPROCESSING

The candidate pool contains 15 public datasets and the in-house SCD dataset. It covers clinical EEG—TUAB (Obeid & Picone, 2016), TUEV (Harati et al., 2015), TUSL (von Weltin et al., 2017), and Siena (Detti et al., 2020); sleep staging—HMC (Alvarez-Estevez & Rijsman, 2021); motor imagery—PhysioNet (Schalk et al., 2004), BCIC-2020/3 (Jeong et al., 2020), SHU-MI (Ma et al., 2022), and BCIC-IV-2a (Tangermann et al., 2012); emotion recognition—SEED (Zheng & Lu, 2015), SEED-IV (Zheng et al., 2018), SEED-V (Liu et al., 2021), SEED-VII (Jiang et al., 2024a), and FACED (Chen et al., 2023); and cognitive tasks—Workload (Wilson & Russell, 2003) and SCD. BCIC20-3 abbreviates BCIC-2020/3 in tables. TUSL evaluates clinical EEG slowing. Table 2 combines the task definitions and training-source metadata.

Recordings undergo a 0.1–75 Hz bandpass filter, a 50 or 60 Hz notch filter according to the collection site's powerline frequency, and resampling to 200 Hz. Amplitude normalization uses z-scores. Each source retains its own window length, channel set, and fixed train/validation/test partitions. Only training partitions supply pretraining data; downstream validation partitions are used for model selection and test partitions for reporting.

Table 2: Dataset tasks and staged training-source metadata. Windows are training segments per epoch, at the listed source-specific lengths; they sum to 637,814. Channels are those used in evaluation. These counts are distinct from the whole-corpus counts used by Size and from Coverage's scoring probes.

Dataset	Task	Classes	Channels	Window (s)	Subjects	Train windows
TUAB	Abnormal detection	2	21	10	1668	298,468
HMC	Sleep staging	5	4	16	100	91,248
TUEV	Event detection	6	21	5	231	68,656
Siena	Seizure detection	2	19	10	8	33,605
SEED-VII	Emotion recognition	7	62	4	19	38,931
SCD	Cognitive task	2	62	4	19	24,964
SEED-IV	Emotion recognition	4	62	4	14	16,352
SEED	Emotion recognition	3	62	4	14	13,972
SEED-V	Emotion recognition	5	62	4	15	12,135
PhysioNet	Motor imagery	4	64	3	84	14,616
SHU-MI	Motor imagery	2	16	4	18	8,683
FACED	Emotion recognition	9	32	10	98	8,232
BCIC-2020/3	Motor imagery	2	64	4	15	4,500
BCIC-IV-2a	Motor imagery	4	22	3	9	1,796
Workload	Cognitive load	2	19	4	26	1,508
TUSL	Slowing	3	21	10	19	148
Full16						637,814

B.2 Models and training

LaBraM Base (Jiang et al., 2024b) uses a 12-layer vision Transformer with dimension 200 and ten attention heads. Its vector-quantized tokenizer has 8,192 codes and is fitted once on the 16-source training pool before being fixed. Pretraining predicts tokenizer targets under 50% random patch masking. CBraMod (Wang et al., 2024b) uses a 12-layer, 200-dimensional encoder with eight heads and criss-cross attention, trained with mean squared reconstruction error and 50% random channel masking. BIOT (Yang et al., 2023) uses a four-layer, 256-dimensional encoder with eight heads and linear attention, trained with InfoNCE on augmented temporal segments. CBraMod and BIOT do not use the LaBraM tokenizer in their objectives.

Table 3: Shared training settings. The effective batch size includes all devices and gradient accumulation. Downstream fine-tuning uses a cosine learning-rate schedule, three warmup epochs, weight decay 0.05, and validation balanced accuracy for checkpoint selection.

Setting	Pretraining	Fine-tuning
Optimizer	AdamW	AdamW
Learning rate	5×10^{-4}	10^{-4}
Epochs	50	30
Effective batch size	64	64

All downstream experiments use full fine-tuning, with the same head design within each architecture and seeds $\{0, 1, 2\}$. These seeds repeat downstream training from the same pretrained model. For the fixed six-source and Full16 capacity comparisons, LaBraM Base and Large share preprocessing, objective, training settings, data partitions, fine-tuning, and model selection; capacity is the changed factor. Experiments run on NVIDIA V100 (32GB) and A100 (40GB/80GB) GPUs through Singularity clusters. A fine-tuning run uses one GPU and takes approximately 10–480 minutes depending on the target.

B.3 Evaluation, configuration inventory, and data exposure

Full16 concatenates all 16 sources; Scratch uses no pretraining. AVG13 first averages the three fine-tuning seeds within each target, then gives equal weight to the 13 target means. SEED-V, SCD, and BCIC-IV-2a remain pretraining candidates but are excluded from AVG13 because their downstream evaluation remains near chance (Appendix I). Progressive source ranking uses validation performance over AVG13 plus BCIC-IV-2a; PDV uses validation gains on the other sources' receiving tasks. Appendix E gives the ranking definitions and scores. Subset-size validation records are unavailable, so the reported sweep maxima are observed test results rather than verified validation-selected estimates.

The overview contains 202 complete architecture–configuration records: 61 LaBraM, 112 CBraMod, and 29 BIOT configurations. Each has 13 targets and three fine-tuning seeds. Full16, Scratch, and the recorded weighting, filtering, curation, and source-removal controls remain in the inventory. LaBraM P-H3 and PDV3 reuse the same checkpoint and evaluations and are counted once; some other configurations also reuse checkpoints. Different configuration records therefore need not be independent pretraining repetitions. The complete task results and configuration rankings appear in Appendices C and H.

A training window is one staged EEG segment at its source's window length. Counts sum to 637,814 windows per Full16 epoch, or 31,890,700 nominal visits over 50 epochs. Counts measure exposure rather than end-to-end computation because window durations and channel counts differ. For ordinary source subsets, exposure is the sum of their sources' training-window counts. Filtering controls and reused pair runs use the candidate-pool counts recorded in the overview when retained-window counts are unavailable. Coverage probe-allocation counts and Size's whole-corpus ranking counts are separate from staged training exposure.

Table 4: **Training-window exposure of the highest observed Coverage and PDV configurations.** All six runs use 50 epochs and effective global batch size 64. Full16 has 637,814 windows. Visits count nominal repetitions; Full16 shares also approximate update-count ratios. The highlighted configurations are observed test maxima.

Architecture	Mixture	Windows/epoch	Visits (50 epochs)	Full16 share	AVG13	Δ (pp)
LaBraM	Cov-H3	371,004	18,550,200	58.2%	61.72	+5.58
LaBraM	PDV6	397,384	19,869,200	62.3%	61.21	+5.06
LaBraM	Full16	637,814	31,890,700	100.0%	56.15	+0.00
CBraMod	Cov-H6	493,220	24,661,000	77.3%	58.36	+3.30
CBraMod	PDV4	472,988	23,649,400	74.2%	58.18	+3.12
CBraMod	Full16	637,814	31,890,700	100.0%	55.06	+0.00

The six configurations in Table 4 share 50 pretraining epochs and effective batch size 64. Their window-count ratios therefore also approximate update-count ratios, keeping the selected Coverage/PDV comparisons within 1.72-fold of Full16. The selected pools use fewer windows and approximate updates; these comparisons control the order of magnitude rather than equalizing exact update counts. Ranking-model training and tokenizer fitting have separate costs.

C Detailed Results by Main-Text Section

This appendix follows the experiments in Sections 4–6. It reports task results for every concrete comparison and all source-count sweeps, followed by the remaining configurations in the complete overview. Together the tables cover all 202 configurations in Figure 2, plus the two LaBraM Large configurations in Section 5.2.

Task entries are test balanced accuracy (%), reported as mean ± sample standard deviation over fine-tuning seeds $\{0, 1, 2\}$. AVG13 equally weights the 13 task means. The two panels of each table together contain all 13 targets. Repeated configurations retain identical scores. LaBraM P-H3 and PDV3 share one checkpoint and appear under a joint name. The three seeds repeat fine-tuning, not pretraining; table entries and observed maxima do not denote significance.

C.1 Data scale and source composition: Section 4

Table 5 links the source-set comparisons in the text to their run names. Tables 6–8 give their full task profiles, together with Full16, Scratch, and each architecture's highest observed AVG13 configuration. The expansion paths are reported in Appendix C.3; the remaining points in the overview are in Appendix C.4.

Table 5: Run names for the concrete comparisons in Section 4. Arrows identify the compared source pools; each pool is pretrained separately. Full memberships are stated in the main text and the ranking recipes in Appendix E.

Section	Architecture	Compared configurations	Windows per epoch
4.1	LaBraM	Cov-H3 → Cov-H4	371,004 → 387,356
4.1	CBraMod	PDV4 → PDV5	472,988 → 506,593
4.2	LaBraM	Cov-H2 → Cov-H3 → Cov-H4	332,073 → 371,004 → 387,356
4.2	LaBraM	P-H7 → P-H8	424,176 → 432,859
4.2	BIOT	Size-L3 → Size-L4	3,452 → 7,952
4.3	LaBraM	Random-3 #7 / Cov-L6	53,552 / 48,783
4.3	CBraMod	Random-5 #0 / P-H4	439,126 / 437,937
4.3	BIOT	Size-H2 / P-L2	389,716 / 367,124

Table 6: LaBraM: full task results for the comparisons in Section 4.

Configuration	TUAB	TUEV	TUSL	Siena	HMC	SEED	SEED-IV
Full16	81.61±0.16	57.57±0.71	58.73±5.50	56.52±0.56	71.23±0.23	64.82±0.19	67.89±0.44
Scratch	77.23±0.32	41.31±1.30	41.56±7.55	55.72±5.16	60.56±0.67	47.04±1.98	41.08±1.65
Cov-H2	80.84±0.75	59.52±0.41	68.51±3.19	69.32±6.02	71.66±0.92	60.67±0.33	68.22±0.36
Cov-H3	82.45±0.12	59.94±0.46	69.96±2.94	69.57±7.07	71.09±0.44	69.13±0.37	75.79±0.21
Cov-H4	81.69±0.28	65.02±1.49	69.04±4.11	69.86±1.99	72.03±0.51	69.21±0.13	71.76±0.25
P-H7	82.74±0.13	58.29±0.81	66.67±0.00	69.80±2.34	72.36±0.69	70.20±0.01	73.52±0.15
P-H8	82.20±0.15	62.55±1.61	63.49±2.75	62.71±2.41	72.69±1.30	70.80±0.30	71.42±0.16
Random-3 #7	77.81±0.21	44.06±1.42	44.59±2.41	61.36±3.29	62.88±0.42	54.05±1.71	67.04±0.36
Cov-L6	80.69±0.25	55.12±1.31	64.95±1.45	62.52±4.20	68.12±0.37	60.72±0.21	63.99±0.55

Configuration	SEED-VII	FACED	PhysioNet	BCIC20-3	SHU-MI	Workload	AVG13
Full16	34.97±0.34	22.85±0.48	42.72±0.70	49.79±0.36	58.12±1.30	63.10±1.79	56.15
Scratch	20.32±0.39	14.92±0.63	28.48±0.68	50.00±0.00	50.78±2.18	53.45±2.49	44.80
Cov-H2	33.48±0.09	25.82±1.39	39.26±2.08	50.10±0.95	60.86±0.67	67.36±1.59	58.12
Cov-H3	38.72±0.70	33.06±1.64	50.43±0.24	52.08±1.83	61.57±1.05	68.62±4.07	61.72
Cov-H4	39.39±0.46	25.48±1.03	46.36±0.54	50.00±0.00	62.23±0.91	68.85±0.87	60.84
P-H7	39.20±0.18	26.12±0.35	49.68±0.86	50.00±0.00	61.58±0.44	63.91±3.09	60.31
P-H8	38.76±0.25	24.17±0.78	46.89±1.09	49.64±1.33	61.15±1.47	67.47±1.39	59.53
Random-3 #7	21.51±1.00	15.62±1.13	30.80±0.51	50.00±0.00	50.51±1.17	54.71±0.40	48.84
Cov-L6	30.39±0.17	27.81±0.61	37.87±0.36	53.59±1.18	62.48±0.20	64.71±1.70	56.38

Table 7: CBraMod: full task results for the comparisons in Section 4.

Configuration	TUAB	TUEV	TUSL	Siena	HMC	SEED	SEED-IV
Full16	81.61±0.65	50.56±1.63	70.61±3.91	64.17±0.35	66.01±0.55	47.94±1.12	67.80±0.27
Scratch	78.60±0.68	47.97±0.42	57.53±0.24	64.41±0.90	68.04±0.41	53.35±0.50	55.55±1.28
PDV4	80.29±0.22	55.54±0.57	77.03±7.43	64.76±1.48	67.87±0.68	59.40±0.34	68.94±0.35
PDV5	81.98±0.45	52.19±2.28	76.28±3.03	64.78±1.77	65.63±0.64	52.72±1.28	68.08±0.34
Cov-H6	80.65±0.62	55.77±1.46	84.37±1.75	65.61±2.56	68.62±1.13	60.40±0.23	69.60±0.11
Random-5 #0	80.42±0.68	51.37±1.49	78.79±5.76	61.31±4.13	66.27±0.55	51.24±2.15	66.22±0.04
P-H4	81.46±0.40	55.30±0.67	75.23±6.36	66.21±1.65	67.84±0.36	54.30±0.88	66.37±0.65

Configuration	SEED-VII	FACED	PhysioNet	BCIC20-3	SHU-MI	Workload	AVG13
Full16	24.22±0.04	37.55±1.90	36.92±1.88	52.76±0.33	61.61±0.97	54.02±1.11	55.06
Scratch	22.91±0.61	26.32±1.83	28.31±1.39	50.05±1.73	51.22±0.90	53.22±2.77	50.57
PDV4	24.90±0.98	46.39±1.70	42.78±1.08	51.56±3.02	61.38±3.76	55.52±1.03	58.18
PDV5	23.35±0.45	37.93±1.38	36.69±1.55	53.12±0.87	61.09±4.31	54.60±0.72	56.04
Cov-H6	25.98±0.56	36.61±0.59	39.80±1.00	51.20±1.00	61.80±1.71	58.28±1.50	58.36
Random-5 #0	22.27±1.11	36.85±1.11	38.69±2.21	51.82±2.67	60.74±2.18	59.89±3.49	55.84
P-H4	23.89±0.22	37.58±1.38	38.27±1.28	52.86±2.84	60.97±0.90	59.31±4.07	56.89

Table 8: BIOT: full task results for the comparisons in Section 4.

Configuration	TUAB	TUEV	TUSL	Siena	HMC	SEED	SEED-IV
Full16	77.29±0.69	45.11±4.39	40.66±5.70	67.97±1.71	68.58±0.52	64.76±0.33	54.51±0.49
Scratch	77.79±0.10	46.31±1.97	72.24±6.12	64.35±4.24	68.12±0.22	68.09±0.68	51.01±1.21
P-H3	78.52±0.11	45.34±3.12	71.43±14.64	64.13±4.89	69.09±0.12	70.04±0.26	65.30±0.16
Size-L3	78.50±0.45	41.40±1.39	80.68±5.29	61.66±5.62	68.32±0.33	67.43±0.28	53.70±0.83
Size-L4	77.13±0.39	43.62±2.71	59.65±6.38	69.92±3.41	69.50±0.82	67.88±0.33	54.13±0.59
Size-H2	75.57±1.40	39.23±5.16	41.48±5.99	65.97±0.98	69.42±0.27	62.60±0.39	44.41±0.74
P-L2	77.22±0.36	43.52±2.62	54.42±11.51	71.74±5.60	69.21±0.28	66.83±0.20	53.51±1.03

Configuration	SEED-VII	FACED	PhysioNet	BCIC20-3	SHU-MI	Workload	AVG13
Full16	34.48±1.41	15.65±1.20	22.93±1.14	50.57±0.72	49.55±0.54	63.79±1.50	50.45
Scratch	36.35±0.39	20.29±0.88	26.58±0.49	49.48±0.90	59.10±0.13	71.84±0.53	54.74
P-H3	35.78±0.72	19.90±1.12	26.23±1.18	50.94±0.27	58.76±1.22	72.30±1.31	55.98
Size-L3	35.98±0.84	19.79±0.71	25.90±0.41	48.49±1.49	58.36±0.50	62.87±2.49	54.08
Size-L4	34.40±1.34	17.24±0.58	22.80±0.37	50.05±0.09	55.52±0.68	55.75±2.02	52.12
Size-H2	31.32±1.19	16.10±0.47	23.66±0.24	51.88±1.96	50.16±0.68	66.90±2.82	49.13
P-L2	35.22±0.12	16.63±0.91	24.05±0.77	50.21±0.90	52.21±1.13	70.11±1.90	52.68

C.2 MODEL CAPACITY: SECTION 5

The Base expansion starts from $A = \{\text{SCD}, \text{SEED-V}, \text{BCIC-IV-2a}, \text{Workload}\}$, adds $B = \{\text{TUSL}\}$, and then $C = \{\text{FACED}\}$ while retaining every previous source. These pools correspond to Cov-L4, P-L5, and Cov-L6. Table 9 gives their task scores; Appendix D.3 gives further nested examples.

Table 9: LaBraM Base: the three nested pools in Section 5.1.

Configuration	TUAB	TUEV	TUSL	Siena	HMC	SEED	SEED-IV
A (Cov-L4)	77.43±0.19	40.73±0.47	62.01±8.07	56.00±1.37	62.52±0.57	46.18±0.37	55.25±0.25
A+B (P-L5)	77.57±0.11	32.15±0.21	62.17±3.67	50.18±0.30	54.89±0.10	40.22±0.05	32.27±0.25
A+B+C (Cov-L6)	80.69±0.25	55.12±1.31	64.95±1.45	62.52±4.20	68.12±0.37	60.72±0.21	63.99±0.55

Configuration	SEED-VII	FACED	PhysioNet	BCIC20-3	SHU-MI	Workload	AVG13
A (Cov-L4)	22.32±0.25	17.99±0.43	30.45±0.81	50.31±0.98	55.62±0.74	57.59±0.91	48.80
A+B (P-L5)	18.40±0.13	13.95±1.01	25.23±0.77	49.22±0.00	49.99±0.24	52.30±0.87	42.96
A+B+C (Cov-L6)	30.39±0.17	27.81±0.61	37.87±0.36	53.59±1.18	62.48±0.20	64.71±1.70	56.38

The capacity comparison fixes either Full16 or the six-source mixture TUAB, Siena, SEED-VII, SEED-IV, BCIC-IV-2a, and FACED (PDV6). Base and Large use the same preprocessing, objective, 50 pretraining epochs, effective global batch size 64, downstream partitions, and fine-tuning and model-selection procedures. Table 10 provides the four matched task profiles. The Large subset–Full16 difference is 2.84 points, with higher means on 11 of 13 tasks; Table 39 gives its paired test and interval.

Table 10: LaBraM Base and Large under the two fixed recipes in Section 5.2.

Configuration	TUAB	TUEV	TUSL	Siena	HMC	SEED	SEED-IV
Base Full16	81.61±0.16	57.57±0.71	58.73±5.50	56.52±0.56	71.23±0.23	64.82±0.19	67.89±0.44
Base six-source	83.49±0.10	58.07±1.45	61.90±4.76	74.18±2.22	71.74±0.90	69.86±0.14	72.95±0.27
Large Full16	82.17±0.04	63.84±0.27	68.12±0.00	65.47±2.51	72.03±0.62	67.92±0.25	74.51±0.34
Large six-source	81.87±0.05	65.61±1.58	75.27±8.40	75.84±2.71	72.00±0.36	71.80±0.08	74.77±0.20

Configuration	SEED-VII	FACED	PhysioNet	BCIC20-3	SHU-MI	Workload	AVG13
Base Full16	34.97±0.34	22.85±0.48	42.72±0.70	49.79±0.36	58.12±1.30	63.10±1.79	56.15
Base six-source	39.38±0.79	29.72±2.37	47.18±0.82	52.29±0.79	61.54±0.60	73.45±1.72	61.21
Large Full16	37.06±0.36	30.14±1.57	46.10±0.87	51.72±2.25	61.10±1.02	63.22±3.35	60.26
Large six-source	37.52±0.27	34.65±0.13	49.78±1.80	53.80±1.27	62.49±1.30	64.94±2.93	63.10

C.3 SOURCE-SELECTION SWEEPS: SECTION 6

The following tables cover every evaluated Progressive, PDV, Size, and Coverage prefix, including the low-ranked controls used in the inventory and the rank calculations in Table 1. Rows are grouped by method and source count within each architecture. They also provide the full task results underlying the expansion curves in Figures 1 and 5. Source scores and memberships appear in Appendix E. BIOT has Progressive and Size sweeps only.

Table 11: LaBraM: complete source-selection task results, part 1 of 2. Rows run from P-H1 to PDV6; all available source counts are retained.

Configuration	TUAB	TUEV	TUSL	Siena	HMC	SEED	SEED-IV
P-H1	83.73±0.18	61.81±1.12	63.49±2.75	61.50±2.48	72.63±0.21	61.66±0.28	64.97±0.36
P-H2	83.41±0.33	58.11±1.71	65.08±2.75	76.93±1.84	72.19±0.63	62.07±0.13	66.01±0.36
P-H3 / PDV3	82.88±0.13	65.07±1.63	64.07±2.41	70.55±1.07	72.06±0.40	69.49±0.10	71.86±0.57
P-H4	82.55±0.01	60.61±2.44	65.08±2.75	66.67±0.55	72.07±0.26	69.34±0.20	73.23±0.11
P-H5	82.59±0.46	64.18±0.44	66.67±0.00	69.60±7.40	72.40±0.32	70.43±0.15	72.70±0.40
P-H6	82.31±0.46	60.23±1.46	66.53±4.56	64.24±5.23	71.63±0.14	69.41±0.33	72.12±0.09
P-H7	82.74±0.13	58.29±0.81	66.67±0.00	69.80±2.34	72.36±0.69	70.20±0.01	73.52±0.15
P-H8	82.20±0.15	62.55±1.61	63.49±2.75	62.71±2.41	72.69±1.30	70.80±0.30	71.42±0.16
P-H9	82.25±0.05	60.24±2.69	66.67±0.00	64.38±3.00	72.84±1.11	69.47±0.27	71.28±0.70
P-H10	82.89±0.15	63.84±1.19	69.08±4.18	59.16±5.00	73.65±1.04	70.20±0.03	73.49±0.34
P-H12	82.76±0.08	62.55±0.25	63.49±2.75	66.46±1.21	73.62±0.52	69.15±0.04	71.08±0.29
P-H14	82.16±0.19	62.81±0.58	66.67±0.00	64.00±4.39	74.46±0.34	68.71±0.13	71.57±0.41
P-L1	76.48±0.05	38.63±2.49	41.56±0.00	55.53±3.94	60.96±0.44	45.42±0.20	45.02±0.37
P-L2	77.40±0.13	37.20±0.78	48.48±3.03	59.36±0.96	59.01±0.53	42.58±0.08	54.69±0.60
P-L3	77.09±0.07	36.79±0.80	57.14±0.00	55.43±0.51	61.32±0.02	45.14±0.08	51.63±0.34
P-L4	77.09±0.15	41.23±1.84	49.78±2.41	59.28±1.33	63.86±0.16	48.10±0.15	63.23±0.12
P-L5	77.57±0.11	32.15±0.21	62.17±3.67	50.18±0.30	54.89±0.10	40.22±0.05	32.27±0.25
P-L6	79.10±0.19	50.61±1.09	52.67±5.50	59.91±0.98	69.26±0.84	58.60±0.19	66.64±0.14
P-L7	77.31±0.10	44.30±1.70	49.06±1.75	62.09±2.07	63.62±0.17	53.29±0.08	63.78±0.28
P-L8	79.43±0.27	45.66±1.74	55.84±0.00	57.19±0.33	69.42±0.50	59.45±0.17	66.97±0.38
P-L9	79.42±0.22	52.30±0.82	58.87±0.00	57.27±0.21	71.13±0.22	61.36±0.12	66.56±0.37
P-L10	82.05±0.12	55.58±1.03	66.67±0.00	61.78±1.79	73.02±0.14	66.35±0.16	66.25±0.34
P-L12	81.18±0.05	57.94±1.13	60.46±2.75	57.81±1.23	71.71±0.23	62.61±0.16	69.19±0.42
PDV2	82.35±0.21	58.62±2.35	72.95±1.67	67.13±5.75	72.42±0.22	69.69±0.43	72.23±0.20
PDV4	82.12±0.73	62.98±0.44	65.08±2.75	76.24±0.83	72.15±0.48	69.25±0.06	74.27±0.40
PDV6	83.49±0.10	58.07±1.45	61.90±4.76	74.18±2.22	71.74±0.90	69.86±0.14	72.95±0.27

Configuration	SEED-VII	FACED	PhysioNet	BCIC20-3	SHU-MI	Workload	AVG13
P-H1	35.02±0.12	27.71±0.84	41.51±4.03	49.32±1.13	62.27±0.24	65.75±1.05	57.80
P-H2	34.21±0.71	25.44±0.55	36.67±0.16	49.53±0.27	60.62±1.96	65.52±0.34	58.14
P-H3 / PDV3	38.41±0.80	25.63±0.31	49.36±1.27	54.01±1.26	61.28±1.05	66.32±4.11	60.85
P-H4	39.01±0.38	24.59±1.36	47.68±1.15	52.19±2.46	62.92±0.69	68.39±1.31	60.33
P-H5	38.91±0.56	27.08±0.98	44.64±1.93	50.00±0.78	59.58±1.10	64.48±2.99	60.25
P-H6	38.38±0.77	25.45±0.64	49.31±0.99	51.15±1.19	60.81±0.04	61.84±0.53	59.49
P-H7	39.20±0.18	26.12±0.35	49.68±0.86	50.00±0.00	61.58±0.44	63.91±3.09	60.31
P-H8	38.76±0.25	24.17±0.78	46.89±1.09	49.64±1.33	61.15±1.47	67.47±1.39	59.53
P-H9	39.65±0.93	24.51±0.47	48.13±1.30	51.77±1.72	61.15±0.81	65.75±1.39	59.85
P-H10	38.89±0.91	25.88±1.17	50.44±1.54	49.84±0.27	61.17±1.13	68.05±1.70	60.51
P-H12	36.92±0.44	24.31±0.35	47.16±1.46	49.48±0.77	59.81±0.04	68.74±2.49	59.66
P-H14	38.25±0.57	20.77±0.75	40.14±0.90	49.06±0.47	60.56±1.27	71.03±2.39	59.25
P-L1	20.06±0.11	12.72±0.56	27.03±0.65	50.21±0.18	50.45±0.09	52.99±1.05	44.39
P-L2	19.80±0.10	13.16±0.71	26.66±0.10	50.00±0.00	49.72±0.56	50.23±1.11	45.25
P-L3	19.45±0.09	13.72±0.02	26.03±0.28	50.00±0.00	49.94±0.11	55.63±1.90	46.10
P-L4	23.00±0.22	13.58±0.31	28.80±0.35	49.43±0.90	52.26±0.54	54.83±0.60	48.04
P-L5	18.40±0.13	13.95±1.01	25.23±0.77	49.22±0.00	49.99±0.24	52.30±0.87	42.96
P-L6	27.38±0.10	20.43±0.74	32.17±0.02	50.00±0.95	56.37±0.95	57.59±0.60	52.36
P-L7	22.46±0.11	16.87±0.48	29.90±0.39	48.02±0.59	54.09±0.50	52.87±1.44	49.05
P-L8	26.04±0.08	20.69±0.25	30.84±0.83	49.79±0.86	55.62±0.89	57.82±0.40	51.91
P-L9	29.55±0.32	21.46±1.41	31.22±0.57	48.70±0.33	59.03±0.36	56.55±0.69	53.34
P-L10	35.27±0.35	27.38±0.72	36.90±1.27	50.31±0.31	59.22±1.21	59.77±2.77	56.97
P-L12	30.64±0.28	23.33±0.42	36.51±1.27	49.69±1.13	60.37±0.91	56.67±1.11	55.24
PDV2	39.76±0.22	27.24±0.26	47.51±2.55	50.83±0.86	61.03±0.67	65.63±0.53	60.57
PDV4	39.40±0.79	27.09±0.90	46.41±0.94	50.00±0.00	62.97±1.16	65.86±0.00	61.06
PDV6	39.38±0.79	29.72±2.37	47.18±0.82	52.29±0.79	61.54±0.60	73.45±1.72	61.21

Table 12: LaBraM: complete source-selection task results, part 2 of 2. Rows run from PDV8 to Cov-L10; all available source counts are retained.

Configuration	TUAB	TUEV	TUSL	Siena	HMC	SEED	SEED-IV
PDV8	83.12±0.23	62.50±0.54	66.67±0.00	72.83±1.76	72.14±0.70	70.29±0.37	73.00±0.29
PDV10	82.48±0.10	62.65±1.35	65.66±1.75	72.60±1.84	74.19±0.25	69.57±0.29	73.74±0.68
Size-H2	82.71±0.69	60.70±0.63	66.67±0.00	60.70±2.15	72.55±0.84	58.77±0.27	65.71±0.30
Size-H3	82.35±0.27	58.33±2.18	66.67±0.00	63.35±2.96	72.60±0.49	57.57±0.17	66.49±0.21
Size-H4	82.30±0.79	59.28±2.77	66.67±0.00	60.36±0.85	73.44±1.01	60.38±0.20	67.93±0.19
Size-H6	82.63±0.23	57.85±1.01	64.65±1.75	60.76±6.29	74.02±0.21	66.06±0.46	70.90±0.40
Size-H8	83.03±0.06	57.55±0.61	66.67±0.00	63.78±6.64	73.89±0.35	69.31±0.54	73.45±0.36
Size-H10	82.30±0.48	58.62±0.34	64.65±1.75	67.52±6.72	73.20±0.87	69.18±0.22	74.44±0.11
Size-L2	76.27±0.08	39.16±1.06	46.46±3.50	59.58±2.68	60.60±0.17	44.90±0.09	50.93±0.26
Size-L3	76.28±0.11	41.63±0.35	45.74±1.75	56.19±3.56	61.76±0.37	46.29±0.07	48.90±0.85
Size-L4	78.77±0.10	44.43±0.89	56.81±0.84	60.96±1.09	65.30±0.19	51.96±0.33	64.52±0.44
Size-L6	81.43±0.19	55.71±0.83	59.77±1.75	61.83±1.52	69.28±0.17	62.67±0.10	68.17±0.29
Size-L8	81.28±0.27	58.56±1.53	60.61±0.00	56.19±2.19	70.61±0.20	64.88±0.21	67.72±0.22
Size-L10	80.55±0.06	65.91±1.30	57.86±10.00	54.41±1.10	70.34±0.45	67.48±0.29	71.21±0.31
Cov-H2	80.84±0.75	59.52±0.41	68.51±3.19	69.32±6.02	71.66±0.92	60.67±0.33	68.22±0.36
Cov-H3	82.45±0.12	59.94±0.46	69.96±2.94	69.57±7.07	71.09±0.44	69.13±0.37	75.79±0.21
Cov-H4	81.69±0.28	65.02±1.49	69.04±4.11	69.86±1.99	72.03±0.51	69.21±0.13	71.76±0.25
Cov-H6	82.55±0.88	61.25±2.57	60.32±2.75	67.28±1.77	73.51±0.92	68.11±0.44	73.02±0.26
Cov-H8	82.12±0.42	61.11±1.18	65.66±1.75	68.08±1.54	74.37±0.42	69.71±0.21	72.57±0.06
Cov-H10	82.90±0.33	60.27±1.29	66.67±0.00	71.25±4.30	74.76±0.49	68.31±0.22	72.73±0.21
Cov-L2	76.54±0.12	39.76±0.80	46.46±4.63	61.48±0.86	61.85±0.36	47.70±0.17	39.49±0.26
Cov-L3	77.86±0.10	32.83±1.93	33.33±0.00	50.00±0.00	54.04±0.46	35.51±0.49	26.03±0.33
Cov-L4	77.43±0.19	40.73±0.47	62.01±8.07	56.00±1.37	62.52±0.57	46.18±0.37	55.25±0.25
Cov-L6	80.69±0.25	55.12±1.31	64.95±1.45	62.52±4.20	68.12±0.37	60.72±0.21	63.99±0.55
Cov-L8	77.77±0.20	44.89±0.75	49.78±0.00	58.74±2.69	65.13±0.50	57.15±0.12	62.25±0.12
Cov-L10	78.62±0.13	48.91±1.38	51.08±0.00	61.43±1.69	64.84±0.44	59.62±0.09	68.47±0.09

Configuration	SEED-VII	FACED	PhysioNet	BCIC20-3	SHU-MI	Workload	AVG13
PDV8	37.68±0.46	25.66±0.09	45.90±3.15	50.52±0.65	61.58±1.00	67.82±1.05	60.75
PDV10	37.84±0.27	24.62±0.77	46.71±0.35	50.00±0.00	61.62±1.35	64.48±1.38	60.47
Size-H2	33.07±0.28	23.83±0.59	34.95±0.68	49.90±1.42	56.62±3.31	64.37±4.39	56.19
Size-H3	33.14±0.37	27.45±1.05	35.59±0.65	50.47±1.22	62.07±1.53	63.68±2.24	56.90
Size-H4	33.35±0.72	25.62±0.59	37.17±1.14	50.42±0.92	60.38±1.37	64.48±3.16	57.06
Size-H6	38.94±1.30	26.83±2.10	40.13±1.51	51.41±0.56	60.95±0.93	63.22±2.42	58.33
Size-H8	37.84±0.43	22.91±0.60	38.95±1.45	50.26±0.59	60.45±0.63	65.17±1.24	58.71
Size-H10	38.35±0.76	23.77±1.18	48.35±1.34	49.06±1.36	60.97±0.40	70.92±0.53	60.10
Size-L2	19.57±0.10	13.86±0.54	26.90±1.12	50.00±0.00	50.41±0.71	55.40±0.53	45.70
Size-L3	19.52±0.35	15.44±0.16	27.21±0.23	50.42±0.72	50.25±0.59	57.13±2.19	45.90
Size-L4	21.83±0.15	17.31±0.20	31.90±0.36	50.78±1.96	57.43±0.93	53.10±1.03	50.39
Size-L6	31.56±0.18	25.47±0.62	36.78±0.24	50.05±0.55	59.82±0.18	61.15±0.80	55.67
Size-L8	34.53±0.30	28.91±0.42	45.69±0.14	49.64±0.63	60.81±0.79	62.76±3.63	57.09
Size-L10	36.42±0.42	27.71±0.51	44.93±0.96	49.48±0.45	60.89±0.49	62.30±0.40	57.65
Cov-H2	33.48±0.09	25.82±1.39	39.26±2.08	50.10±0.95	60.86±0.67	67.36±1.59	58.12
Cov-H3	38.72±0.70	33.06±1.64	50.43±0.24	52.08±1.83	61.57±1.05	68.62±4.07	61.72
Cov-H4	39.39±0.46	25.48±1.03	46.36±0.54	50.00±0.00	62.23±0.91	68.85±0.87	60.84
Cov-H6	38.32±0.36	25.68±2.26	47.62±0.45	50.00±0.00	61.34±1.03	71.38±1.19	60.03
Cov-H8	38.24±0.67	22.12±1.29	48.14±1.48	50.00±0.00	61.55±0.94	66.90±1.58	60.04
Cov-H10	37.09±0.63	21.94±1.83	42.98±0.63	51.72±0.72	63.19±0.78	67.01±0.87	60.06
Cov-L2	20.77±0.07	12.84±0.60	26.84±0.41	50.00±0.00	50.22±1.35	56.32±1.77	45.41
Cov-L3	17.26±0.14	12.63±0.19	21.64±0.20	50.00±0.00	52.16±0.99	51.15±3.01	39.57
Cov-L4	22.32±0.25	17.99±0.43	30.45±0.81	50.31±0.98	55.62±0.74	57.59±0.91	48.80
Cov-L6	30.39±0.17	27.81±0.61	37.87±0.36	53.59±1.18	62.48±0.20	64.71±1.70	56.38
Cov-L8	25.77±0.08	17.64±0.38	30.99±0.60	50.57±0.33	56.57±0.42	53.79±2.74	50.08
Cov-L10	26.16±0.08	20.35±0.32	31.33±0.83	47.76±0.18	60.10±0.46	53.79±1.03	51.73

Table 13: CBraMod: complete source-selection task results, part 1 of 2. Rows run from P-H2 to Size-H4; all available source counts are retained.

Configuration	TUAB	TUEV	TUSL	Siena	HMC	SEED	SEED-IV
P-H2	81.53±0.43	53.82±0.45	76.24±5.50	64.61±2.02	66.42±0.26	52.77±2.00	64.74±0.66
P-H3	81.15±0.28	54.07±2.06	74.13±3.19	63.68±4.61	66.35±0.44	52.46±2.23	66.60±0.35
P-H4	81.46±0.40	55.30±0.67	75.23±6.36	66.21±1.65	67.84±0.36	54.30±0.88	66.37±0.65
P-H6	81.85±0.47	53.06±3.67	79.27±6.62	63.90±1.27	66.58±0.82	55.45±0.89	66.77±0.49
P-H8	81.54±0.13	52.53±1.17	84.45±3.77	67.12±2.68	66.55±0.84	55.25±0.08	67.59±0.50
P-H10	80.93±0.09	52.39±1.07	71.57±1.45	64.66±2.67	66.57±1.07	53.97±0.51	65.50±0.28
P-L2	80.28±0.57	49.09±2.34	41.90±5.56	61.61±0.93	65.63±1.27	49.51±1.79	32.65±0.92
P-L3	79.49±0.57	49.78±1.08	47.76±8.38	63.40±1.49	65.04±0.24	55.52±0.80	34.67±0.86
P-L4	79.50±0.41	49.88±1.67	48.48±8.65	59.74±4.18	65.01±0.29	55.60±0.42	32.42±0.85
P-L6	79.51±0.46	47.86±0.71	71.33±5.65	63.77±1.61	66.08±0.04	52.74±0.24	66.71±1.43
P-L8	80.30±0.18	50.17±2.32	77.91±4.01	64.61±2.99	65.90±0.20	50.52±0.65	67.25±0.28
P-L10	80.25±0.15	49.38±0.78	78.83±4.57	61.91±6.10	64.72±0.76	51.71±0.71	66.81±0.57
PDV2	81.09±0.11	50.68±1.98	81.99±2.21	63.83±3.23	67.27±0.13	57.86±0.64	62.56±0.09
PDV3	80.87±0.46	53.61±1.62	72.63±6.39	64.37±3.52	66.04±0.28	50.89±0.61	65.40±0.86
PDV4	80.29±0.22	55.54±0.57	77.03±7.43	64.76±1.48	67.87±0.68	59.40±0.34	68.94±0.35
PDV5	81.98±0.45	52.19±2.28	76.28±3.03	64.78±1.77	65.63±0.64	52.72±1.28	68.08±0.34
PDV6	80.85±0.26	53.48±1.44	81.77±4.63	66.23±0.43	66.38±0.42	56.00±1.02	64.62±1.21
PDV8	80.91±0.66	51.11±0.91	79.88±3.98	66.71±1.70	67.90±0.85	56.82±0.47	66.15±0.84
PDV10	81.71±0.49	51.47±3.15	74.26±5.36	67.10±1.96	67.39±0.19	56.48±1.20	66.09±1.13
Size-H2	80.53±0.14	52.41±1.07	79.31±6.66	66.78±2.02	66.43±0.52	54.34±1.11	64.94±0.47
Size-H3	80.62±0.35	52.20±0.70	75.27±6.31	62.92±3.63	65.61±0.10	50.31±1.00	67.82±0.46
Size-H4	80.61±0.17	53.38±1.12	84.85±3.54	63.23±1.71	66.70±0.43	55.77±0.25	65.59±0.96

Configuration	SEED-VII	FACED	PhysioNet	BCIC20-3	SHU-MI	Workload	AVG13
P-H2	23.14±0.27	36.40±1.47	37.43±1.68	50.89±1.76	63.20±0.47	55.75±5.90	55.92
P-H3	22.95±0.33	37.44±0.82	37.74±0.69	51.82±1.95	60.03±2.28	56.44±6.19	55.76
P-H4	23.89±0.22	37.58±1.38	38.27±1.28	52.86±2.84	60.97±0.90	59.31±4.07	56.89
P-H6	23.73±0.57	37.25±0.56	38.18±1.75	50.83±2.42	61.77±1.02	60.69±1.03	56.87
P-H8	24.33±0.14	38.11±1.04	36.39±0.67	53.23±2.12	60.65±2.61	58.16±0.87	57.38
P-H10	25.16±0.61	38.64±0.82	39.30±0.78	51.88±0.87	61.17±1.33	60.46±1.31	56.32
P-L2	23.09±1.36	13.88±1.57	32.80±0.68	49.95±0.09	49.42±0.85	46.55±1.72	45.87
P-L3	22.51±0.64	12.92±1.06	33.67±0.36	50.00±0.00	50.00±0.00	49.08±1.59	47.22
P-L4	23.18±0.80	14.60±1.71	34.16±1.07	50.00±0.00	49.84±0.18	48.39±0.80	46.99
P-L6	21.80±0.44	27.98±1.32	32.52±0.64	48.91±2.74	51.09±0.11	55.06±2.59	52.72
P-L8	22.09±0.23	29.57±0.88	31.80±0.50	50.52±3.80	52.10±1.80	52.30±2.42	53.46
P-L10	22.31±0.44	30.25±0.30	34.74±0.85	51.04±3.54	53.78±2.99	53.56±1.44	53.79
PDV2	24.80±0.11	30.49±1.77	42.11±0.59	57.24±2.02	57.78±0.71	56.44±3.98	56.47
PDV3	23.02±0.35	37.56±1.05	39.95±1.15	52.86±2.67	60.88±1.59	63.33±1.11	56.26
PDV4	24.90±0.98	46.39±1.70	42.78±1.08	51.56±3.02	61.38±3.76	55.52±1.03	58.18
PDV5	23.35±0.45	37.93±1.38	36.69±1.55	53.12±0.87	61.09±4.31	54.60±0.72	56.04
PDV6	24.77±0.15	37.84±1.53	40.18±0.70	53.75±2.82	60.94±2.18	56.90±7.95	57.21
PDV8	25.08±0.48	37.21±1.54	39.30±0.82	51.35±0.59	60.77±2.88	57.36±2.87	56.97
PDV10	24.44±0.83	36.87±1.36	39.78±0.14	50.78±1.08	60.63±2.41	58.51±1.44	56.58
Size-H2	24.70±0.41	37.97±0.46	39.93±1.45	52.66±1.65	60.65±0.18	55.98±3.65	56.66
Size-H3	24.32±0.53	39.15±1.24	39.43±0.97	51.30±1.18	61.03±2.26	55.63±2.11	55.82
Size-H4	23.77±0.24	37.65±0.99	39.28±1.39	54.38±2.97	62.41±2.03	58.16±0.80	57.37

Table 14: CBraMod: complete source-selection task results, part 2 of 2. Rows run from Size-H6 to Cov-L10; all available source counts are retained.

Configuration	TUAB	TUEV	TUSL	Siena	HMC	SEED	SEED-IV
Size-H6	80.70±0.90	49.39±1.87	70.09±7.04	66.95±2.40	67.45±0.60	56.68±0.20	65.10±1.09
Size-H8	80.03±1.88	53.20±0.38	76.24±3.98	67.58±3.45	66.31±0.66	55.83±0.23	66.73±0.29
Size-H10	80.76±1.00	52.71±2.69	76.14±3.96	62.31±1.61	65.66±0.36	57.31±0.92	66.41±0.81
Size-L2	79.50±0.38	48.34±1.99	72.19±13.96	63.74±1.46	68.30±0.29	53.90±1.22	55.57±0.72
Size-L3	80.14±0.15	48.46±1.30	79.21±6.14	60.29±3.36	67.58±0.35	54.45±0.79	57.71±0.47
Size-L4	79.87±0.29	54.25±0.91	55.30±7.25	66.96±1.30	67.48±0.85	57.16±0.52	61.41±0.49
Size-L6	80.42±0.34	50.99±0.93	49.68±9.06	68.58±3.25	66.42±0.66	53.90±0.54	52.71±1.67
Size-L8	79.77±0.26	50.82±1.83	70.12±3.55	65.11±2.31	66.07±0.42	54.47±0.71	54.85±0.77
Size-L10	80.56±0.67	51.86±0.33	76.90±3.78	64.95±2.39	66.07±0.54	51.75±0.65	58.23±1.24
Cov-H2	80.48±0.42	52.18±1.90	70.18±2.59	65.28±2.07	66.43±0.51	54.12±0.81	66.48±0.14
Cov-H3	81.79±0.30	53.15±1.67	81.38±4.56	65.78±0.90	66.23±0.72	52.29±3.30	61.74±0.81
Cov-H4	79.82±0.47	52.56±2.15	78.26±0.91	64.80±1.80	65.54±0.60	58.95±1.93	69.57±0.44
Cov-H6	80.65±0.62	55.77±1.46	84.37±1.75	65.61±2.56	68.62±1.13	60.40±0.23	69.60±0.11
Cov-H8	81.01±0.12	53.00±2.49	75.22±7.01	67.18±2.06	67.64±0.49	58.99±0.90	62.96±0.81
Cov-H10	80.33±0.40	51.06±1.89	74.70±3.03	66.06±0.78	66.93±1.09	56.99±0.56	64.38±0.74
Cov-L2	78.73±0.39	50.93±1.68	50.91±3.65	65.73±4.51	68.95±0.57	55.94±0.56	47.16±1.22
Cov-L3	79.40±0.13	53.68±1.08	61.42±12.03	64.93±2.20	69.60±0.47	57.69±0.48	41.75±0.70
Cov-L4	79.91±0.37	46.85±4.30	75.54±2.23	65.00±2.01	68.37±0.79	56.37±0.68	64.55±0.39
Cov-L6	80.16±0.29	45.09±3.10	44.65±1.73	61.46±0.55	64.52±0.27	48.25±1.17	33.16±0.78
Cov-L8	80.10±0.50	48.71±2.35	82.96±5.71	63.11±0.66	65.18±0.76	48.01±0.44	68.54±0.23
Cov-L10	80.40±0.13	47.95±0.98	81.95±3.72	63.07±2.31	66.80±0.26	46.61±1.73	68.45±0.31

Configuration	SEED-VII	FACED	PhysioNet	BCIC20-3	SHU-MI	Workload	AVG13
Size-H6	24.32±0.55	37.35±0.90	38.69±0.29	50.99±3.21	60.89±0.57	59.89±1.00	56.04
Size-H8	25.48±0.37	38.59±1.56	39.12±0.90	50.21±1.37	61.43±0.36	56.67±3.52	56.72
Size-H10	24.58±0.61	38.74±0.89	39.39±0.93	52.14±1.06	62.06±1.54	61.95±1.70	56.93
Size-L2	21.48±0.36	22.23±0.68	33.40±0.63	50.05±0.09	50.34±0.61	50.46±1.90	51.50
Size-L3	21.37±0.36	22.25±1.65	33.47±0.10	50.26±1.53	52.18±1.67	50.34±3.60	52.13
Size-L4	25.30±0.20	31.22±2.71	35.96±1.14	50.52±3.24	51.87±1.01	57.59±2.26	53.45
Size-L6	21.15±0.11	24.56±1.01	34.78±1.15	48.39±0.55	51.32±0.81	56.09±1.21	50.69
Size-L8	23.15±0.53	26.51±3.14	40.61±0.39	51.15±2.40	53.02±1.42	52.30±1.00	52.92
Size-L10	23.98±0.21	29.59±0.35	38.86±0.95	50.16±2.15	52.98±3.24	51.49±3.80	53.64
Cov-H2	22.91±0.53	36.03±0.28	39.80±1.42	51.30±2.78	61.59±1.73	54.83±0.91	55.51
Cov-H3	22.98±0.21	36.02±0.83	39.18±1.56	52.29±2.22	61.78±1.41	60.69±3.29	56.56
Cov-H4	23.81±0.69	37.85±0.24	38.95±1.83	50.05±1.69	63.45±0.66	58.85±0.72	57.11
Cov-H6	25.98±0.56	36.61±0.59	39.80±1.00	51.20±1.00	61.80±1.71	58.28±1.50	58.36
Cov-H8	26.10±0.57	38.81±2.06	39.90±0.74	50.26±1.60	61.24±3.37	56.67±5.87	56.84
Cov-H10	25.08±0.40	38.14±0.83	38.91±0.89	51.09±2.78	61.68±0.83	58.62±1.38	56.46
Cov-L2	25.67±0.17	21.26±1.94	32.31±1.40	49.79±0.36	49.77±0.27	53.68±1.21	50.06
Cov-L3	27.12±1.52	19.16±1.51	33.85±0.22	49.90±0.18	50.02±0.14	58.16±0.40	51.28
Cov-L4	24.11±0.08	23.00±0.65	36.02±0.66	52.92±2.35	56.40±2.72	62.64±0.40	54.74
Cov-L6	20.94±0.54	17.55±0.38	34.97±0.71	50.00±0.00	49.94±0.11	47.93±6.06	46.05
Cov-L8	22.07±0.25	30.68±0.56	32.53±1.02	51.67±1.33	54.99±1.07	52.87±1.39	53.96
Cov-L10	22.19±0.53	30.69±0.87	36.29±0.69	53.59±1.80	57.32±1.76	55.29±1.55	54.66

Table 15: BIOT: complete source-selection task results, part 1 of 1. Rows run from P-H2 to Size-L10; all available source counts are retained.

Configuration	TUAB	TUEV	TUSL	Siena	HMC	SEED	SEED-IV
P-H2	77.62±0.18	44.77±0.43	65.07±13.22	61.17±3.06	69.53±1.14	70.14±0.25	64.43±0.10
P-H3	78.52±0.11	45.34±3.12	71.43±14.64	64.13±4.89	69.09±0.12	70.04±0.26	65.30±0.16
P-H4	78.45±0.15	45.58±1.18	68.20±9.60	68.45±1.90	69.35±0.46	69.80±0.15	64.97±0.45
P-H6	78.49±0.16	45.05±3.02	66.62±9.71	68.37±2.45	69.49±0.13	69.11±0.31	64.59±0.25
P-H8	78.20±0.31	46.67±0.53	63.96±7.21	70.49±5.34	68.80±0.71	69.65±0.37	61.32±0.29
P-H10	77.87±0.18	45.83±1.79	58.42±16.58	66.18±2.23	69.30±0.14	63.92±0.86	49.26±0.68
P-L2	77.22±0.36	43.52±2.62	54.42±11.51	71.74±5.60	69.21±0.28	66.83±0.20	53.51±1.03
P-L3	76.83±0.20	44.55±3.52	58.30±6.87	68.60±5.74	70.61±0.44	67.50±0.54	54.32±0.51
P-L4	76.55±0.49	44.71±2.43	35.23±9.97	66.19±2.38	68.89±0.24	64.08±0.75	49.72±0.85
P-L6	77.72±0.25	47.87±0.49	48.31±19.27	70.26±2.51	68.93±0.21	66.51±0.39	48.53±0.84
P-L8	77.80±0.01	45.88±1.42	45.56±6.83	68.29±1.07	69.51±0.52	65.27±0.53	44.70±0.61
P-L10	77.51±0.37	46.49±0.49	44.86±8.91	65.21±4.19	69.19±0.85	65.04±0.15	47.56±0.50
Size-H2	75.57±1.40	39.23±5.16	41.48±5.99	65.97±0.98	69.42±0.27	62.60±0.39	44.41±0.74
Size-H3	75.91±0.52	40.01±1.57	49.08±16.18	62.88±3.46	70.24±0.45	63.31±0.45	43.14±2.82
Size-H4	76.52±0.39	39.32±0.78	40.22±3.42	66.21±4.21	69.38±0.44	63.32±0.57	42.30±1.43
Size-H6	76.62±0.65	39.44±0.86	50.72±4.77	68.75±1.66	68.48±0.22	62.82±1.18	43.92±0.72
Size-H8	76.46±0.14	42.14±3.93	44.75±3.24	68.47±3.50	69.03±1.16	63.85±0.79	48.49±0.47
Size-H10	75.93±1.19	40.76±3.92	48.79±5.20	66.19±1.13	67.96±0.42	62.43±0.31	44.50±3.88
Size-L2	77.68±0.24	42.54±0.41	80.68±5.32	70.00±3.00	68.22±0.16	67.47±0.61	56.83±0.70
Size-L3	78.50±0.45	41.40±1.39	80.68±5.29	61.66±5.62	68.32±0.33	67.43±0.28	53.70±0.83
Size-L4	77.13±0.39	43.62±2.71	59.65±6.38	69.92±3.41	69.50±0.82	67.88±0.33	54.13±0.59
Size-L6	76.23±0.83	43.87±5.20	62.48±7.93	68.57±0.44	68.37±0.58	65.70±1.44	50.89±4.22
Size-L8	76.87±0.28	44.94±2.96	54.41±14.73	65.81±8.81	69.46±0.60	63.52±1.17	45.69±3.15
Size-L10	77.33±0.97	43.99±1.91	54.21±9.08	68.52±1.44	69.50±0.22	60.61±0.73	40.84±0.54

Configuration	SEED-VII	FACED	PhysioNet	BCIC20-3	SHU-MI	Workload	AVG13
P-H2	36.40±0.16	20.11±0.81	27.02±0.68	52.08±0.33	58.83±0.59	68.39±2.19	55.04
P-H3	35.78±0.72	19.90±1.12	26.23±1.18	50.94±0.27	58.76±1.22	72.30±1.31	55.98
P-H4	36.16±0.50	21.43±0.51	24.97±0.63	50.68±0.86	59.25±0.47	63.45±4.07	55.44
P-H6	36.02±0.50	19.59±1.44	23.53±0.78	49.64±0.63	55.63±1.08	64.83±1.24	54.69
P-H8	35.89±0.45	21.11±0.73	24.66±0.25	50.42±1.15	54.12±0.94	60.11±2.24	54.26
P-H10	35.39±1.39	17.66±0.53	23.30±0.51	50.36±0.70	50.43±1.53	64.71±0.40	51.74
P-L2	35.22±0.12	16.63±0.91	24.05±0.77	50.21±0.90	52.21±1.13	70.11±1.90	52.68
P-L3	35.76±0.46	17.43±0.91	23.95±1.45	50.10±0.48	54.58±2.61	66.32±8.17	52.99
P-L4	35.39±0.72	16.41±0.40	23.87±0.78	49.90±1.88	51.46±0.84	63.79±5.01	49.71
P-L6	33.82±0.78	16.55±1.16	24.13±0.48	50.94±3.43	51.15±1.12	66.09±2.54	51.60
P-L8	33.23±0.53	17.66±1.25	24.39±0.71	52.60±4.78	51.86±1.05	61.72±1.50	50.65
P-L10	33.75±0.87	16.40±0.49	21.46±0.75	51.56±1.65	50.06±0.32	67.36±1.59	50.50
Size-H2	31.32±1.19	16.10±0.47	23.66±0.24	51.88±1.96	50.16±0.68	66.90±2.82	49.13
Size-H3	31.57±0.87	15.80±0.67	23.89±0.50	52.76±1.49	49.49±0.91	67.70±1.05	49.68
Size-H4	32.68±1.39	15.57±0.64	23.68±0.33	49.32±0.63	49.12±0.52	65.98±5.81	48.74
Size-H6	33.41±0.26	15.21±1.37	23.89±0.85	51.67±1.88	49.55±0.32	60.69±8.21	49.63
Size-H8	31.52±0.45	14.68±1.44	24.13±0.54	49.06±3.69	50.09±1.22	67.24±2.82	49.99
Size-H10	31.52±0.41	15.63±1.58	22.43±1.89	49.27±4.96	49.67±0.81	57.47±7.37	48.66
Size-L2	35.75±0.49	20.83±0.61	26.15±0.33	47.71±2.35	57.67±0.33	65.06±1.11	55.12
Size-L3	35.98±0.84	19.79±0.71	25.90±0.41	48.49±1.49	58.36±0.50	62.87±2.49	54.08
Size-L4	34.40±1.34	17.24±0.58	22.80±0.37	50.05±0.09	55.52±0.68	55.75±2.02	52.12
Size-L6	34.20±0.02	19.52±1.53	23.92±1.07	49.22±1.77	56.14±1.88	68.85±1.55	52.92
Size-L8	32.46±1.05	16.33±0.84	23.77±0.69	49.84±0.78	54.39±0.85	66.67±5.53	51.09
Size-L10	30.56±0.22	15.98±1.18	22.78±2.16	50.99±1.02	49.63±0.52	57.93±4.48	49.45

C.4 REMAINING CONFIGURATIONS IN THE OVERVIEW

These tables complete the 202-configuration inventory in Figures 2 and 3. They retain the random subsets and weighting, filtering, curation, and source-removal results that are not individually discussed in the main text. The tables retain both improvements and declines relative to Full16. Random-k identifies a k-source subset and its number identifies the sampled composition. "Without" identifies the omitted source. The source-removal export does not specify its mixing protocol, and the filtering labels do not establish a single-factor intervention; these rows are descriptive configuration results.

Table 16: LaBraM: remaining overview configurations, part 1 of 1. Rows follow decreasing observed AVG13.

Configuration	TUAB	TUEV	TUSL	Siena	HMC	SEED	SEED-IV
DoReMi (ref16)	82.64±0.20	61.00±0.60	65.08±2.75	61.83±1.65	73.03±0.29	65.34±0.12	68.05±0.12
GCAW	81.57±0.05	52.69±1.20	58.30±2.00	58.33±2.74	72.06±0.16	64.71±0.27	69.11±0.10
DoReMi (ref8)	81.07±0.08	54.37±0.38	66.67±0.00	61.63±1.46	72.40±0.05	63.71±0.25	64.67±0.56
Random-3 #5	77.87±0.29	45.05±1.44	44.59±2.41	63.10±1.44	63.10±0.27	53.95±1.43	67.02±0.37
Random-3 #8	77.84±0.23	45.52±1.15	44.59±2.41	61.72±2.43	63.24±0.83	53.92±1.39	67.30±0.35
Random-3 #1	77.86±0.30	44.06±1.42	44.59±2.41	61.30±2.22	63.09±0.21	54.05±1.71	67.28±0.38

Configuration	SEED-VII	FACED	PhysioNet	BCIC20-3	SHU-MI	Workload	AVG13
DoReMi (ref16)	36.20±0.24	20.54±0.57	36.33±0.90	50.00±0.00	59.33±1.67	63.68±1.55	57.16
GCAW	33.62±0.25	22.67±1.44	39.85±1.41	50.00±0.00	60.07±1.10	64.37±1.39	55.95
DoReMi (ref8)	34.27±0.39	19.92±0.73	32.98±1.00	50.52±1.58	57.02±0.84	60.46±1.31	55.36
Random-3 #5	21.56±1.04	15.62±1.13	31.00±1.23	49.64±0.63	51.24±0.81	54.37±0.72	49.09
Random-3 #8	21.60±1.24	15.72±1.04	30.65±1.28	49.64±0.63	51.25±0.82	54.60±1.11	49.05
Random-3 #1	21.49±1.06	15.62±1.13	31.57±0.40	50.00±0.00	51.01±1.16	54.71±0.40	48.97

Table 17: BIOT: remaining overview configurations, part 1 of 1. Rows follow decreasing observed AVG13.

Configuration	TUAB	TUEV	TUSL	Siena	HMC	SEED	SEED-IV
Memory filter	77.14±0.26	44.62±1.44	41.88±8.98	68.79±3.77	70.06±0.10	64.85±0.33	53.57±1.29
Self-score filter	77.46±0.46	44.49±1.33	49.35±11.29	63.45±5.58	69.53±0.45	64.52±0.74	50.88±0.34
GCAW	77.73±0.34	40.86±1.71	47.56±23.26	66.03±2.82	69.92±0.45	61.43±0.29	48.05±2.04

Configuration	SEED-VII	FACED	PhysioNet	BCIC20-3	SHU-MI	Workload	AVG13
Memory filter	34.92±1.03	17.54±0.62	22.63±1.76	50.47±1.92	50.47±0.19	66.44±3.26	51.03
Self-score filter	33.54±1.05	16.52±0.62	22.91±2.18	52.14±2.23	50.35±1.27	59.43±3.81	50.35
GCAW	30.59±1.36	15.58±0.75	23.79±0.77	50.26±0.59	50.79±0.77	62.87±3.40	49.65

Table 18: CBraMod: remaining overview configurations, part 1 of 3. Rows follow decreasing observed AVG13.

Configuration	TUAB	TUEV	TUSL	Siena	HMC	SEED	SEED-IV
Random-15 #1	81.14±0.38	49.90±1.88	79.80±5.71	66.57±1.88	65.72±0.32	52.75±0.51	68.06±0.52
Without BCIC-IV-2a	80.42±0.11	52.69±0.82	75.23±2.29	68.16±2.22	65.24±0.21	52.18±0.09	68.51±0.31
Random-14 #2	81.19±0.34	51.45±1.61	77.25±4.41	66.49±1.09	66.64±0.89	52.08±0.70	67.79±0.89
Random-7 #0	80.80±0.20	52.49±1.75	78.30±1.75	66.16±1.75	65.56±0.94	53.10±1.43	65.61±0.35
Without Siena	80.61±0.17	53.76±3.16	82.30±3.98	63.94±2.63	64.99±1.47	49.68±1.87	67.96±0.48
Without SHU-MI	81.88±0.24	52.23±1.95	83.79±5.25	64.34±4.02	66.59±0.50	54.80±0.23	67.20±0.97
Without SEED-VII	81.54±0.37	51.77±1.88	72.68±1.74	67.11±3.47	66.56±0.56	52.47±1.56	68.36±0.65
Random-7 #2	80.85±0.27	52.56±2.60	75.71±4.50	66.52±1.74	65.75±0.61	51.91±0.58	67.77±0.13
Random-11 #0	81.12±0.05	54.08±2.23	80.28±3.14	64.00±3.97	64.99±0.47	52.97±0.29	67.57±0.30
Without HMC	81.26±0.23	53.29±0.99	82.30±4.55	65.81±2.78	65.36±0.45	49.30±1.52	67.33±0.50
Without BCIC-2020/3	81.52±0.13	53.08±5.30	72.15±3.25	67.19±1.12	66.28±0.53	53.01±1.14	67.74±0.25
Random-13 #2	81.06±0.12	50.51±0.98	75.27±9.83	64.49±1.30	66.48±0.94	50.97±1.00	68.22±0.35
Curated 7	80.92±0.29	52.08±2.57	77.34±8.39	65.36±1.92	65.59±0.52	51.27±0.42	65.48±0.39
Without FACED	81.60±0.33	50.19±1.27	69.17±2.29	65.90±0.90	65.87±0.29	54.63±0.50	68.01±0.59
GCAW	80.95±0.40	50.89±1.38	77.73±3.03	67.28±1.31	65.95±0.38	52.32±1.63	68.05±0.38
Without PhysioNet	81.48±0.30	50.03±1.01	82.30±4.03	63.24±0.84	66.67±0.30	53.60±0.74	68.04±0.10
Random-13 #0	81.23±0.10	52.84±2.15	77.73±4.09	60.73±3.02	66.55±0.71	53.48±1.60	66.51±0.58
Random-6 #0	80.91±0.38	54.39±0.93	75.18±4.78	62.98±3.68	66.09±0.41	52.43±0.45	65.27±0.34
Random-11 #2	81.07±0.73	53.25±1.37	77.73±3.03	65.64±2.28	65.70±0.39	51.30±1.55	67.95±0.81
Without TUEV	81.11±0.23	53.18±2.59	75.71±9.74	66.99±1.38	65.95±0.43	51.37±0.15	66.32±0.21
Without SCD	81.34±0.19	49.59±2.62	79.27±1.52	62.25±0.38	65.96±0.73	50.05±0.25	68.17±0.49
Without Workload	81.27±0.60	51.65±1.56	78.74±0.87	65.27±0.97	66.53±0.19	49.05±0.66	68.11±0.23

Configuration	SEED-VII	FACED	PhysioNet	BCIC20-3	SHU-MI	Workload	AVG13
Random-15 #1	23.79±0.45	39.81±0.45	37.86±0.93	53.18±5.51	63.58±1.56	56.32±2.81	56.81
Without BCIC-IV-2a	24.73±0.66	37.42±1.38	40.61±1.60	52.81±2.98	61.25±1.32	55.63±4.68	56.53
Random-14 #2	23.94±0.39	37.96±1.71	36.88±1.79	54.32±1.86	61.94±0.94	55.86±4.19	56.45
Random-7 #0	23.80±0.25	36.72±0.28	37.65±1.27	51.98±0.55	60.72±0.86	60.46±1.55	56.41
Without Siena	23.40±0.73	38.64±0.68	38.81±1.25	51.88±2.05	62.11±3.00	55.17±1.19	56.40
Without SHU-MI	23.94±0.43	35.35±1.22	36.57±1.35	48.49±3.16	60.92±0.61	56.67±3.26	56.37
Without SEED-VII	23.54±0.06	36.53±1.33	38.40±0.57	54.22±2.98	59.43±2.95	59.77±1.39	56.34
Random-7 #2	23.64±0.70	38.48±1.41	39.89±0.76	52.29±2.08	60.03±2.71	56.78±4.55	56.32
Random-11 #0	23.82±0.71	38.48±1.33	38.15±1.39	51.93±1.04	60.17±0.87	54.48±1.50	56.31
Without HMC	23.95±0.70	37.56±0.75	36.81±0.34	52.14±2.29	61.75±0.32	54.60±1.00	56.27
Without BCIC-2020/3	24.11±0.11	38.83±1.77	37.16±1.53	52.60±2.13	61.56±2.78	54.71±2.45	56.15
Random-13 #2	23.65±0.82	38.69±0.46	37.91±0.65	52.66±3.46	62.25±0.14	57.47±2.02	56.13
Curated 7	24.40±0.44	37.77±2.14	40.40±0.27	52.08±1.94	61.49±3.00	55.40±5.87	56.12
Without FACED	24.01±0.58	38.59±1.10	37.05±0.83	54.22±1.59	64.02±0.78	56.21±1.79	56.11
GCAW	23.08±0.65	37.80±0.61	37.45±1.11	52.29±2.81	62.11±1.80	53.33±3.11	56.10
Without PhysioNet	23.72±0.54	34.36±1.86	37.26±1.17	51.72±0.31	58.90±0.81	57.47±1.90	56.06
Random-13 #0	23.12±0.69	36.67±0.61	38.62±1.58	51.30±1.02	62.51±2.36	57.47±4.00	56.06
Random-6 #0	23.37±0.23	38.98±0.53	37.61±0.89	51.77±1.11	62.90±1.94	56.78±3.13	56.05
Random-11 #2	23.33±0.47	38.22±1.30	37.08±1.20	51.35±1.86	61.68±0.86	54.14±4.31	56.03
Without TUEV	23.79±0.04	36.86±1.65	37.06±1.98	52.50±0.94	60.15±1.74	57.13±1.77	56.01
Without SCD	24.07±0.73	38.97±1.54	36.60±1.00	54.58±2.27	62.21±2.20	54.48±1.82	55.97
Without Workload	23.04±0.26	39.45±0.49	38.69±0.50	48.85±3.17	60.29±0.49	56.09±3.72	55.93

Table 19: CBraMod: remaining overview configurations, part 2 of 3. Rows follow decreasing observed AVG13.

Configuration	TUAB	TUEV	TUSL	Siena	HMC	SEED	SEED-IV
Random-11 #1	81.74±0.15	51.34±2.23	74.13±4.53	64.96±2.34	66.66±0.79	50.16±0.64	67.42±0.23
Random-14 #0	81.02±0.18	50.18±3.42	73.16±7.05	66.81±3.53	66.41±0.52	52.48±1.24	68.58±0.27
Without TUSL	82.15±0.15	51.70±2.82	73.21±2.59	64.25±3.51	65.31±0.44	49.05±0.40	68.17±0.13
Memory filter	81.99±0.43	52.05±2.23	76.67±4.56	65.24±0.45	66.24±0.96	50.78±1.36	68.05±0.32
Without SEED-IV	80.94±0.05	51.14±1.97	77.82±5.43	64.42±3.27	65.72±1.11	50.67±1.28	68.33±0.53
Random-15 #0	81.38±0.20	51.47±1.96	72.73±8.92	65.52±0.36	66.04±0.87	52.61±0.34	68.16±0.08
Random-4 #0	80.88±0.26	52.30±1.70	76.91±3.91	66.18±1.87	65.72±0.24	51.92±1.40	64.41±2.60
Without TUAB	80.43±0.44	49.18±0.65	83.79±1.66	64.23±3.52	66.24±0.81	51.02±0.15	68.06±0.88
Random-10 #0	81.24±0.24	50.39±0.90	77.73±10.93	65.33±1.73	66.25±0.93	50.30±0.40	67.58±0.90
Without SEED-V	81.66±0.12	51.87±1.30	69.08±8.45	64.46±0.80	66.79±0.31	52.38±1.41	68.53±0.23
Random-14 #1	82.01±0.53	52.78±3.27	70.04±4.01	66.52±3.78	65.33±0.49	49.57±1.88	67.92±0.20
Curated 4	81.17±0.33	51.72±1.40	72.11±4.92	66.99±1.52	66.78±0.73	52.50±0.56	65.77±0.43
Random-6 #1	80.89±0.12	53.11±2.83	73.03±4.56	67.70±3.99	66.24±0.11	57.97±0.36	65.84±0.88
Random-15 #2	81.09±0.18	50.48±0.95	63.40±0.84	67.10±1.43	65.42±0.28	51.66±0.96	68.25±0.17
Random-3 #0	80.31±0.16	53.36±1.44	81.51±3.24	62.36±0.98	65.94±0.09	54.65±0.42	66.63±0.59
Random-10 #1	80.33±0.34	51.84±2.42	73.69±6.12	65.33±0.14	65.97±0.75	52.50±0.27	68.41±0.55
Random-8 #4	80.72±0.17	48.62±1.88	79.53±2.55	62.59±0.92	66.39±0.67	54.69±0.71	62.45±2.59
LaBraM-score filter	80.82±0.65	53.05±0.95	67.49±8.70	67.70±2.02	65.91±0.34	48.95±0.58	67.60±0.43
Random-12 #2	80.35±0.86	51.05±2.08	81.38±1.79	59.94±3.45	67.04±0.23	50.70±1.10	67.79±0.52
Random-3 #3	79.94±0.35	52.02±0.47	71.18±8.46	62.12±0.93	65.67±0.18	60.74±0.34	64.33±0.98
Random-5 #2	81.85±0.17	51.47±0.21	74.70±6.06	64.77±1.43	66.43±0.54	51.72±1.48	66.70±0.58
Without SEED	81.44±0.09	49.85±2.32	67.05±7.18	66.84±2.16	66.03±0.46	52.39±0.78	68.12±0.41

Configuration	SEED-VII	FACED	PhysioNet	BCIC20-3	SHU-MI	Workload	AVG13
Random-11 #1	23.22±0.22	37.18±3.42	38.99±0.79	52.14±1.81	59.88±1.72	59.20±3.45	55.92
Random-14 #0	23.42±0.41	34.39±2.04	37.83±0.83	54.32±1.45	60.57±2.28	57.82±4.58	55.92
Without TUSL	23.82±0.12	39.17±1.31	39.19±1.53	52.60±2.08	61.30±1.75	56.67±2.68	55.89
Memory filter	23.10±0.10	38.83±1.49	38.66±2.10	50.99±1.94	60.10±1.91	53.68±1.21	55.88
Without SEED-IV	24.28±0.74	37.73±1.09	36.90±1.24	52.08±3.30	61.99±1.75	53.79±3.98	55.83
Random-15 #0	23.72±0.22	39.35±0.50	37.67±0.80	52.66±2.30	60.47±0.14	53.91±3.81	55.82
Random-4 #0	22.88±1.42	36.13±2.53	35.90±1.30	52.14±0.55	60.32±2.97	59.66±4.24	55.80
Without TUAB	23.56±0.59	33.49±1.05	36.35±1.33	51.88±3.33	60.43±2.65	56.21±4.24	55.76
Random-10 #0	23.39±0.52	37.73±0.50	37.64±0.62	51.04±1.57	60.58±1.20	54.83±1.19	55.69
Without SEED-V	24.32±0.41	37.22±3.26	37.41±2.05	52.14±2.94	60.25±2.20	57.47±4.00	55.66
Random-14 #1	23.96±0.29	37.84±0.75	39.57±2.31	50.94±0.95	62.19±2.24	54.60±1.39	55.63
Curated 4	23.79±0.38	36.55±2.03	36.28±0.18	51.09±1.51	60.35±4.64	57.36±5.05	55.57
Random-6 #1	24.89±0.61	33.62±1.32	37.04±2.04	53.85±0.70	57.26±1.31	50.23±2.54	55.51
Random-15 #2	24.01±0.11	37.97±1.64	38.88±2.22	52.40±2.04	62.60±1.18	58.39±2.11	55.51
Random-3 #0	22.66±0.33	32.20±0.82	35.48±0.78	53.70±1.42	53.32±3.53	58.51±2.08	55.43
Random-10 #1	23.08±1.11	33.96±1.89	34.92±1.13	52.55±0.90	59.66±1.77	57.36±2.61	55.36
Random-8 #4	25.40±0.58	34.14±0.81	40.06±1.30	52.03±0.56	60.66±1.44	52.18±4.68	55.34
LaBraM-score filter	23.95±0.26	36.95±2.26	38.42±0.99	50.78±2.67	60.95±1.00	56.44±2.02	55.31
Random-12 #2	22.27±0.84	31.82±1.58	37.75±2.32	53.33±0.39	57.74±0.65	57.70±1.55	55.30
Random-3 #3	25.64±0.49	32.99±0.91	39.42±0.57	53.44±1.09	53.28±3.11	57.47±7.94	55.25
Random-5 #2	22.38±1.23	32.23±0.60	37.44±0.70	51.72±3.89	59.35±3.09	57.36±1.21	55.24
Without SEED	23.52±0.45	39.23±1.06	36.26±1.78	49.90±2.82	62.52±2.02	54.37±3.13	55.19

Table 20: CBraMod: remaining overview configurations, part 3 of 3. Rows follow decreasing observed AVG13.

Configuration	TUAB	TUEV	TUSL	Siena	HMC	SEED	SEED-IV
Random-8 #1	81.09±0.60	51.24±1.66	80.76±4.55	62.91±1.92	65.61±1.00	49.61±1.10	68.46±0.22
Random-3 #2	80.30±0.20	52.09±1.96	82.43±2.70	65.50±1.39	66.46±0.38	56.66±0.84	65.35±0.86
Random-13 #1	81.54±0.44	53.17±2.67	68.64±6.68	62.79±3.41	65.75±0.49	49.88±1.60	67.67±0.51
Random-9 #0	80.79±0.09	49.62±1.52	81.81±4.97	60.73±5.09	66.48±0.16	49.47±0.31	68.02±0.09
Self-B3	79.65±0.79	54.19±1.28	67.36±5.38	66.05±1.96	65.69±1.16	56.32±0.40	63.47±1.83
Random-12 #1	81.56±0.30	51.63±1.83	65.89±0.90	62.76±1.09	65.59±0.93	50.69±0.54	68.31±0.17
Random-9 #2	81.16±0.07	50.09±3.07	78.69±6.39	60.94±2.34	66.65±0.11	50.05±1.62	66.73±0.39
Random-9 #1	80.50±0.25	47.41±0.60	76.67±8.39	62.54±1.79	66.25±1.30	51.31±1.81	68.28±0.34
Random-8 #3	80.18±0.50	52.48±2.03	78.87±0.65	65.56±1.39	65.08±0.58	49.69±1.33	66.81±1.19
Random-5 #1	80.43±0.11	50.88±0.71	80.32±5.36	65.98±0.90	65.88±0.38	52.43±0.78	57.24±2.54
Random-4 #1	79.98±1.32	52.23±0.28	76.37±4.55	62.79±3.02	65.25±0.24	52.05±3.17	63.47±1.70
Random-8 #0	80.63±0.56	49.57±3.93	69.78±11.66	64.80±6.78	65.29±0.90	49.92±0.83	68.45±0.60
Random-12 #0	80.01±0.74	49.75±0.44	71.62±7.72	66.11±2.18	65.60±0.35	49.95±1.47	68.30±0.33
Random-3 #1	80.86±0.27	49.63±1.36	71.57±6.27	64.45±0.37	64.98±0.64	53.14±0.34	62.44±1.20
Random-7 #1	79.83±0.38	50.57±0.48	79.75±7.28	60.64±2.24	66.30±0.43	50.78±1.36	66.96±1.74
Random-6 #2	79.73±0.15	48.48±1.03	82.39±2.66	61.44±1.49	65.52±1.40	50.21±0.95	67.77±0.77
Random-10 #2	80.27±0.04	47.67±1.54	77.29±4.63	62.75±3.18	65.22±0.76	51.20±0.45	67.84±0.52
Random-3 #4	78.84±0.38	51.15±2.53	60.45±8.06	64.51±3.77	67.60±0.70	56.68±0.37	60.40±0.45
Self-S8	79.49±0.53	50.08±1.40	77.38±6.26	63.63±1.40	65.24±0.55	47.07±0.35	58.45±3.45
Random-4 #2	79.91±0.87	52.15±1.17	69.01±4.57	62.52±2.80	66.87±0.13	53.56±0.94	58.15±3.16
Random-8 #2	80.49±0.25	50.58±1.18	64.77±1.67	64.15±2.03	67.15±0.18	52.30±0.80	62.12±1.25
Self-S3	80.27±1.49	46.30±2.14	64.49±10.37	65.77±2.28	66.09±0.54	48.23±1.15	51.61±0.49

Configuration	SEED-VII	FACED	PhysioNet	BCIC20-3	SHU-MI	Workload	AVG13
Random-8 #1	22.96±0.74	33.89±0.80	38.12±0.60	53.65±5.17	53.92±0.89	54.83±1.50	55.16
Random-3 #2	23.07±0.34	29.26±1.71	36.69±0.07	51.56±2.31	51.78±1.12	55.40±1.55	55.12
Random-13 #1	23.61±0.45	38.60±0.71	36.26±2.05	51.56±4.64	61.13±1.13	55.86±1.24	55.11
Random-9 #0	23.71±0.53	33.45±1.22	37.59±0.74	53.18±1.30	57.27±1.77	54.02±4.31	55.09
Self-B3	24.68±0.59	31.20±1.04	35.53±1.06	52.55±2.91	56.54±2.22	62.30±2.35	55.04
Random-12 #1	23.45±0.37	35.82±0.89	38.41±1.38	50.57±2.12	59.71±1.83	60.92±5.74	55.02
Random-9 #2	23.57±0.77	33.77±1.21	37.12±1.20	52.81±2.46	57.09±4.00	55.29±0.87	54.92
Random-9 #1	23.64±0.65	31.75±2.19	36.61±2.45	53.75±0.81	59.78±1.39	55.06±5.28	54.89
Random-8 #3	23.31±0.26	32.56±1.61	35.01±3.15	52.40±0.79	54.82±2.84	55.06±2.77	54.75
Random-5 #1	23.82±0.44	27.98±2.11	37.34±3.18	51.72±2.74	59.15±1.42	56.67±2.61	54.60
Random-4 #1	23.36±1.10	31.36±2.31	38.25±0.32	54.53±2.17	54.78±2.67	55.17±8.39	54.58
Random-8 #0	23.00±0.80	34.69±1.60	35.47±1.10	52.71±3.53	57.08±0.74	55.75±1.39	54.39
Random-12 #0	23.70±0.45	32.18±1.53	35.64±0.55	51.46±2.53	58.41±0.36	50.92±3.35	54.13
Random-3 #1	23.63±0.29	30.79±2.98	38.12±1.29	55.42±3.21	52.73±2.16	55.40±4.69	54.09
Random-7 #1	22.86±0.66	31.09±0.95	34.44±1.18	52.34±2.63	55.82±3.64	51.38±1.72	54.06
Random-6 #2	22.22±0.35	32.77±2.40	32.77±0.81	52.81±2.84	54.59±2.38	51.61±1.63	54.03
Random-10 #2	22.53±0.91	30.64±1.20	34.37±1.58	51.88±1.49	56.66±3.28	53.45±1.03	53.98
Random-3 #4	23.39±0.54	30.48±1.38	34.59±0.62	50.62±2.30	52.79±2.54	61.03±2.76	53.27
Self-S8	21.65±0.26	28.97±0.64	38.61±1.18	49.95±1.18	52.97±3.14	56.78±5.16	53.10
Random-4 #2	21.79±0.27	29.88±1.55	36.49±1.16	51.25±2.63	50.70±0.49	56.44±2.02	52.98
Random-8 #2	23.81±0.18	27.97±1.66	36.81±0.58	51.93±0.63	50.54±0.74	50.92±3.54	52.58
Self-S3	22.26±0.49	25.70±1.73	35.63±1.64	50.68±1.17	50.60±0.88	57.24±2.26	51.14

D Supporting Evidence for Data Composition and Capacity

D.1 The complete configuration inventory

Figure 2 contains the same configurations for all 13 targets within each architecture. Of the 202 configurations, 126 have higher observed AVG13 than Full16, and 121 also use fewer plotted windows. Nineteen of the latter use less than one quarter of Full16's pool: three LaBraM, four CBraMod, and twelve BIOT configurations. The threshold is strictly below $637{,}814/4 = 159{,}453.5$ windows. This group includes BIOT Scratch (zero windows) and CBraMod PDV2 (105,864 candidate-pool windows); the other 17 use recorded training-window counts.

Table 21 gives the per-task counts underlying Figure 3a. All 39 architecture–task combinations have a higher-scoring configuration with fewer windows. These are comparisons between observed means; the evaluated configurations share sources and sometimes checkpoints. The counts describe the tested inventory rather than independent trials or a probability of success for arbitrary subsets.

Table 21: **Configurations with fewer windows and higher task means than Full16.** Each cell gives the number of qualifying configurations over all configurations evaluated for that architecture. Comparisons use the mean of three fine-tuning seeds per task. Full16 and Scratch are included in the denominators. The inventory and recorded window counts follow Figure 2, including candidate-pool counts for pool-based configurations. These counts are descriptive.

Task	LaBraM	CBraMod	BIOT
BCIC20-3	40/61	27/112	10/29
SEED	26/61	106/112	15/29
SEED-IV	29/61	34/112	6/29
SEED-VII	26/61	26/112	12/29
FACED	32/61	37/112	21/29
HMC	30/61	64/112	19/29
PhysioNet	22/61	64/112	21/29
SHU-MI	36/61	22/112	22/29
Siena	48/61	68/112	13/29
TUAB	29/61	10/112	13/29
TUEV	31/61	73/112	8/29
TUSL	37/61	83/112	23/29
Workload	29/61	85/112	17/29

Among the reported subset–Full16 tests, 20 of 23 distinct LaBraM configurations and five of six BIOT configurations pass their respective within-family Holm corrections; none of the 19 CBraMod configurations does. The LaBraM count combines ten Progressive, five Coverage, and six PDV results and counts the shared P-H3/PDV3 run once. Appendix F retains the full comparison families, including nonsignificant results. The complete task tables in Appendix C also retain all configurations below Full16.

D.2 Adding sources from the target task family

Table 22 records the task-family additions used in Section 4.2. Each smaller pool already contains a source from the target family, and each expansion retains all previous sources. Their family-average changes are -2.72, -1.46, and -1.20 points. These descriptive three-seed means show that adding another source from a matching task family need not improve that family's targets consistently. Complete profiles, including targets outside the illustrated family, are in Tables 6 and 8.

The similar-volume comparisons in Section 4.3 complement these expansions: their source counts and memberships differ, while window counts are close. Table 5 identifies every compared configuration, and its following task tables make the opposing responses within emotion and clinical tasks visible alongside AVG13. Neither comparison treats a shared task category as a guarantee of positive transfer.

Table 22: **Adding sources from the target task family.** Three-seed mean score changes after adding one source to the stated pool, in percentage points. Each starting pool already contains a source from the target family. Mean equally weights the targets within each panel. Differences and averages are calculated before rounding.

LaBraM, emotion recognition. Starting pool: TUAB + Siena + SEED-VII. Added source: SEED-IV.

SEED	SEED-IV	SEED-VII	FACED	Mean
$+0.08$	-4.03	$+0.67$	-7.58	-2.72

BIOT, motor imagery. Starting pool: BCIC-IV-2a + Workload + TUSL. Added source: BCIC-2020/3.

PhysioNet	SHU-MI	BCIC20-3	Mean
-3.10	-2.84	$+1.56$	-1.46

LaBraM, motor imagery. Starting pool: TUAB + Siena + SEED-IV + SEED-VII + SEED + PhysioNet + FACED. Added source: SHU-MI.

PhysioNet	SHU-MI	BCIC20-3	Mean
-2.79	-0.43	-0.36	-1.20

D.3 Further nested expansions at Base capacity

The main-text LaBraM path uses $A = \{\text{SCD}, \text{SEED-V}, \text{BCIC-IV-2a}, \text{Workload}\}$, $B = \{\text{TUSL}\}$, and $C = \{\text{FACED}\}$. Its three pools contain 40,403, 40,551, and 48,783 windows and reach 48.80%, 42.96%, and 56.38% AVG13. Table 9 reports every task result. All three pools are pretrained separately with the same Base architecture and 50-epoch protocol; they are not successive checkpoints from one training trajectory.

Table 23 adds five nested comparisons: two LaBraM, two CBraMod, and one BIOT. Every stage retains all previous sources and satisfies the observed ordering $\text{AVG13}(A+B+C) > \text{AVG13}(A) > \text{AVG13}(A+B)$. They show recovery after an initial decline without a capacity increase. The paths were selected from the evaluated inventory by this ordering, sometimes reuse configurations, and do not constitute independent repetitions or isolate exposure from composition. Their complete task scores are retained in Appendix C.3.

Table 23: **Additional strictly nested source-expansion examples.** Within each panel, A lists the complete starting pool; B and C list successive additions. Each step retains all earlier sources. Windows counts the cumulative pool per epoch, and Δ is its AVG13 difference from A in percentage points. Each pool is pretrained independently for 50 epochs. Scores are descriptive means over 13 targets and three fine-tuning seeds, without a significance claim.

Cumulative pool / new sources	Windows per epoch	AVG13 (%)	Δ from A (pp)
(a) LaBraM Base: 2 → 3 → 4 sources			
A: SCD + SEED-V	37,099	45.41	0.00
$A \cup B$: add BCIC-IV-2a	38,895	39.57	−5.83
$A \cup B \cup C$: add Workload	40,403	48.80	+3.39
(b) LaBraM Base: 1 → 2 → 6 sources			
A: TUAB	298,468	57.80	0.00
$A \cup B$: add HMC	389,716	56.19	−1.60
$A \cup B \cup C$: add Siena + SEED-VII + SEED-IV + PhysioNet	493,220	60.03	+2.23
(c) CBraMod: 3 → 6 → 8 sources			
A: BCIC-IV-2a + SCD + SEED-V	38,895	51.28	0.00
$A \cup B$: add FACED + Workload + TUSL	48,783	46.05	−5.23
$A \cup B \cup C$: add TUEV + SHU-MI	126,122	53.96	+2.67
(d) CBraMod: 2 → 3 → 4 sources			
A: TUAB + HMC	389,716	56.66	0.00
$A \cup B$: add TUEV	458,372	55.82	−0.84
$A \cup B \cup C$: add PhysioNet	472,988	58.18	+1.52
(e) BIOT: 3 → 4 → 6 sources			
A: BCIC-IV-2a + Workload + TUSL	3,452	54.08	0.00
$A \cup B$: add BCIC-2020/3	7,952	52.12	−1.96
$A \cup B \cup C$: add SEED-IV + SEED	38,276	54.69	+0.60

D.4 COMPLETE TASK-LEVEL DATA-VOLUME PLOTS

The following panels display every configuration on every target. They extend Figure 3 beyond its counts and selected examples, retaining both gains and losses relative to Full16. Numerical values and within-configuration seed variation are reported in Appendix C.

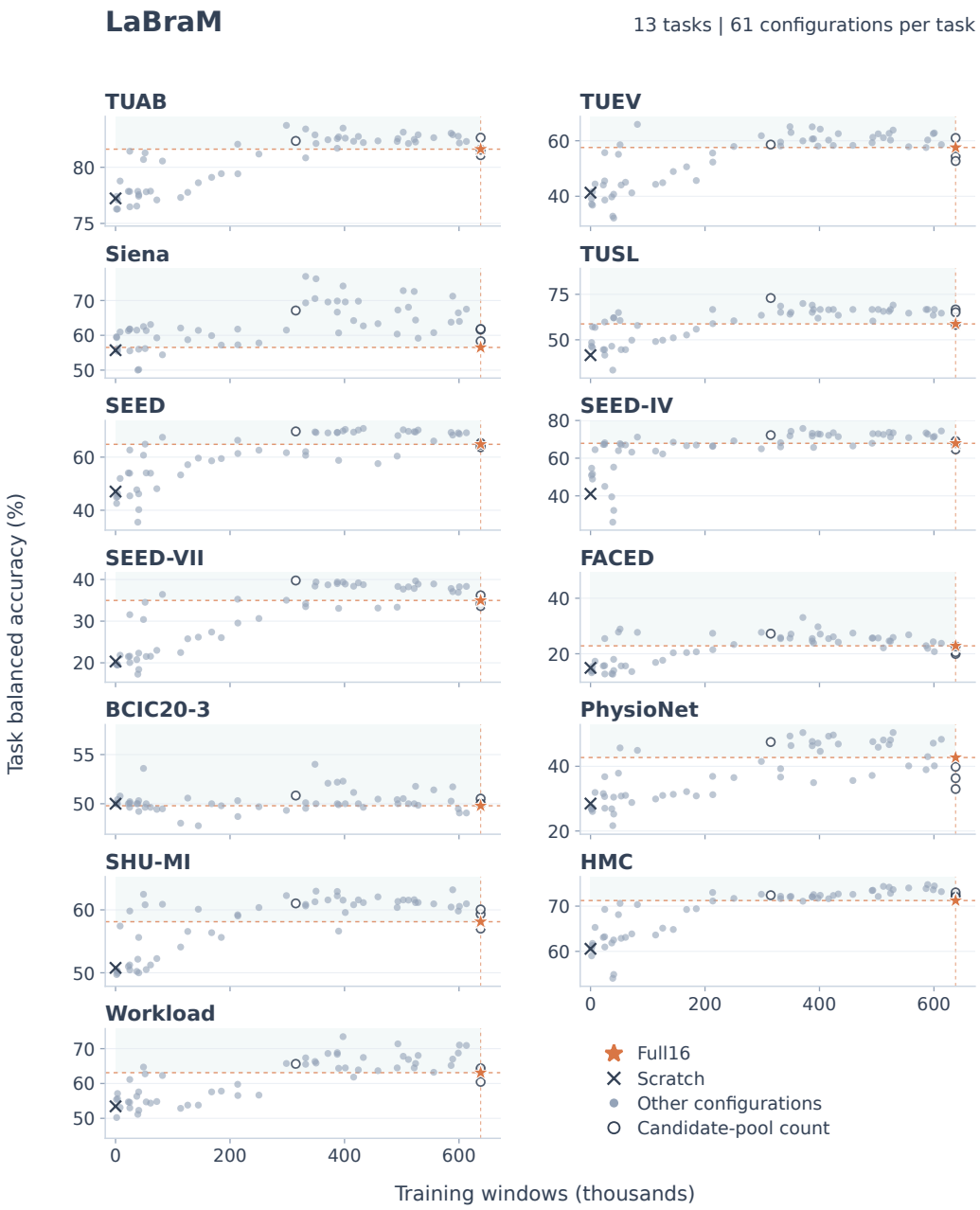


Figure 6: LaBraM: all 13 tasks and 61 configurations per task. Points are three-seed means. Stars and dashed lines mark Full16; crosses mark Scratch. Shading identifies fewer windows and a higher mean. Open circles use candidate-pool counts, which need not equal retained training windows.

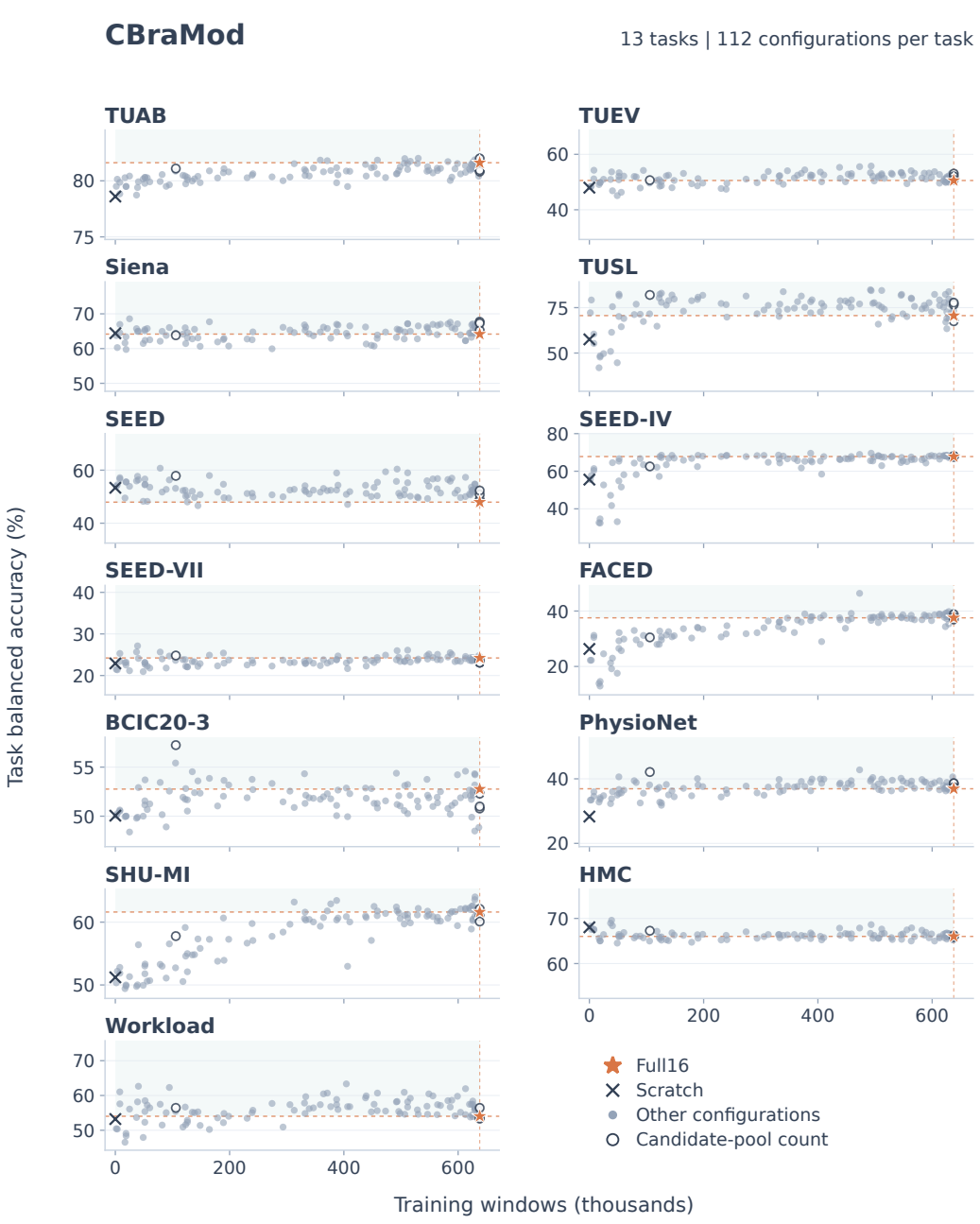


Figure 7: CBraMod: all 13 tasks and 112 configurations per task, using the same markers and exposure convention as the LaBraM panels. Means do not encode statistical significance.

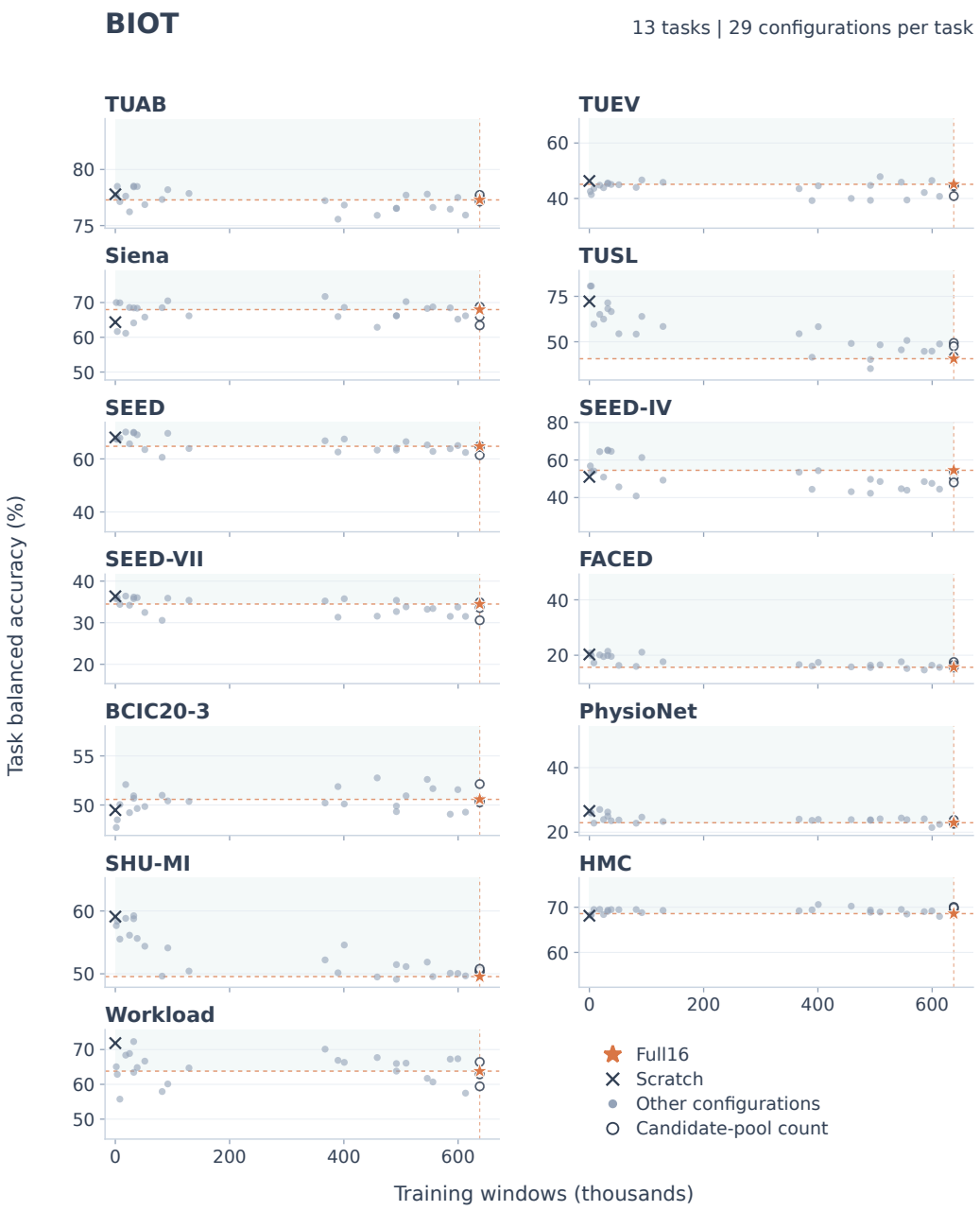


Figure 8: BIOT: all 13 tasks and 29 configurations per task, using the same markers and exposure convention as the LaBraM panels. Means do not encode statistical significance.

E SOURCE RANKINGS AND SELECTION RECIPES

This section provides the scores and source memberships needed to reconstruct the four methods in Section 6. Source orders are computed once; adding sources does not update their scores. The full task results for every evaluated mixture are in Appendix C.3.

E.1 PROGRESSIVE ORDERING

Progressive scores a source by mean validation balanced accuracy after pretraining on that source alone, averaging available fine-tuning seeds within each task and then averaging equally over 14 tasks: the 13 AVG13 targets plus BCIC-IV-2a. Table 24 lists the sources in descending score order. P-Hk takes the first k entries in its architecture's sequence; CBraMod and BIOT P-Lk take the last k. LaBraM P-L uses the explicit source sets in that table, which do not form one nested sequence across every reported source count.

Table 24: **Progressive source orders.** Sources follow the order used in the evaluated mixtures. P-H takes prefixes; CBraMod and BIOT P-L take suffixes. LaBraM P-L memberships appear in Table 25.

Order	LaBraM source	CBraMod source	BIOT source
1	TUAB	PhysioNet	SEED-IV
2	Siena	TUAB	BCIC-IV-2a
3	SEED-IV	Siena	SEED
4	SEED-VII	HMC	TUSL
5	SEED	BCIC20-3	BCIC20-3
6	PhysioNet	SEED-IV	Workload
7	FACED	SEED-VII	PhysioNet
8	SHU-MI	SEED	SEED-VII
9	HMC	SEED-V	SEED-V
10	BCIC20-3	SCD	SCD
11	TUEV	TUEV	SHU-MI
12	Workload	Workload	FACED
13	TUSL	TUSL	HMC
14	BCIC-IV-2a	BCIC-IV-2a	Siena
15	SEED-V	FACED	TUEV
16	SCD	SHU-MI	TUAB

Table 25: **LaBraM P-L source sets used for evaluation.** P-Hk for all three architectures takes the first k sources in Table 24; CBraMod and BIOT P-Lk take the last k sources in that table. LaBraM P-Lk uses the sets listed below.

Configuration	Sources
LaBraM P-L1	SCD
LaBraM P-L2	BCIC-IV-2a, TUSL
LaBraM P-L3	BCIC-IV-2a, TUSL, Workload
LaBraM P-L4	BCIC-IV-2a, TUSL, Workload, TUEV
LaBraM P-L5	SCD, SEED-V, BCIC-IV-2a, TUSL, Workload
LaBraM P-L6	BCIC-IV-2a, TUSL, Workload, TUEV, BCIC20-3, HMC
LaBraM P-L7	SCD, SEED-V, BCIC-IV-2a, TUSL, Workload, TUEV, BCIC20-3
LaBraM P-L8	BCIC-IV-2a, TUSL, Workload, TUEV, BCIC20-3, HMC, SHU-MI, FACED
LaBraM P-L9	SCD, SEED-V, BCIC-IV-2a, TUSL, Workload, TUEV, BCIC20-3, HMC, SHU-MI
LaBraM P-L10	BCIC-IV-2a, TUSL, Workload, TUEV, BCIC20-3, HMC, SHU-MI, FACED, PhysioNet, SEED
LaBraM P-L12	SCD, SEED-V, BCIC-IV-2a, TUSL, Workload, TUEV, BCIC20-3, HMC, SHU-MI, FACED, PhysioNet, SEED

E.2 PAIRWISE DATASET VALUATION

Let $\mathcal{D}$ contain the $N = 14$ sources other than SCD and SEED-V. Write $\bar{a}_s^q(S)$ for balanced accuracy after pretraining on source set S, fine-tuning on receiving dataset s, and evaluating split

$q \in \{\text{val}, \text{test}\}$. Each pair-versus-singleton difference uses the fine-tuning seeds shared by its two records. PDV fixes a contributing source d and averages its gains over the other receiving datasets:

$$v^q(d) = \frac{1}{N-1} \sum_{s \in \mathcal{D} \setminus \{d\}} \left[\bar{a}_s^q(\{d, s\}) - \bar{a}_s^q(\{s\}) \right], \qquad v(d) = v^{\text{val}}(d). \tag{1}$$

For example, TUAB's contribution to HMC compares TUAB+HMC with HMC-only pretraining, followed in both cases by HMC fine-tuning and validation. Each pair is evaluated on its two constituent tasks, yielding two directed gains. Fourteen singleton models and 91 pair models therefore supply 182 directed comparisons before seed repetitions. Scoring every pair on all 13 benchmark tasks is not required.

PDVk jointly pretrains on the k highest-scoring sources in Table 26. Its receiving-task average differs from the fixed AVG13 evaluation: each source contributes to the other 13 tasks, including BCIC-IV-2a when it is a receiver. The score measures directed transfer against a singleton baseline, rather than a non-additive interaction or a Shapley/Banzhaf value for one shared coalition utility (Ghorbani & Zou, 2019; Kwon & Zou, 2021; Wang & Jia, 2023). LaBraM PDV3 reuses the P-H3 checkpoint and evaluations; PDV2 reuses pair pretraining. The quadratic scoring-model count is not a measured wall-clock ratio relative to Progressive.

Table 26: **Complete PDV source rankings and validation scores.** $v(d)$ is the mean directed pair-versus-singleton gain in percentage points over the other 13 tasks in the common 14-source pool (Equation 1). Wins counts positive mean gains among these targets. The pool excludes SCD and SEED-V. PDVk takes the first k sources in each architecture-specific order. Scores use validation data.

Rank	LaBraM source	$v(d)$ (pp)	Wins	CBraMod source	$v(d)$ (pp)	Wins
1	TUAB	+1.1998	7/13	PhysioNet	+4.4563	10/13
2	SEED-IV	+0.0861	9/13	HMC	+3.7461	10/13
3	Siena	+0.0489	9/13	TUAB	+2.3519	7/13
4	BCIC-IV-2a	-0.0681	7/13	TUEV	+1.3957	9/13
5	FACED	-0.1486	6/13	Siena	+1.3946	8/13
6	SEED-VII	-0.1927	6/13	SEED-VII	+0.9433	8/13
7	HMC	-0.4332	6/13	BCIC20-3	+0.7038	6/13
8	SEED	-0.5381	4/13	SEED	+0.4090	7/13
9	PhysioNet	-0.9911	4/13	SEED-IV	+0.0215	6/13
10	BCIC20-3	-1.0424	5/13	BCIC-IV-2a	-0.1715	5/13
11	SHU-MI	-1.1070	4/13	SHU-MI	-0.5841	3/13
12	TUEV	-3.2095	4/13	TUSL	-0.8884	6/13
13	TUSL	-6.6974	3/13	Workload	-1.1762	4/13
14	Workload	-7.0535	3/13	FACED	-2.6573	2/13

Table 27: **PDV subset results.** Three-seed test AVG13 over 13 targets. Observed test maxima are descriptive. LaBraM PDV3 reuses the P-H3 checkpoint and evaluations.

k	LaBraM	CBraMod
2	60.57	56.47
3	60.85	56.26
4	61.06	58.18
5	—	56.04
6	61.21	57.21
8	60.75	56.97
10	60.47	56.58

E.3 SIZE ORDERING

Size uses one order for all three architectures. Ranking scores count each entire corpus, including its training, validation, and test partitions, with that source's own window length. Counts are rounded to thousands except TUSL's listed value of 245; ties retain the recorded order. Size-Hk/Size-Lk take the

first/last k sources in Table 28. These metadata scores construct the ranking; only training partitions are used for pretraining. Size-H4 contains TUAB, HMC, TUEV, and Siena. Actual training exposure is computed separately as described in Appendix B.

Table 28: **Size source order and window counts.** The ranking score counts the entire corpus using each source's own window length, then rounds to thousands; TUSL retains its listed count of 245. Tied scores retain the listed order. Training windows are counted separately and determine the exposure reported in the data-volume figures. Window lengths appear in Appendix B.1.

Rank	Source	Whole-corpus score (windows)	Training windows
1	TUAB	409,000	298,468
2	HMC	274,000	91,248
3	TUEV	111,000	68,656
4	Siena	51,000	33,605
5	SEED-VII	48,000	38,931
6	SCD	38,000	24,964
7	SEED-IV	34,000	16,352
8	SEED	24,000	13,972
9	PhysioNet	18,000	14,616
10	SEED-V	18,000	12,135
11	SHU-MI	12,000	8,683
12	FACED	10,000	8,232
13	BCIC20-3	5,000	4,500
14	BCIC-IV-2a	2,000	1,796
15	Workload	2,000	1,508
16	TUSL	245	148

E.4 Coverage ordering

For a source probe P_d, let $n_i(P_d)$ count assignments to code i by the fixed LaBraM tokenizer fitted on the 16-source training pool. Coverage is the fraction of its 8,192 entries activated at least twice:

$$c(d) = \frac{|\{i \in \{1, \ldots, 8192\} : n_i(P_d) \geq 2\}|}{8192}. \tag{2}$$

Table 29: **Cov-H/Cov-L source ordering and coverage scores.** The score is the fraction of 8,192 tokenizer codes occurring at least twice in the source probe. Cov-H takes the first k sources and Cov-L the last k. The order is shared across LaBraM and CBraMod. For scoring-pool size M_d, the recorded probe count is $96\lceil 0.008 M_d/96 \rceil$ windows. This scoring batch is distinct from the pretraining batch; counts differ across sources. Ties retain the listed order.

Rank	Dataset	Probe windows	Active codes	Coverage
1	TUAB	3,360	8,190	0.999756
2	Siena	480	7,563	0.923218
3	SEED-VII	384	7,215	0.880737
4	SEED-IV	288	7,060	0.861816
5	HMC	2,208	4,709	0.574829
6	PhysioNet	192	4,135	0.504761
7	SEED	192	3,740	0.456543
8	BCIC20-3	96	3,529	0.430786
9	SHU-MI	96	2,873	0.350708
10	TUEV	960	2,856	0.348633
11	FACED	96	2,751	0.335815
12	TUSL	96	2,259	0.275757
13	Workload	96	1,485	0.181274
14	BCIC-IV-2a	96	12	0.001465
15	SCD	384	1	0.000122
16	SEED-V	192	1	0.000122

Cov-Hk/Cov-Lk take the first/last k sources in this shared LaBraM/CBraMod order. Cov-H3 contains TUAB, Siena, and SEED-VII; Cov-H6 adds SEED-IV, HMC, and PhysioNet. Cov-L6 contains FACED, TUSL, Workload, BCIC-IV-2a, SCD, and SEED-V. Both six-source pools add SEED, BCIC20-3, SHU-MI, and TUEV at ten sources. Tied SCD and SEED-V scores retain their listed order. Reused pair and pretraining checkpoints are counted as shared observations.

After fitting the tokenizer, source scoring needs only forward passes over probes. The same order is used for CBraMod, whose pretraining objective does not use that tokenizer. Probe sizes differ, however, so coverage depends on both the source and its opportunities to activate codes; it is not a direct measure of physiological diversity. The peak configurations improve 12 of 13 LaBraM task means and 10 of 13 CBraMod means, as shown in Appendix C. Their paired statistical comparisons are in Appendix F. Neither Coverage nor Size needs downstream labels to construct its ranking; choosing a subset size still requires evaluation.

E.5 Complete expansion curves and ranking comparisons

Figure 9 includes every recorded Progressive source count and the low-ranked controls. Together with the complete task tables in Appendix C.3, it supplies the additional prefixes and controls omitted from the main-text plots. Untested source counts are not interpolated. LaBraM's P-L sets need not be nested, so its line is a sequence of evaluated configurations rather than one cumulative expansion path.

Table 30 recomputes all score ranks on their common 14-source pool. Progressive and PDV correlate at $\rho = 0.68$ for LaBraM and 0.84 for CBraMod; Coverage correlates with Progressive at 0.90 and 0.78, and with Size at 0.80. The rankings nevertheless differ within task families: TUEV ranks third by Size and tenth by Coverage, while Siena ranks fourth and second. These are descriptive agreements among source scores, not tests that one method causes better transfer or that coverage is independent of source size.

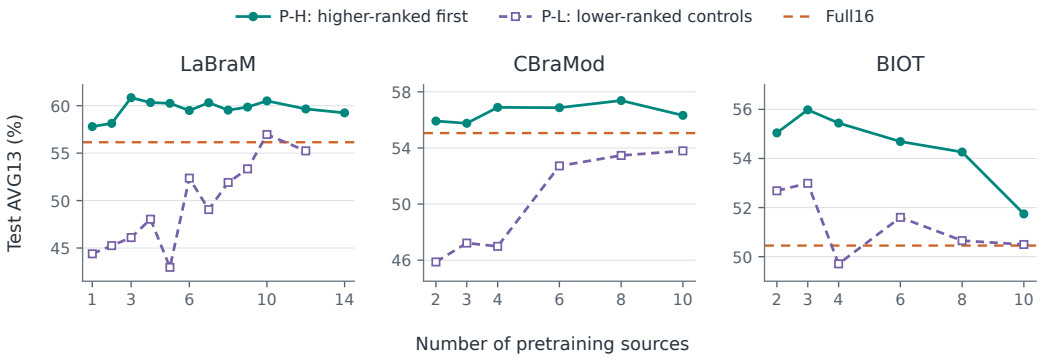


Figure 9: Complete Progressive source-count sweeps. Test AVG13 for all recorded P-H and P-L configurations; dashed lines mark Full16. Lines connect evaluated configurations and do not encode significance.

Table 30: **Spearman correlations among the four source scores.** All comparisons use the same 14 sources with available Progressive and PDV validation scores; SCD and SEED-V are excluded. Tied scores receive average ranks. Size and Coverage scores are shared across architectures.

Scoring signals	LaBraM	CBraMod
Progressive–PDV	0.68	0.84
Progressive–Size	0.67	0.60
Progressive–Coverage	0.90	0.78
PDV–Size	0.49	0.75
PDV–Coverage	0.61	0.67
Size–Coverage	0.80	

F STATISTICAL TESTS

For the main benchmark, paired differences are formed within each target and fine-tuning seed, then averaged over the three seeds within that target. The two-sided exact Wilcoxon signed-rank test uses the 13 target means, excludes exact zeros, gives average ranks to ties, and enumerates the sign distribution. An all-zero contrast has $p = 1$. Targets, rather than the 39 target–seed observations, are the units in this test. Related targets such as the SEED datasets limit extrapolation to an independent sample of all EEG tasks.

Holm correction is applied separately within each architecture and comparison family: Cov-H–Full16, Cov-H–Cov-L, Cov-H–Size-H, Cov-H–P-H, PDV–Full16, PDV–Cov-H, P-H–Full16, and PDV–P-H. Each Cov-H family has six contrasts. LaBraM has six in each PDV family; CBraMod has seven PDV–Full16 and six PDV–Cov-H contrasts because Cov-H5 is unreported. P-H–Full16 includes all twelve LaBraM prefixes and six each for CBraMod and BIOT. Matched PDV–P-H includes five LaBraM and six CBraMod contrasts; the identical LaBraM P-H3/PDV3 observation is excluded. Corrections are within these families, not across the whole study.

Pointwise 95% percentile intervals use 10,000 hierarchical bootstrap draws. Each draw samples 13 targets with replacement, then three paired seed indices with replacement within each sampled target. Both configurations use the same resampled indices. Deterministic comparison-specific random seeds allow reproduction. These intervals describe the mean difference, while the signed-rank test addresses a different distributional hypothesis. Neither procedure estimates variation across independently repeated pretraining runs.

F.1 SUBSET COMPARISONS WITH FULL16

Table 31: **Single-source performance selection versus Full16.** Differences are percentage points. p is the two-sided exact signed-rank test over 13 target means; p_H applies Holm within each architecture and comparison family. Intervals use 10,000 task/paired-seed bootstrap draws. The correction includes all 12 available LaBraM P-H prefixes and all six CBraMod or BIOT prefixes.

Architecture	Comparison	Δ	Wins	p	p_H	95% CI
LaBraM	P-H1–Full16	+1.65	9/13	0.080	0.161	[+0.01, +3.36]
LaBraM	P-H2–Full16	+1.99	8/13	0.340	0.340	[-0.80, +5.73]
LaBraM	P-H3–Full16	+4.70	13/13	< 0.001	0.003	[+2.98, +6.76]
LaBraM	P-H4–Full16	+4.18	13/13	< 0.001	0.003	[+2.83, +5.73]
LaBraM	P-H5–Full16	+4.10	13/13	< 0.001	0.003	[+2.29, +6.35]
LaBraM	P-H6–Full16	+3.34	12/13	0.001	0.005	[+1.82, +5.04]
LaBraM	P-H7–Full16	+4.17	13/13	< 0.001	0.003	[+2.30, +6.30]
LaBraM	P-H8–Full16	+3.38	12/13	< 0.001	0.003	[+2.21, +4.55]
LaBraM	P-H9–Full16	+3.70	13/13	< 0.001	0.003	[+2.50, +5.09]
LaBraM	P-H10–Full16	+4.36	13/13	< 0.001	0.003	[+2.88, +6.14]
LaBraM	P-H12–Full16	+3.51	12/13	< 0.001	0.003	[+2.10, +5.09]
LaBraM	P-H14–Full16	+3.10	10/13	0.010	0.031	[+1.16, +5.01]
CBraMod	P-H2–Full16	+0.86	8/13	0.340	0.732	[-0.56, +2.58]
CBraMod	P-H3–Full16	+0.70	6/13	0.635	0.732	[-0.57, +2.14]
CBraMod	P-H4–Full16	+1.83	9/13	0.068	0.341	[+0.34, +3.57]
CBraMod	P-H6–Full16	+1.81	8/13	0.244	0.732	[+0.02, +4.04]
CBraMod	P-H8–Full16	+2.32	9/13	0.048	0.287	[+0.40, +4.82]
CBraMod	P-H10–Full16	+1.26	9/13	0.080	0.341	[-0.03, +2.71]
BIOT	P-H2–Full16	+4.59	11/13	0.014	0.028	[+1.10, +8.96]
BIOT	P-H3–Full16	+5.53	12/13	0.005	0.019	[+1.62, +10.73]
BIOT	P-H4–Full16	+4.99	12/13	< 0.001	0.004	[+1.64, +9.58]
BIOT	P-H6–Full16	+4.24	11/13	0.003	0.017	[+1.15, +8.75]
BIOT	P-H8–Full16	+3.81	11/13	0.008	0.024	[+0.99, +7.81]
BIOT	P-H10–Full16	+1.29	9/13	0.273	0.273	[-0.92, +4.67]

Table 32: **Cov-H–Full paired comparisons.** Differences are percentage points. p is the two-sided exact signed-rank test; p_H applies Holm within architecture and comparison family. Intervals resample targets and paired fine-tuning seeds.

Architecture	Prefix	Δ	Wins	p	p_H	95% CI
LaBraM	Cov-H2	+1.98	9/13	0.244	0.244	[-0.40, +4.78]
LaBraM	Cov-H3	+5.58	12/13	< 0.001	0.002	[+3.39, +8.03]
LaBraM	Cov-H4	+4.69	13/13	< 0.001	0.001	[+2.79, +6.91]
LaBraM	Cov-H6	+3.88	13/13	< 0.001	0.001	[+2.46, +5.54]
LaBraM	Cov-H8	+3.89	12/13	0.001	0.002	[+2.31, +5.64]
LaBraM	Cov-H10	+3.91	12/13	< 0.001	0.002	[+2.05, +6.23]
CBraMod	Cov-H2	+0.45	6/13	0.946	0.946	[-0.76, +1.80]
CBraMod	Cov-H3	+1.50	9/13	0.191	0.471	[-0.69, +3.91]
CBraMod	Cov-H4	+2.05	9/13	0.094	0.471	[+0.15, +4.27]
CBraMod	Cov-H6	+3.30	10/13	0.010	0.063	[+1.00, +6.00]
CBraMod	Cov-H8	+1.78	9/13	0.110	0.471	[-0.28, +4.00]
CBraMod	Cov-H10	+1.40	10/13	0.127	0.471	[-0.20, +3.18]

Table 33: **PDV-Full paired comparisons.** Differences are percentage points. p is the two-sided exact signed-rank test; p_H applies Holm within architecture and comparison family. Intervals resample targets and paired fine-tuning seeds.

Architecture	Prefix	Δ	Wins	p	p_H	95% CI
LaBraM	PDV2	+4.42	13/13	< 0.001	0.001	[+2.54, +6.82]
LaBraM	PDV3	+4.70	13/13	< 0.001	0.001	[+2.98, +6.74]
LaBraM	PDV4	+4.92	13/13	< 0.001	0.001	[+2.77, +7.90]
LaBraM	PDV6	+5.06	13/13	< 0.001	0.001	[+2.89, +7.86]
LaBraM	PDV8	+4.60	13/13	< 0.001	0.001	[+2.74, +7.00]
LaBraM	PDV10	+4.33	13/13	< 0.001	0.001	[+2.51, +6.70]
CBraMod	PDV2	+1.41	8/13	0.414	0.509	[-1.36, +4.38]
CBraMod	PDV3	+1.20	9/13	0.216	0.509	[-0.29, +3.01]
CBraMod	PDV4	+3.12	10/13	0.017	0.120	[+1.09, +5.42]
CBraMod	PDV5	+0.98	9/13	0.146	0.509	[-0.12, +2.34]
CBraMod	PDV6	+2.15	10/13	0.068	0.409	[+0.15, +4.49]
CBraMod	PDV8	+1.91	8/13	0.094	0.471	[+0.10, +3.92]
CBraMod	PDV10	+1.52	9/13	0.127	0.509	[-0.06, +3.32]

Ten of twelve LaBraM Progressive prefixes, five of six Cov-H prefixes, and all six PDV prefixes have significant positive differences after their respective corrections. Counting P-H3/PDV3 once gives the 20 of 23 configurations reported in Section 4.1. Five of six BIOT prefixes pass correction; none of the 19 CBraMod prefixes does. These counts concern the specified comparisons, not all 202 configurations.

F.2 COMPARISONS BETWEEN SELECTION METHODS AND RANKING DIRECTIONS

Table 34: **PDV versus single-source performance selection at matched source counts.** Differences are percentage points. p is the two-sided exact signed-rank test over 13 target means; p_H applies Holm within each architecture and comparison family. Intervals use 10,000 task/paired-seed bootstrap draws. The families contain five LaBraM and six CBraMod comparisons. LaBraM PDV3 and P-H3 share the same checkpoint and evaluations and are excluded from this test family.

Architecture	Comparison	Δ	Wins	p	p_H	95% CI
LaBraM	PDV2–P-H2	+2.43	11/13	0.048	0.239	[-0.45, +5.12]
LaBraM	PDV4–P-H4	+0.73	7/13	0.735	0.735	[-0.67, +2.54]
LaBraM	PDV6–P-H6	+1.72	10/13	0.244	0.671	[-0.51, +4.31]
LaBraM	PDV8–P-H8	+1.21	8/13	0.216	0.671	[-0.11, +3.06]
LaBraM	PDV10–P-H10	-0.03	5/13	0.168	0.671	[-1.95, +2.70]
CBraMod	PDV2–P-H2	+0.55	7/13	0.685	1.000	[-1.73, +2.92]
CBraMod	PDV3–P-H3	+0.51	7/13	0.839	1.000	[-0.63, +1.98]
CBraMod	PDV4–P-H4	+1.29	9/13	0.244	1.000	[-0.73, +3.37]
CBraMod	PDV6–P-H6	+0.34	8/13	0.455	1.000	[-1.04, +1.51]
CBraMod	PDV8–P-H8	-0.41	5/13	0.455	1.000	[-1.59, +0.70]
CBraMod	PDV10–P-H10	+0.25	7/13	0.735	1.000	[-0.71, +1.34]

Table 35: **Cov-H–Size-H paired comparisons.** Differences are percentage points. p is the two-sided exact signed-rank test; p_H applies Holm within architecture and comparison family. Intervals resample targets and paired fine-tuning seeds.

Architecture	Prefix	Δ	Wins	p	p_H	95% CI
LaBraM	Cov-H2	+1.93	10/13	0.021	0.086	[+0.43, +3.68]
LaBraM	Cov-H3	+4.82	11/13	0.002	0.015	[+2.33, +7.53]
LaBraM	Cov-H4	+3.78	9/13	0.010	0.052	[+1.73, +5.93]
LaBraM	Cov-H6	+1.69	7/13	0.273	0.649	[-0.29, +3.78]
LaBraM	Cov-H8	+1.33	8/13	0.216	0.649	[-0.09, +3.14]
LaBraM	Cov-H10	-0.04	7/13	0.893	0.893	[-1.61, +1.64]
CBraMod	Cov-H2	-1.15	2/13	0.040	0.239	[-3.01, +0.09]
CBraMod	Cov-H3	+0.74	9/13	0.305	0.916	[-1.02, +2.57]
CBraMod	Cov-H4	-0.26	7/13	0.946	1.000	[-1.86, +1.20]
CBraMod	Cov-H6	+2.32	9/13	0.068	0.341	[+0.35, +4.96]
CBraMod	Cov-H8	+0.12	7/13	0.569	1.000	[-1.31, +1.41]
CBraMod	Cov-H10	-0.48	3/13	0.191	0.764	[-1.44, +0.56]

Table 36: **Cov-H–P-H paired comparisons.** Differences are percentage points. p is the two-sided exact signed-rank test; p_H applies Holm within architecture and comparison family. Intervals resample targets and paired fine-tuning seeds.

Architecture	Prefix	Δ	Wins	p	p_H	95% CI
LaBraM	Cov-H2	-0.01	8/13	0.588	1.000	[-1.72, +1.39]
LaBraM	Cov-H3	+0.88	7/13	0.588	1.000	[-0.95, +2.80]
LaBraM	Cov-H4	+0.51	6/13	0.787	1.000	[-0.59, +1.74]
LaBraM	Cov-H6	+0.54	8/13	0.588	1.000	[-1.33, +2.61]
LaBraM	Cov-H8	+0.51	7/13	0.542	1.000	[-0.51, +1.67]
LaBraM	Cov-H10	-0.44	5/13	0.376	1.000	[-2.62, +2.23]
CBraMod	Cov-H2	-0.41	6/13	0.787	1.000	[-1.81, +0.78]
CBraMod	Cov-H3	+0.80	8/13	0.305	1.000	[-0.87, +2.65]
CBraMod	Cov-H4	+0.22	6/13	0.787	1.000	[-1.17, +1.70]
CBraMod	Cov-H6	+1.49	10/13	0.040	0.239	[+0.21, +2.90]
CBraMod	Cov-H8	-0.53	8/13	0.946	1.000	[-2.62, +1.24]
CBraMod	Cov-H10	+0.14	5/13	0.839	1.000	[-0.68, +1.07]

Table 37: **PDV–Cov-H paired comparisons.** Differences are percentage points. p is the two-sided exact signed-rank test; p_H applies Holm within architecture and comparison family. Intervals resample targets and paired fine-tuning seeds.

Architecture	Prefix	Δ	Wins	p	p_H	95% CI
LaBraM	PDV2	+2.44	10/13	0.068	0.409	[+0.29, +4.54]
LaBraM	PDV3	-0.88	6/13	0.588	1.000	[-2.79, +0.94]
LaBraM	PDV4	+0.22	9/13	0.470	1.000	[-1.09, +1.62]
LaBraM	PDV6	+1.18	9/13	0.127	0.636	[-0.21, +2.65]
LaBraM	PDV8	+0.70	10/13	0.168	0.671	[-0.42, +1.85]
LaBraM	PDV10	+0.41	7/13	0.497	1.000	[-0.69, +1.53]
CBraMod	PDV2	+0.96	8/13	0.588	1.000	[-1.34, +3.55]
CBraMod	PDV3	-0.30	7/13	1.000	1.000	[-2.38, +1.21]
CBraMod	PDV4	+1.07	8/13	0.305	1.000	[-0.65, +2.90]
CBraMod	PDV6	-1.15	5/13	0.110	0.659	[-2.53, +0.23]
CBraMod	PDV8	+0.12	5/13	0.787	1.000	[-1.15, +1.56]
CBraMod	PDV10	+0.12	6/13	0.787	1.000	[-0.72, +0.96]

Table 38: **Cov-H–Cov-L paired comparisons.** Differences are percentage points. p is the two-sided exact signed-rank test; p_H applies Holm within architecture and comparison family. Intervals resample targets and paired fine-tuning seeds.

Architecture	Prefix	Δ	Wins	p	p_H	95% CI
LaBraM	Cov-H2	+12.72	13/13	< 0.001	0.001	[+8.85, +16.79]
LaBraM	Cov-H3	+22.15	13/13	< 0.001	0.001	[+15.41, +29.08]
LaBraM	Cov-H4	+12.04	12/13	< 0.001	0.001	[+8.19, +15.98]
LaBraM	Cov-H6	+3.65	9/13	0.021	0.021	[+0.99, +6.25]
LaBraM	Cov-H8	+9.96	12/13	< 0.001	0.001	[+7.03, +12.68]
LaBraM	Cov-H10	+8.33	13/13	< 0.001	0.001	[+5.97, +10.67]
CBraMod	Cov-H2	+5.44	9/13	0.110	0.471	[+1.32, +9.89]
CBraMod	Cov-H3	+5.28	9/13	0.127	0.471	[+0.80, +10.21]
CBraMod	Cov-H4	+2.37	7/13	0.244	0.471	[-0.17, +5.25]
CBraMod	Cov-H6	+12.31	13/13	< 0.001	0.001	[+6.43, +19.33]
CBraMod	Cov-H8	+2.89	10/13	0.094	0.471	[-0.27, +5.66]
CBraMod	Cov-H10	+1.80	9/13	0.146	0.471	[-0.80, +4.28]

The matched PDV–Progressive comparisons do not establish an additional AVG13 benefit after correction. Neither architecture has an adjusted Cov-H–P-H or PDV–Cov-H contrast below 0.05. These results do not establish equivalence or a uniformly strongest selector. High/low comparisons fix source count but not training-window exposure.

F.3 Capacity and external-target comparisons

Table 39: **LaBraM Large paired comparison.** Differences are percentage points. The two-sided exact Wilcoxon test uses the 13 paired task means. Holm adjustment uses the full family of three evaluated Large contrasts. The pointwise 95% interval uses 10,000 hierarchical task/paired-seed bootstrap resamples (Appendix F).

Comparison	Δ	Task wins	p	p_H	95% CI
PDV6–Full16	+2.84	11/13	0.00171	0.00513	[+1.21, +4.76]

The LaBraM Large six-source–Full16 comparison uses the same 13-target procedure, retaining the full correction family specified in Table 39. Its Holm-adjusted $p = 0.00513$ tests the within-Large contrast, not the Base–Large interaction. The six-target external panel uses separate within-architecture families and is reported with its complete results in Appendix G.

The unresolved near-chance tasks in Appendix I are not added to AVG13 without complete run-level evidence; no complete AVG16 sensitivity result is asserted. Task-win counts and the descriptive rankings are not significance tests for individual tasks or guarantees for new source pools.

G Transfer to Pool-External Tasks

G.1 Evaluation protocol

Benchmark targets absent from a mixture may still have informed validation-based selection. The external panel instead comprises EEGMAT, CHB-MIT, ISRUC-S3, Mumtaz2016, BI2014a, and Lee2019-SSVEP. All six are absent from the 16-source pool, LaBraM tokenizer training, and selector construction. We evaluate the configurations with the highest observed test AVG13 within each selection method: LaBraM Cov-H3/PDV6 and CBraMod Cov-H6/PDV4. Each mixture is fixed across external targets. Target training partitions are used for fine-tuning, validation partitions for checkpoint selection, and test partitions for reporting; target labels do not redefine the source ranking.

Each architecture has three configurations, including Full16, evaluated on six targets with seeds $\{0, 1, 2\}$: $6 \times 3 \times 3 = 54$ evaluations per architecture and 108 in total. AVG6 first averages the three seed scores within each target, then weights the six targets equally. Table 41 reports every target mean and sample SD. Three fine-tuning seeds do not constitute independent pretrained models.

G.2 Transfer performance

All four configurations improve every external-task mean (Figure 10). AVG6 rises from 77.68% to 81.61% and 80.31% for LaBraM, and from 71.39% to 76.55% and 76.44% for CBraMod, for Coverage and PDV respectively (Table 40).

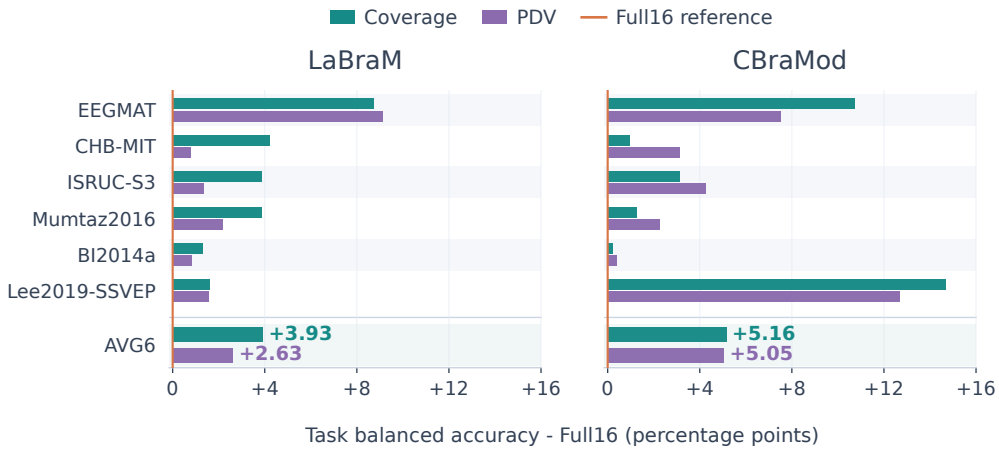


Figure 10: **Fixed subsets improve all six external-task means.** Bars show three-seed mean differences in balanced accuracy from Full16. LaBraM uses Cov-H3/PDV6 and CBraMod uses Cov-H6/PDV4, fixed before external fine-tuning. AVG6 weights targets equally. No significance is encoded.

The consistent direction masks uneven gains. CBraMod Cov-H6 gains 14.67 points on Lee2019-SSVEP and 10.71 on EEGMAT, but only 0.22 on BI2014a. LaBraM's largest improvements are on EEGMAT, with Cov-H3 gaining more there than on BI2014a. CBraMod PDV4 exceeds Cov-H6 on four of six targets despite its slightly lower AVG6. Thus, even outside the original benchmark, a higher average need not imply a preferable recipe for most targets.

G.3 Statistical comparison and scope

To assess uncertainty in these gains, we apply the two-sided exact Wilcoxon signed-rank test to the six target-level mean differences, using the same zero and tie rules as Appendix F. Holm correction

Table 40: **Pool-external adaptation of fixed mixtures.** Test balanced accuracy (%). AVG6 averages three fine-tuning seeds within each target, then the six targets equally. Configurations are fixed across targets. Per-target scores and paired comparisons are in Appendix G.

Architecture	Coverage ordering	PDV	Full16
CBraMod	Cov-H6: 76.55	PDV4: 76.44	71.39
LaBraM	Cov-H3: 81.61	PDV6: 80.31	77.68

Table 41: **Per-target pool-external results.** Test balanced accuracy (%), mean ± sample SD over three fine-tuning seeds. AVG6 equally weights the six target means.

	CBraMod			LaBraM		
Target	Cov-H6	PDV4	Full16	Cov-H3	PDV6	Full16
EEGMAT	75.79 ± 4.96	72.62 ± 5.95	65.08 ± 1.82	82.54 ± 3.83	82.94 ± 0.69	73.81 ± 1.19
CHB-MIT	79.35 ± 4.37	81.53 ± 1.46	78.39 ± 0.93	86.07 ± 1.01	82.63 ± 0.87	81.85 ± 2.08
ISRUC-S3	68.74 ± 1.41	69.88 ± 1.20	65.61 ± 0.48	71.64 ± 1.10	69.11 ± 2.01	67.77 ± 1.49
Mumtaz2016	88.96 ± 0.80	89.99 ± 0.80	87.71 ± 2.32	93.41 ± 0.10	91.73 ± 0.48	89.55 ± 1.04
BI2014a	62.19 ± 0.40	62.35 ± 0.47	61.97 ± 0.50	64.59 ± 0.54	64.10 ± 0.26	63.28 ± 0.66
Lee2019-SSVEP	84.24 ± 1.93	82.27 ± 1.73	69.58 ± 1.18	91.39 ± 0.76	91.36 ± 0.57	89.79 ± 0.29
AVG6	76.55	76.44	71.39	81.61	80.31	77.68

covers the two selected-versus-Full16 contrasts within each architecture, separately from the main source-count sweeps. Each contrast has $p = 0.03125$ and $p_H = 0.0625$; none meets the corrected 0.05 threshold. Pointwise 95% percentile intervals use 10,000 hierarchical bootstrap resamples of six targets and three paired seed indices within each sampled target; seeds are resampled jointly across configurations. These intervals describe mean differences, whereas the signed-rank test addresses a different distributional hypothesis. Neither analysis estimates variation across independently pretrained models. Table 42 reports the contrasts without treating the 18 target–seed evaluations as independent tasks.

Table 42: **Pool-external paired comparisons.** Differences are percentage points. p is the two-sided exact task-level signed-rank test; p_H applies Holm to the two selected-versus-Full16 contrasts within each architecture. Pointwise 95% intervals use 10,000 hierarchical target/paired-seed bootstrap resamples.

Architecture	Comparison	Δ	Wins	p	p_H	95% CI
CBraMod	Cov-H6–Full16	+5.16	6/6	0.03125	0.0625	[+1.05, +9.92]
CBraMod	PDV4–Full16	+5.05	6/6	0.03125	0.0625	[+1.99, +8.67]
LaBraM	Cov-H3–Full16	+3.93	6/6	0.03125	0.0625	[+2.12, +6.33]
LaBraM	PDV6–Full16	+2.63	6/6	0.03125	0.0625	[+0.71, +5.45]

This panel evaluates adaptation of fixed mixtures after target fine-tuning. It does not evaluate zero-shot performance, prove equivalence, or test selection among newly assembled candidate sources. Transfer to a new source pool would require recomputing rankings and pretraining the newly selected mixtures; that experiment has not been completed.

H Complete Configuration Rankings

The following plots rank all 202 configurations in the main overview by test AVG13: 61 LaBraM, 112 CBraMod, and 29 BIOT configurations. Each score averages three fine-tuning seeds within each of the same 13 targets and then weights the targets equally. Ranks use the unrounded means; they summarize the evaluated configurations and do not indicate statistical significance. Full16 and Scratch remain in each inventory. The shared LaBraM P-H3/PDV3 checkpoint appears once. Random-k denotes a k-source subset, and its following number identifies the sampled composition. "Without" names the omitted source in a source-removal configuration. Horizontal segments connect each score to Full16; colors distinguish higher and lower means, with Full16 and Scratch highlighted separately.

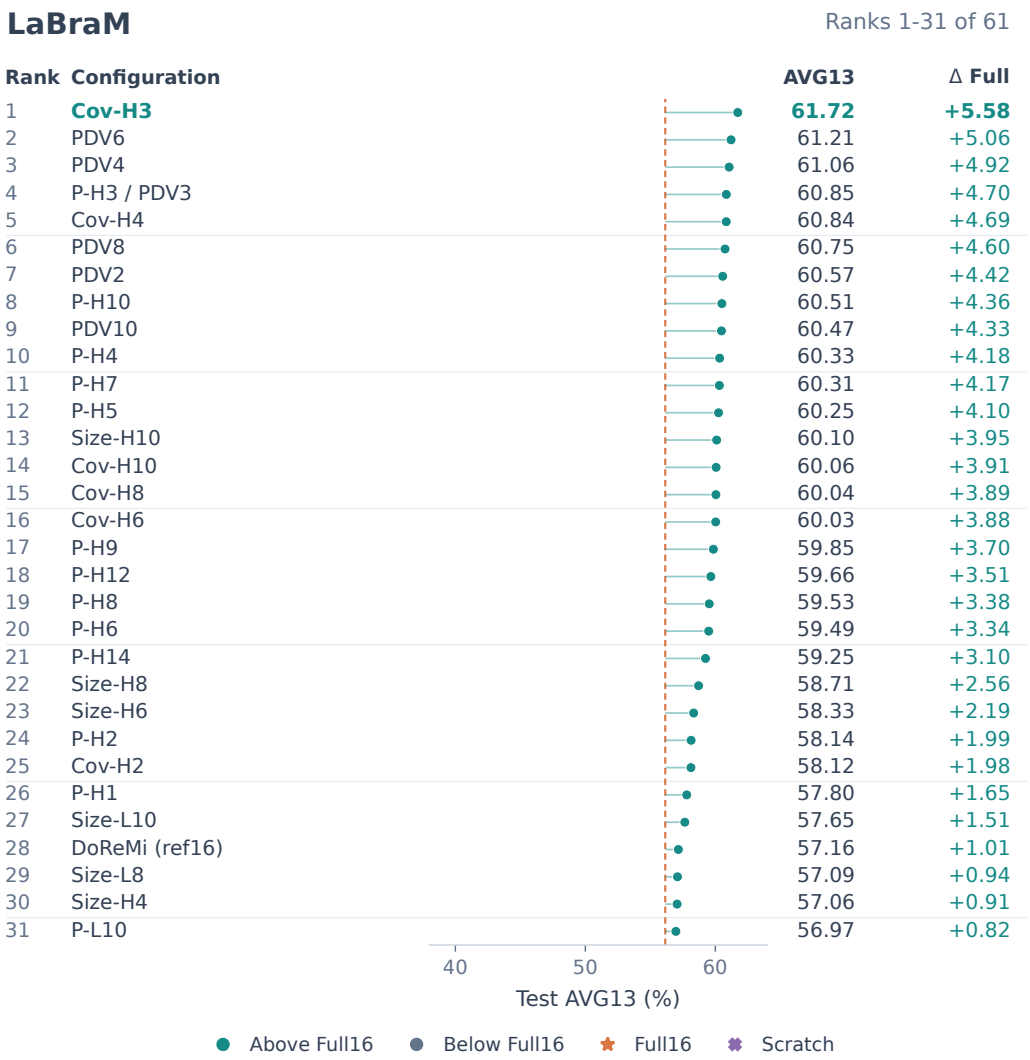


Figure 11: **LaBraM configuration ranking, ranks 1–31 of 61.** Scores are test AVG13 (%); Δ Full is the difference from the architecture-matched Full16 in percentage points.

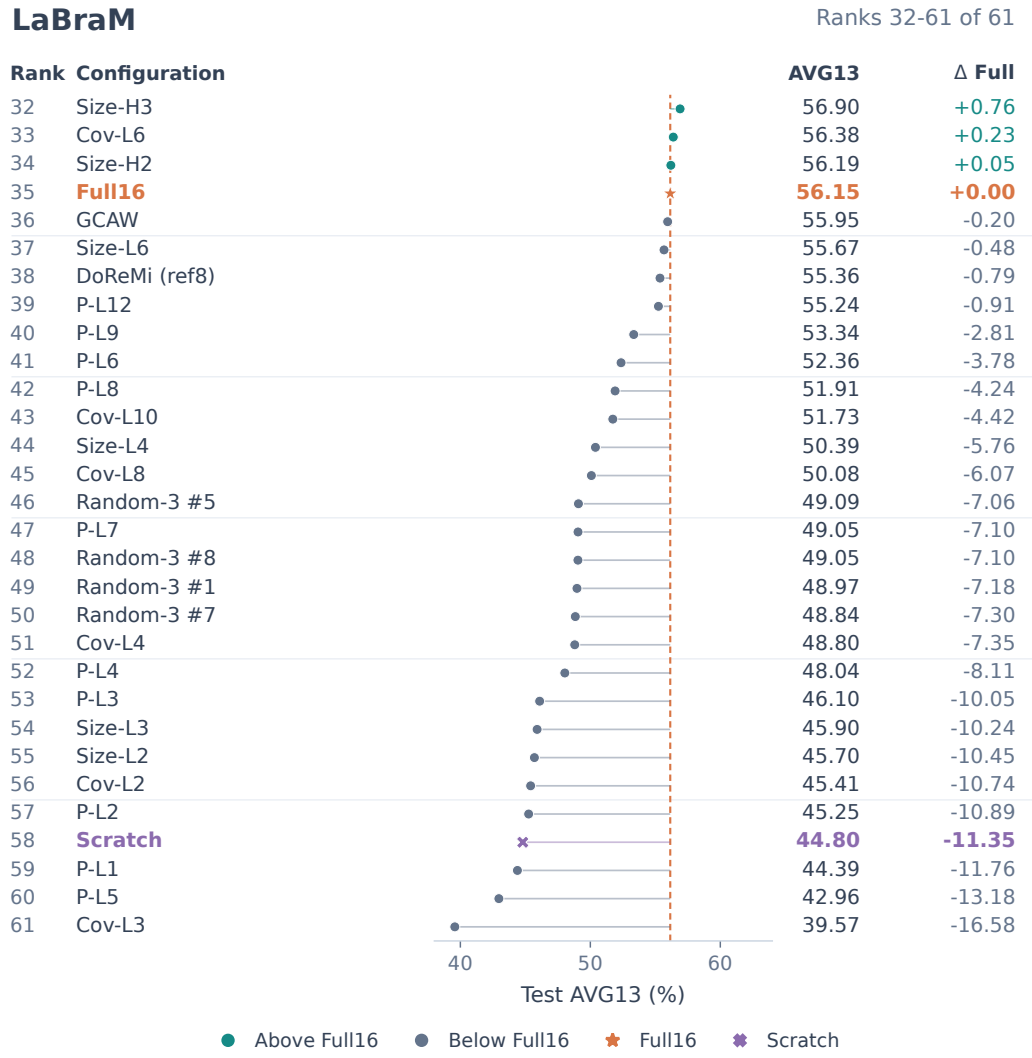


Figure 12: **LaBraM configuration ranking, ranks 32–61 of 61.** Scores are test AVG13 (%); Δ Full is the difference from the architecture-matched Full16 in percentage points.



Figure 13: **CBraMod configuration ranking, ranks 1–38 of 112.** Scores are test AVG13 (%); Δ Full is the difference from the architecture-matched Full16 in percentage points.



Figure 14: **CBraMod configuration ranking, ranks 39–76 of 112.** Scores are test AVG13 (%); Δ Full is the difference from the architecture-matched Full16 in percentage points.

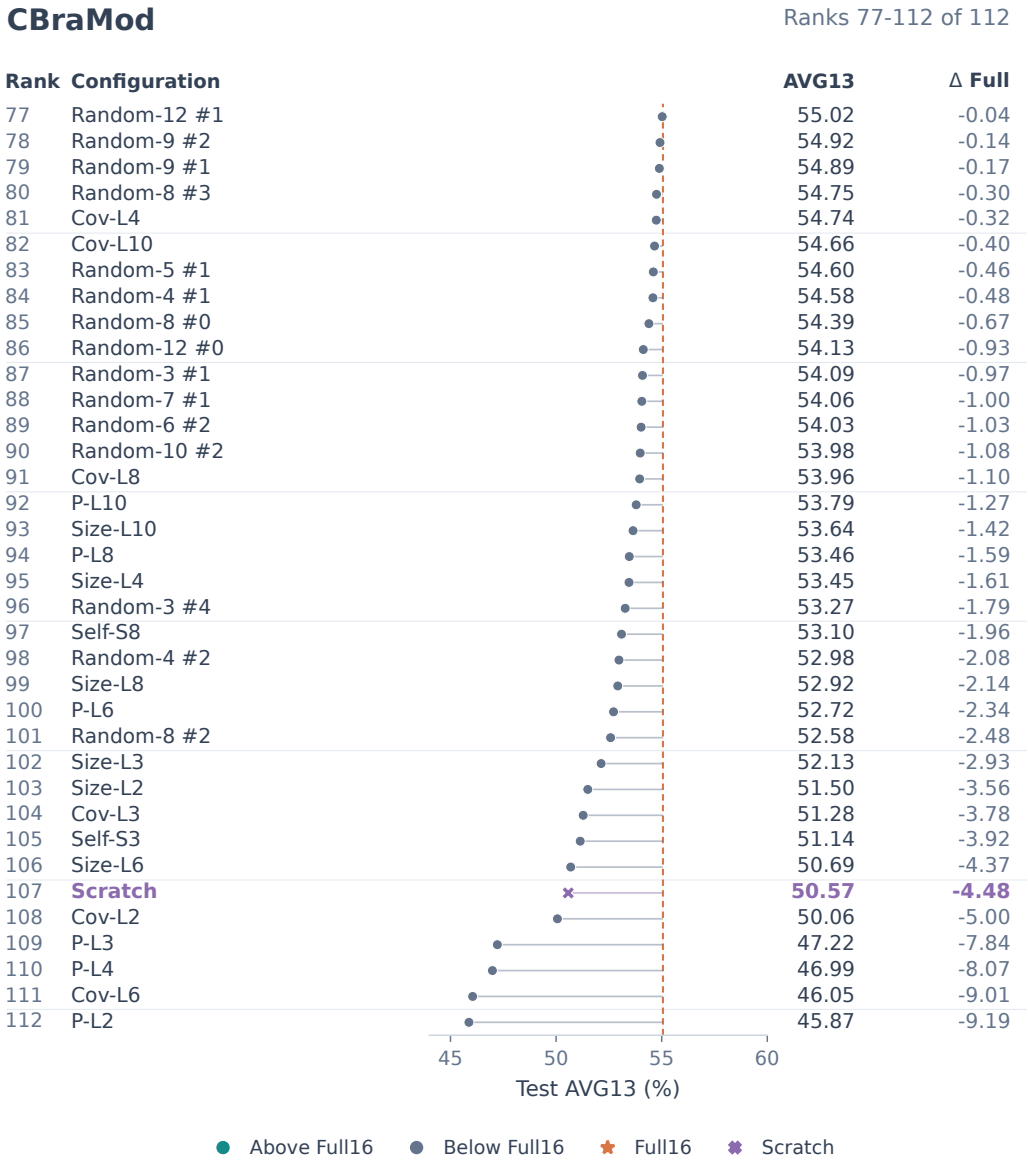


Figure 15: **CBraMod configuration ranking, ranks 77–112 of 112.** Scores are test AVG13 (%); Δ Full is the difference from the architecture-matched Full16 in percentage points.

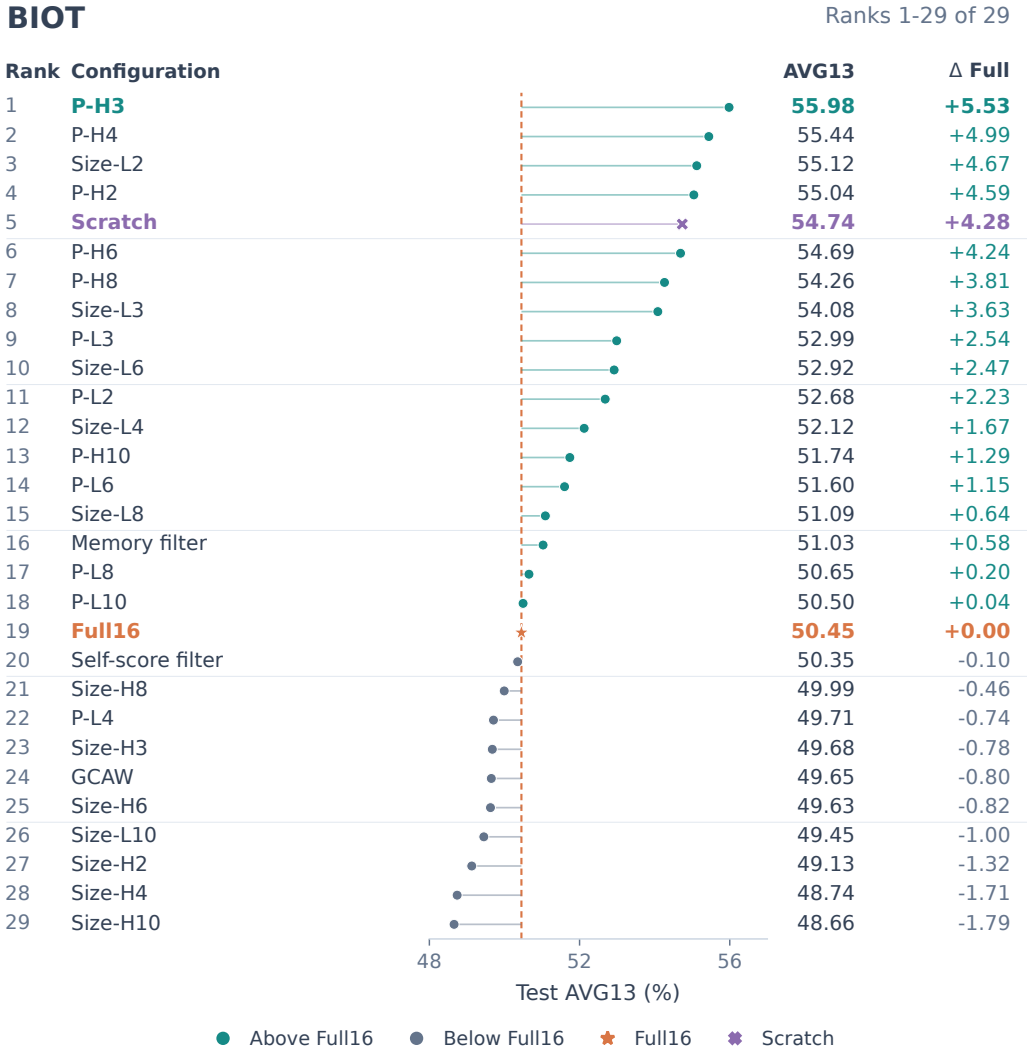


Figure 16: **BIOT configuration ranking, ranks 1–29 of 29.** Scores are test AVG13 (%); Δ Full is the difference from the architecture-matched Full16 in percentage points.

I EXCLUDED TASK DIAGNOSTICS

SEED-V, SCD, and BCIC-IV-2a remain close to chance under the current downstream implementation, including scratch. The reported ranges are 20.0–20.5% for five-class SEED-V, 48.5–50.5% for binary SCD, and 24.5–25.5% for four-class BCIC-IV-2a. We have not yet reproduced reliable above-chance performance for these tasks under this pipeline. This is an unresolved evaluation limitation, not evidence that the tasks are inherently too difficult or that their source recordings cannot be useful for pretraining.

Their training partitions remain in the candidate pretraining pool; their downstream scores are excluded from AVG13. They are not silently treated as successful downstream evaluations. No complete AVG16 sensitivity result is asserted (Appendix F). The task count in AVG13 and the source count in the PDV comparison refer to different sets.

Table 43: **Selected BIOT task-level declines under Full-16.** Entries show Full-16 minus scratch in percentage points for five tasks with notable declines. Task types use the corresponding dataset definitions.

Task	Task type	Δ
TUSL	Clinical slowing	-31.6
SHU-MI	Motor imagery	-9.6
Workload	Cognitive load	-8.0
FACED	Emotion recognition	-4.6
PhysioNet	Motor imagery	-3.6