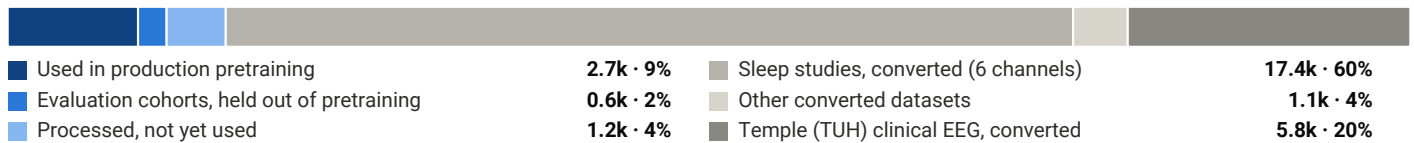


Pretraining progress

Production EEG encoder (internal name d10) · status as of Oct 30, 2026

29.4M parameters 512-wide, 8-layer transformer that reads up to 64 EEG channels	2.7k h pretraining data 17 datasets, 5.5k people, no labels used	131k h training exposure 2,525 unique hours × 52 passes, about 21 hours on 4 GPUs	45 assessm ents evaluated 30 of 45 reach AUC 0.70 or higher and 15 reach 0.85 or higher (only 2 of those with 30+ test patients)
---	---	---	---

Where our hours are



28.8k unique hours of EEG are on hand. The production model has been pretrained on 2.7k of them (9%). The rest is processed or converted and available for the next pretraining run.

How the model is trained and checked

- **Self-supervised pretraining.** The encoder reads EEG and learns to predict the next 160 ms of signal, plus 171 standard signal features. No diagnoses or other labels are used.
- **Checking it on an assessment.** The encoder is frozen: each recording becomes a vector of numbers, and a light classifier (TabPFN or XGBoost) is trained on labeled patients and scored on held-out patients.
- **Fixed, audited test sets.** Test cohorts stay the same across model versions, every split is checked for patient overlap (76 of 77 clean; the one leaky split, TBI, was retired), and every model gets the same protocol.
- **Against standard EEG features.** Each assessment is also scored with 28 standard EEG features. On the 17 assessments compared head to head, deep learning features lead on 5, EEG features lead on 4 and 8 are ties (a lead is 0.03 AUC or more; mean difference +0.01 AUC).

Does more pretraining data help? Checks so far

What we changed	Result	How to read it
Larger corpus across model generations: 665 h (v6_r3) to 3,261 h (d10), 4.9×	Mean AUC change −0.03 (95% CI −0.07 to +0.01) over 19 classification assessments; 9 up, 10 down.	No measurable gain. Test sets are small (median 20 patients), so the interval is wide.
Controlled sweep on the d10 data: 10%, 25%, 50% and 100% of the pretraining hours (253 to 2,525 h), same 3,000 training steps	Validation loss 0.853, 0.866, 0.841, 0.846.	Flat within about ±0.01. One seed; the passes over the data fall as hours rise; short runs.
Add Temple clinical EEG to pretraining: 1.65× and 2.8× the hours (3,000-step runs), 31 assessments	AUC change vs the d10-only run: −0.010 (CI −0.043 to +0.023) and +0.002 (CI −0.037 to +0.043). Two d10-only runs differ by −0.030. Abnormal-EEG and seizure assessments are also flat.	No effect beyond run-to-run noise. These short runs score well below the production model (for example 0.70 vs 0.95 AUC on an Alzheimer's assessment), so they show direction, not a ceiling.

So far, more pretraining data has not improved downstream accuracy in the short tests we have run. A full-length test is next.

What we are doing next

- **Validation loss versus samples seen:** one model on all the data we have, on the full training schedule, with validation loss recorded throughout; compared with models trained on half and a tenth of the data.
- **Then pretrain on the full Temple set** and re-run the frozen-feature evaluation (XGBoost and TabPFN) on the same 45 assessments.
- **In parallel:** a cross-dataset preprocessing audit (reference, units, normalization).

Hours are unique EEG time from file headers (no augmented copies or overlapping windows). Training exposure = unique training hours × passes over them. AUC and the assessments are defined in the slide “How to read these slides”. Sources: Oct 2026 board figures; Temple screening results, Oct 28 2026.