

# Structural-Functional Integration for Enhanced Detection of Subtle Epileptogenic Lesions: A Framework for a Multimodal EEG-MRI Clinically-Informed Approach

**Saanvi S. Subramanian**

California Institute of Technology

**Mateus A. Esmeraldo**

Stanford University

**Matheus Tonholo Ikedo**

Stanford University

**Sergios Gatidis**

Stanford University

**Stefanie Chambers**

Medical University of Vienna

**Robbie Holland**

Stanford University

**Ashwin Kumar**

Stanford University

**Shreyas Vasanawala**

Stanford University

**Bruno P. Soares**

**`bpsouares@stanford.edu`**

Stanford University

---

## Research Article

**Keywords:** MRI, epilepsy, focal cortical dysplasia, nnU-Net, lesion detection, localization prior

**Posted Date:** September 7th, 2026

**DOI:** <https://doi.org/10.21203/rs.3.rs-10781100/v1>

**License:** © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.  
[Read Full License](#)

**Additional Declarations:** No competing interests reported.

---

# Structural-Functional Integration for Enhanced Detection of Subtle Epileptogenic Lesions: A Framework for a Multimodal EEG-MRI Clinically-Informed Approach

Saanvi S. Subramanian<sup>1\*</sup>, Mateus A. Esmeraldo<sup>2,3</sup>,  
Matheus Tonholo Ikedo<sup>2,3</sup>, Sergios Gatidis<sup>2,3</sup>,  
Stefanie Chambers<sup>4</sup>, Robbie Holland<sup>5</sup>, Ashwin Kumar<sup>5</sup>,  
Shreyas Vasanaawala<sup>3</sup>, Bruno P. Soares<sup>2,3\*</sup>

<sup>1\*</sup>Division of Biology and Biological Engineering, California Institute of Technology, Pasadena, CA, USA.

<sup>2</sup>Lab for Early-Age Neuroimaging at Stanford, Stanford, CA, USA.

<sup>3</sup>Department of Radiology, Stanford University, Stanford, CA, USA.

<sup>4</sup>Medical University of Vienna, Vienna, Austria.

<sup>5</sup>Stanford Center for Artificial Intelligence in Medicine and Imaging, Stanford University, Stanford, CA, USA.

\*Corresponding author(s). E-mail(s): [ssubramanian@caltech.edu](mailto:ssubramanian@caltech.edu);  
[bsoares@stanford.edu](mailto:bsoares@stanford.edu);

Contributing authors: [emerald3@stanford.edu](mailto:emerald3@stanford.edu); [mtonholo@stanford.edu](mailto:mtonholo@stanford.edu);  
[sgatidis@stanford.edu](mailto:sgatidis@stanford.edu); [stefanie.chambers@meduniwien.ac.at](mailto:stefanie.chambers@meduniwien.ac.at);  
[rholland@stanford.edu](mailto:rholland@stanford.edu); [akkumar@stanford.edu](mailto:akkumar@stanford.edu);  
[vasanaawala@stanford.edu](mailto:vasanaawala@stanford.edu);

## Abstract

Purpose: Focal cortical dysplasia (FCD) is a leading cause of treatment-resistant epilepsy and one of the most diagnostically challenging diseases in neuroimaging. Even the best deep learning models available to detect it rely on imaging data alone and achieve only suboptimal accuracy. A key insight motivates the approach taken here: experienced neuroradiologists and epilepsy specialists do not search the entire brain blindly. Instead, they integrate electroencephalography (EEG) and clinical findings that pinpoint the region of abnormal activity

where seizures originate, and use them to focus and prioritize the lesion search. This study tests whether an automated pipeline can mirror this search pattern by dynamically adjusting the detection threshold of a 3D nnU-Net according to the lesion location based on different localization scales (hemisphere, lobe, and sub-region). This design measures the upper-bound performance improvement that EEG or clinical priors could offer. Methods: A 3D nnU-Net was trained on 3D T2-FLAIR images from 175 FCD cases. Anatomical search zones at varying scales were built with SynthSeg, an atlas-based parcellation that runs in seconds per case, far below the roughly eight hours of model training per fold. Rather than applying the fixed 0.5 probability threshold everywhere, the pipeline lowered it inside the localization zone to surface faint sub-threshold peaks and raised it to 0.90 outside the zone to suppress false positives, scoring a case as recovered when any of the top five candidate peaks overlapped the lesion. Results: The model detected 86.3% of lesions (mean Dice 0.57), missing 24 cases (13.7%). Restricting a multi-peak search to the lesion lobe recovered 14 of the 24 misses (58%), which a chance baseline confirms is well above the 1.4 expected from a spatially uninformative search ( $P < 0.001$ ). Most of the benefit came from restriction to a moderate scale, the correct hemisphere or lobe; finer subregional restriction added no significant improvement over the lobe. Conclusion: This study shows that sub-threshold, spatially coherent signal often persists in cases where a standard nnU-Net reports the case as negative. Future work should focus on integrating real EEG and clinical data to improve deep learning detection pipelines.

**Keywords:** MRI, epilepsy, focal cortical dysplasia, nnU-Net, lesion detection, localization prior

## 1 Introduction

Focal cortical dysplasia (FCD) is one of the most common structural causes of treatment-resistant epilepsy, yet many pathologically confirmed cases show no visible abnormality on conventional brain MRI. Deep learning segmentation is now the strongest tool for automated FCD detection [1], and MELD Graph, a graph-based extension of the U-Net, is the best-validated approach [2]. It still errs on 12% to 25% of predictions, and it relies mainly on T1-weighted imaging even though T2-FLAIR is the sequence most informative to human readers for FCD [3, 4].

The 3D nnU-Net, a self-configuring segmentation method that adapts its preprocessing, architecture, and training to each dataset [7], is increasingly used for FCD [5, 6], but evaluations of such detectors report false-negative rates as high as 40% [3, 5]. Untreated refractory epilepsy carries a heavy cost, including developmental decline and functional impairment in up to 80% of affected children [8]. A complete miss may mean no model-detectable abnormality exists, or it may mean the abnormality is present as sub-threshold signal the model never surfaces. Clinicians resolve the same ambiguity by not searching blindly: EEG and clinical findings tell them which region to scrutinize on MRI. No FCD detection pipeline yet applies that kind of localization prior to a trained model’s sub-thresholded output. The purpose of this study is to quantify how much of that missed signal an idealized, error-free localization prior

can recover, establishing the ceiling that a real EEG- or clinically-derived prior could approach.

## 2 Materials and methods

This study adds anatomical localization information to a trained FCD detector and quantifies how much it recovers of the lesions the detector missed. For each case the lobe of the confirmed lesion is used for localization, an error-free stand-in for the region an EEG-guided work-up would flag. No EEG signal or clinical report is processed. Using the true lesion lobe sets the ceiling on what any real localization prior could add, and this quantity is worth knowing before building the full EEG pipeline. The output of a baseline 3D nnU-Net T2-FLAIR model is restricted to that lobe, and the number of missed lesions recovered is counted. Figure 1 summarizes the pipeline.

The model was trained on 175 radiologically confirmed FCD cases: 157 from three public cohorts and 18 institutional cases from a single tertiary epilepsy center. The public cohorts were the Bonn Open Presurgery MRI Dataset of people with epilepsy and FCD type II [9], the Imaging Database for Epilepsy and Surgery (IDEAS) [10], and the MR-RATE open source dataset from Forthmus, the University of Zurich, with NVIDIA and Istanbul Medipol University [11]. The Bonn dataset holds 85 FCD type II cases; 10 were dropped during quality control and coregistration (no usable T2-FLAIR, failed coregistration, or a label that could not be confirmed against the report), leaving the 75 used here. Table 1 gives the per-cohort counts.

**Table 1:** Number of confirmed FCD cases per dataset (FCD-only training cohort)

| Dataset       | Number of confirmed FCD cases |
|---------------|-------------------------------|
| Bonn          | 75                            |
| IDEAS         | 31                            |
| MR-RATE       | 51                            |
| Institutional | 18                            |
| Total         | 175                           |

All 175 cases are confirmed FCD, 157 from the open-source repositories (Bonn, IDEAS, MR-RATE) and 18 institutional cases from a single tertiary epilepsy center. The Bonn count is 75 of the 85 available FCD type II cases, after 10 were excluded during quality control and coregistration.

Training and inference used nnU-Net v2 in its 3D full-resolution configuration on an NVIDIA GeForce RTX 5090 GPU. T2-FLAIR was the single input channel (labels: 0 background, 1 FCD), since it is the sequence most sensitive to the cortical abnormalities of FCD [4]. Images were resampled to 1.0 mm isotropic spacing, Z-score normalized by nnU-Net, and processed at a patch size of  $160 \times 160 \times 96$  and batch size 2 by a six-stage PlainConvUNet (feature widths up to 320). Five-fold cross-validation was used throughout (1000 epochs per fold, roughly 8 h per fold), so each case was scored only by a fold that had not seen it.

The training procedure, architecture search, loss function, and validation of this baseline nnU-Net, including its healthy-control specificity, are the subject of a companion methods paper. Here the trained model is fixed. For each case it produces a voxel-level FCD probability map over the T2-FLAIR volume, trained with a differentiable pseudo-Dice loss that keeps sub-threshold probabilities in the gradient. That map is the only input to the recovery step.

For each case, the lobe holding the confirmed lesion is the localization prior, standing in for the lobe an EEG-guided work-up would identify. SynthSeg parcellates each case on the Desikan–Killiany–Tourville (DKT) atlas [12] in seconds on a GPU, so the parcellation would not slow a clinical read. The DKT parcels were grouped into eight regions, four per hemisphere: frontal, temporal, occipital, and parietal. The insula was not used as an independent zone, since scalp EEG rarely localizes it as the seizure onset zone; a case with an insular lesion was assigned a combined frontal and temporal zone (left insula becomes left frontal and temporal), given the insula’s adjacency to both. Finer DKT subdivision is possible but exceeds what EEG and clinical localization can specify with confidence. The selected region was expanded by a 10 mm margin for anatomical uncertainty to give the final zone. Annotation was performed by a radiologist with 5 years of experience (M.A.E.) and reviewed by a pediatric neuroradiologist with 21 years of experience (B.P.S.).

Inside the zone, the pipeline finds the highest-probability voxels, the peaks. Recovery was scored at top-1, top-3, and top-5, which differ only in how many spatially distinct local maxima are checked before a case counts as recovered: the single highest voxel, the three highest distinct maxima, or the five highest.

A single lesion can show up as several adjacent weak clusters rather than one strong one. To draw such cases as a single candidate, a similarity-weighted fusion step gives each peak a composite probability raised by nearby peaks, with each neighbor’s contribution decaying with its Euclidean distance  $d_{ij}$  from peak  $i$ :

$$p_{\text{comp}}(i) = p(i) \times \left( 1 + \sum_{j \neq i} \frac{p(j)}{p(i)} e^{-d_{ij}/\sigma} \right), \quad \sigma = 15.0 \text{ mm}. \quad (1)$$

The decay constant  $\sigma = 15.0$  mm was fixed a priori at the approximate spatial scale over which adjacent sub-peaks plausibly belong to the same lesion, not tuned to recovery. Two weak peaks that lie close together reinforce one another while an isolated weak peak is barely boosted. This composite value orders the peaks and drives the figure overlays. The reported recovery counts use the simpler test of whether a top- $k$  peak’s cluster overlaps the lesion, so this fusion changes the display, not the statistics. Each peak became a candidate region by thresholding the map and running a 6-connected flood fill from the peak, dropping any suprathreshold voxels not contiguous with it.

A dual threshold balances recovery against out-of-zone false positives. Inside the zone the search runs below the standard 0.5 threshold to surface faint signal. The out-of-zone threshold was fixed a priori at 0.90, before any recovery evaluation, as a deliberately conservative near-ceiling confidence level so that only very high-confidence out-of-zone signal survives; it was not selected by tuning against recovery. This keeps the model’s ability to flag a confident lesion crossing the zone edge while suppressing

background. Out-of-zone false-positive burden was not separately quantified for this model.

Two screens were used. The single-peak screen restricted the map to the indicated lobe and hemisphere and checked only the highest location there. The multi-peak screen checked the three (top-3) or five (top-5) highest distinct peaks, counting a case as recovered if any overlapped the lesion.

The granularity analysis used a high-difficulty cohort, defined objectively as every case with a baseline (unrestricted) Dice below 0.30 ( $n = 33$ : the 24 complete misses plus 9 faint-signal cases). This same cohort is used for the granularity comparison in the Results. Each case was scored at whole-hemisphere, lobar, and subregional granularity against the unrestricted baseline. Granularities were compared with McNemar’s exact test for paired binary outcomes, with significance at  $p < 0.05$ .

In this proof-of-concept study, every localization zone was simulated from the lesion’s true location rather than from real clinical or EEG data. For 155 of the 175 cases the zone was the lobe recorded during lesion annotation. For the 20 without a recorded localization (the 18 institutional cases and two public-cohort cases; listed in the Supplementary Information) the lobe was taken from the intersection of the ground-truth mask with the SynthSeg parcellation. Both routes place the zone on the confirmed lesion, so the prior is error-free by construction. A case whose recorded localization did not contain the lesion was to be excluded, though none met that rule. The two routes agreed closely: the ground-truth-derived hemisphere matched the recorded annotation in 100% of annotated cases and the lobe in 92%.

Recovery was measured two ways. Strict recovery required the highest-probability peak to overlap the lesion and to exceed ten times the 99.9th percentile of background probability. Lenient recovery required any of the top-five clusters to share at least one voxel with the lesion. Candidate lesions were shown by overlaying a 2 cm-radius probability heatmap and connected-component contours on native T2-FLAIR. Each case figure shows the unannotated and annotated T2-FLAIR side by side (Figures 3 and 4); windowing and lesion depiction was reviewed by M.T.I.

To test whether recovery reflects true signal, two chance baselines that carry no lesion information were run, 300 draws per case, scored by the same overlap test. The permutation null shuffles the probability values inside the zone, destroying spatial structure while the value distribution and the zone stay fixed, then reruns the top-5 recovery. The random-cluster null grows a connected region the size of the real recovered cluster from a random in-zone seed.

## 3 Results

### 3.1 FCD-only model: cross-validation detection

Across five-fold cross-validation on 175 cases, the T2-FLAIR-only nnU-Net reached an aggregate Dice of 0.5703 (per-fold 0.5075–0.6243) and a mean voxel recall of 0.5359. At the lesion level (any-voxel overlap), it detected 151 of 175 lesions (86.3%) and missed 24 (13.7%). Detection varied by cohort (Table 2).

The 24 missed lesions were smaller than the 151 detected (missed median 1,852 mm<sup>3</sup>, IQR 1,194–3,493; detected median 3,487 mm<sup>3</sup>, IQR 2,215–6,043; one-sided

**Table 2:** Lesion-level detection by cohort under five-fold cross-validation

| Cohort        | Detected / total | Detection rate | Mean Dice |
|---------------|------------------|----------------|-----------|
| Bonn          | 61 / 75          | 81.3%          | 0.513     |
| IDEAS         | 28 / 31          | 90.3%          | 0.585     |
| MR-RATE       | 46 / 51          | 90.2%          | 0.642     |
| Institutional | 16 / 18          | 88.9%          | 0.580     |
| All           | 151 / 175        | 86.3%          | 0.570     |

Detection counts a lesion as found when the model output overlaps the ground truth by at least one voxel. Mean Dice is over all cases in the cohort. The aggregate Dice over the full cohort is 0.5703.

Mann–Whitney  $U = 1,195$ ,  $p = 0.0037$ ), 18 of them below the median detected size, and 14 came from Bonn. They split into two modes. Seventeen were *silent*: the model produced no suprathreshold foreground anywhere. Seven were *mislocalized*: the model flagged an FCD-like region 34 to 139 mm from the lesion, a distant false positive while the lesion went unmarked. Six of the misses were large (about 3% of the dataset, the two largest at 16,561 and 11,352 mm<sup>3</sup>, both silent); on radiological review (M.T.I.) these sat in challenging locations near the amygdala or the horns of the lateral ventricles. The two problems the recovery step targets are faint or absent signal in the smallest lesions and confident but misplaced predictions in others.

### 3.2 Localization-guided recovery

Recovery ran on the FCD-only model’s probability maps. The zone came from the recorded clinical annotation for 155 cases and from the ground-truth intersection with the parcellation for the 20 without one (Section 2). No case was excluded for an annotation that missed the lesion. The insula was reassigned to a combined frontal and temporal zone throughout, as in Section 2.

On the high-difficulty cohort defined in Section 2 (baseline Dice below 0.30,  $n = 33$ ), turning on anatomical restriction raised recovery (Dice above 0) from 27.3% (9/33) unrestricted to 45.5% (15/33) at the hemisphere, 51.5% (17/33) at the lobe, and 54.5% (18/33) at the subregion. Restriction to the lobe beats no restriction (McNemar’s exact test,  $p < 0.01$ ), so the anatomical prior helps. The benefit concentrates at a moderate granularity and then saturates. The one sizeable step is switching restriction on at the hemisphere (no restriction to hemisphere  $p = 0.07$ ); finer restriction adds nothing detectable (hemisphere to lobar  $p = 0.50$ , lobar to subregional  $p = 1.00$ ; Figure 2). Tightening the zone can even remove signal, since the subregional zone dropped two cases the lobe had recovered. A hemisphere-to-lobe prior captures the available gain, and sublobar precision is not required.

Under the multi-peak (top-5) setting, restricting to the lobar zone recovered 14 of the 24 complete misses (58.3%) by the lenient any-voxel-overlap criterion, of which 12 hold up against the chance baseline and 8 clear it individually (quantified below). The two modes behaved differently. Ten of the 17 silent misses were recovered, so sub-threshold signal that the standard pipeline reports as nothing is present and spatially coherent in most of them. Four of the 7 mislocalized misses were recovered, where

confining the search to the correct lobe pulled detection off the distant false positive and onto the lesion. Recovery was spread across the frontal, temporal, and occipital zones and the reassigned insula case, with small per-lobe counts (Supplementary Table S3). Several recoveries reached moderate overlap from a baseline Dice of 0.0, the strongest at top-5 Dice 0.447, 0.431, and 0.291 (Supplementary Table S2). The other 10 misses recovered under no top-five peak, so a share of complete misses hold no model-detectable signal in the zone or hold it outside.

The soft out-of-zone penalty (threshold 0.90) recovered the same cases as a hard exclusionary penalty while keeping lesion boundaries that ran past the zone, so it was adopted as the single rule and the hard-versus-soft comparison is not reported separately. The 7 mislocalized cases, where the model made a confident prediction 34 to 139 mm from the lesion, are the mode lobe restriction is best placed to fix, and 4 of the 7 recovered once the search was confined to the correct lobe (Supplementary Table S1).

Figures 3 and 4 show how the pipeline renders a recovered lesion on native T2-FLAIR: the pale blue outline is the localization zone, the hot-colormap discs are the recovered sub-threshold probability, and each peak carries its composite probability, so the lesion can be judged before the overlay is read.

Every one of the 14 recoveries first appeared within the top four peaks: 8 at the highest peak, 3 at the second, 2 at the third, and 1 at the fourth. The lone fourth-peak recovery is a mislocalized miss whose wider search space puts several higher-probability peaks ahead of the lesion-overlapping one. Top-4 therefore captures the full top-5 benefit and is the practical setting; top-3 recovers 13 of the 14. The reassigned insula case recovered at the first peak.

The 14 recoveries were then tested against the two chance baselines. Under the shuffled-probability null, a case recovered by chance 5.7% of the time on average, so 1.4 of the 24 misses would be expected against the 14 observed (Monte-Carlo  $P(\text{null} \geq 14) < 0.001$ ). The random-cluster null is more permissive, since a cluster placed anywhere in the correct lobe can strike a large lesion; it recovered 16.6% of the time, or 4.0 of 24 expected, still far below 14 ( $P < 0.001$ ). At the case level, 8 of the 14 recoveries beat their own shuffled-probability null at  $p < 0.05$ , and these carried the highest overlap (Supplementary Table S2). The recoveries that did not clear chance were the three largest missed lesions (all above 10,000 mm<sup>3</sup>; Supplementary Table S2), where a random cluster overlaps 60 to 86% of the time, so those specific recoveries are not separable from geometry. For two of the three the observed overlap fell below the permutation-null mean, so dropping them gives a conservative count of 12 of 24. The 14 of 24 headline is thus the lenient upper bound, 12 is the null-defensible count, and 8 is the number of recoveries that individually clear their own shuffled-probability null; the 8 carry the highest overlap and are the firmly supported cases. Per-case null values are in Supplementary Table S2.

## 4 Discussion

The T2-FLAIR nnU-Net detected 86.3% of the 175 lesions (aggregate Dice 0.5703) and missed 13.7% (24 cases), 17 silent and 7 mislocalized, with the missed lesions smaller

than the detected ones ( $p = 0.0037$ ). Restricting a multi-peak search to the correct lobe recovered 14 of the 24 (58%) by the lenient criterion, 12 after dropping two large lesions whose overlap fell below the null and 8 that clear their own null individually, all well above the 1.4 expected from a spatially uninformative search ( $p < 0.001$ ). The two failure modes answered to the prior in different ways. Ten of 17 silent misses recovered, so faint sub-threshold signal persists where the standard pipeline reports nothing, and the zone-restricted search surfaces it. Four of 7 mislocalized misses recovered, where the prior pulled detection off the distant false positive and onto the lesion. Every recovery came within the top four peaks (8, 3, 2, 1 at ranks one through four), so top-4 is the practical setting and top-3 recovers 13. On the high-difficulty cohort, lobe restriction beat no restriction ( $p < 0.01$ ), but the gain saturated at a moderate granularity: the meaningful step was switching restriction on at the hemisphere, and neither the lobe nor the subregion added significantly beyond it ( $p = 0.50$  and  $p = 1.00$ ). This plateau is expected. Once the search is confined to the correct hemisphere or lobe, the sub-threshold signal that exists already lies inside the zone, so a tighter sublobar boundary mainly trims true signal while removing few competing peaks, and with only 33 cases the incremental tests have little power to detect a small further gain. The unrecovered misses hold no model-detectable signal in the zone, or signal outside it, and lobe restriction cannot reach them.

Most of the gain comes from the correct hemisphere, with small further gains from finer restriction. The lobe-level spread in recovery rests on a handful of cases and is a question for later work on regional EEG reliability, not a result.

The localization prior here is the confirmed lesion’s own lobe, an idealized stand-in that sets the ceiling on what an anatomical prior can add. A real EEG- or clinically-derived prior would carry different, and in important respects richer, information. Scalp EEG, and interictal recordings in particular, localizes the irritative zone and the seizure onset zone, which need not coincide with the lesion itself. That difference is clinically decisive: resection limited to the visible lesion is often insufficient for seizure freedom, whereas removal of the broader epileptogenic zone, which extends beyond the lesion, gives the best chance of a seizure-free outcome. Because a moderate, larger search zone gave the best recovery here, an EEG-defined region that is deliberately broader than the lesion is well matched to what the pipeline needs, and adds information a lesion mask alone cannot. Recent deep learning tools that project ictal and interictal EEG onset onto structural MRI [13] offer a route to supplying such a prior and are a natural next step.

The insula was folded into a combined frontal and temporal zone here because scalp EEG rarely localizes it, but its characteristic seizure semiology means a model that incorporates clinical and EEG data could reasonably treat the insula as a separate region.

The method attaches a localization prior to a raw probability map, and it works with any model that outputs a probability per voxel and any source of anatomical localization, not only EEG. The prior is applied after training rather than as an extra input channel by design: because the prior here is the true lesion lobe, feeding it as a channel would leak the label and inflate performance, so post-hoc restriction is what isolates the value of the prior. A learned mask channel becomes appropriate once a

real, imperfect EEG prior is available. With a real prior in place, the recovery step could flag faint sub-threshold candidates inside a clinically informed search region during presurgical planning and ease the reading burden on neuroradiologists. The soft out-of-zone penalty (threshold 0.90) raised recovery from 27.3% to 51.5% at the lobe while holding down the false-positive spike a looser threshold would cause, and it kept the model’s ability to flag a confident lesion outside the zone. The native-T2-FLAIR overlay, a probability heatmap with contours, makes the recovered signal legible for review.

Two robustness checks support the design. First, the recovery count is insensitive to the pipeline’s hyperparameters: it holds at 14 of 24 across zone margins from 5 to 20 mm and cluster hot-fractions from 0.15 to 0.35, and the out-of-zone threshold from 0.70 to 0.95 leaves the soft-rule recovery unchanged. These hyperparameters were fixed a priori, before the recovery evaluation, and the sweep confirms the result does not depend on the specific values chosen. Only the number of candidate peaks changes the count, and that is reported as a full curve that saturates by the fourth peak (Supplementary Table S5). Second, the benefit depends on the prior being correct. Replacing the true lobe with the contralateral-homotopic lobe drops recovery from 14 to 1 of 24, and a random wrong lobe to about 1.5 of 24, both below the 9 of 24 recovered with no restriction at all (Supplementary Table S4). A wrong localization therefore suppresses true lesions instead of merely failing to help, which sets a reliability bar for any authentic EEG prior and is the failure mode to watch in clinical use.

## Limitations

The localization information is simulated from the confirmed lesion lobe, not from raw source localization or a real EEG report. Therefore, the recovery rates here set an upper bound on localization accuracy. A real EEG-derived localization information will often be less precise about the exact lobe, though it also carries information about the epileptogenic zone beyond the lesion. Extending the pipeline to raw EEG is the next step. For the 20 cases without a recorded clinical localization, the zone came from the lesion’s own lobe rather than an independent annotation. This ground-truth-derived hemisphere matched clinical annotation in 100% of annotated cases and the lobe in 92%, but it may modestly inflate recovery for those 20. Per-lobe comparisons rest on one or two cases each and are only suggestive.

Two points bear on how the central result reads. The first is signal magnitude. In the silent misses the recovered signal sits at very low absolute probability ( $10^{-5}$  to  $10^{-6}$ ). The overlap is spatially coherent but weak in magnitude, and the best cluster was often a lower-ranked peak. This is why the multi-peak search is needed. The second is zone size. Because the zone is a whole lobe plus a 10 mm margin, a cluster placed anywhere in the correct lobe can overlap a large lesion by chance. The chance baselines address this directly. The 14 recoveries far exceed the 1.4 and 4.0 expected under the two nulls ( $P < 0.001$  for both), so the effect is not an artifact of zone size. Even so, three recoveries, all of the largest lesions, are not separable from chance, and only 8 of the 14 individually clear the shuffled-probability null at  $p < 0.05$ . The recoveries with the highest overlap are the ones that most clearly beat chance.

The study also lacks an external test set and a specificity check of the recovery step on healthy controls, and the recovery hyperparameters ( $\sigma$ , the flood-fill thresholds, the 10 mm margin) were set on this cohort with no held-out split, though the detection estimates are already out-of-fold under cross-validation and the recovery count is stable across those hyperparameters (above); a separate external test set remains the stronger design and is planned. The insula reassignment to a combined frontal and temporal zone is on clinical grounds and the one reassigned case recovered at the first peak.

## Conclusion

Restricting a T2-FLAIR nnU-Net’s search to the lobe known to hold the lesion recovered 14 of 24 FCD cases the image-only model missed, 10 of them silent misses with no suprathreshold output at all, with every recovery inside the top four peaks and the cohort-level effect above chance ( $P < 0.001$ ). The simulated nature of the EEG localization information means these rates are the ceiling a localization prior could approach. This proof-of-concept study shows that spatially coherent sub-threshold signal often survives in cases a standard nnU-Net calls negative. Misses that may fall into silent or mislocalized modes would need to be corrected when using real EEG localization information. Integrating text from clinical EEG reports or, even better, using raw EEG data is the logical next step and a key direction for future work.

## Supplementary Information

**Table S1:** Mislocalized misses: cases missed by the baseline (zero overlap) in which the model produced a confident but spatially distant prediction, and whether lobe-restricted multi-peak search recovered the lesion

| Case                 | Zone     | GT (vox) | Dist. (mm) | Recovered (rank) |
|----------------------|----------|----------|------------|------------------|
| Bonn_sub-00066       | F+T left | 2,863    | 139.3      | yes (1)          |
| Bonn_sub-00095       | F left   | 1,767    | 120.2      | yes (1)          |
| Inst_001             | O right  | 10,159   | 80.9       | no               |
| MR-RATE_subtle-00040 | T left   | 12,584   | 62.8       | no               |
| Bonn_sub-00116       | F right  | 1,441    | 42.3       | yes (4)          |
| Bonn_sub-00065       | F left   | 1,177    | 40.6       | yes (3)          |
| MR-RATE_subtle-00317 | T left   | 730      | 34.5       | no               |

Zone codes are F frontal, T temporal, O occipital (F+T is the reassigned insula zone). “Dist.” is the distance between the model’s out-of-zone predicted cluster and the lesion. Inst\_001 is a de-identified institutional case that used the ground-truth-derived zone; the rest used the recorded clinical localization.

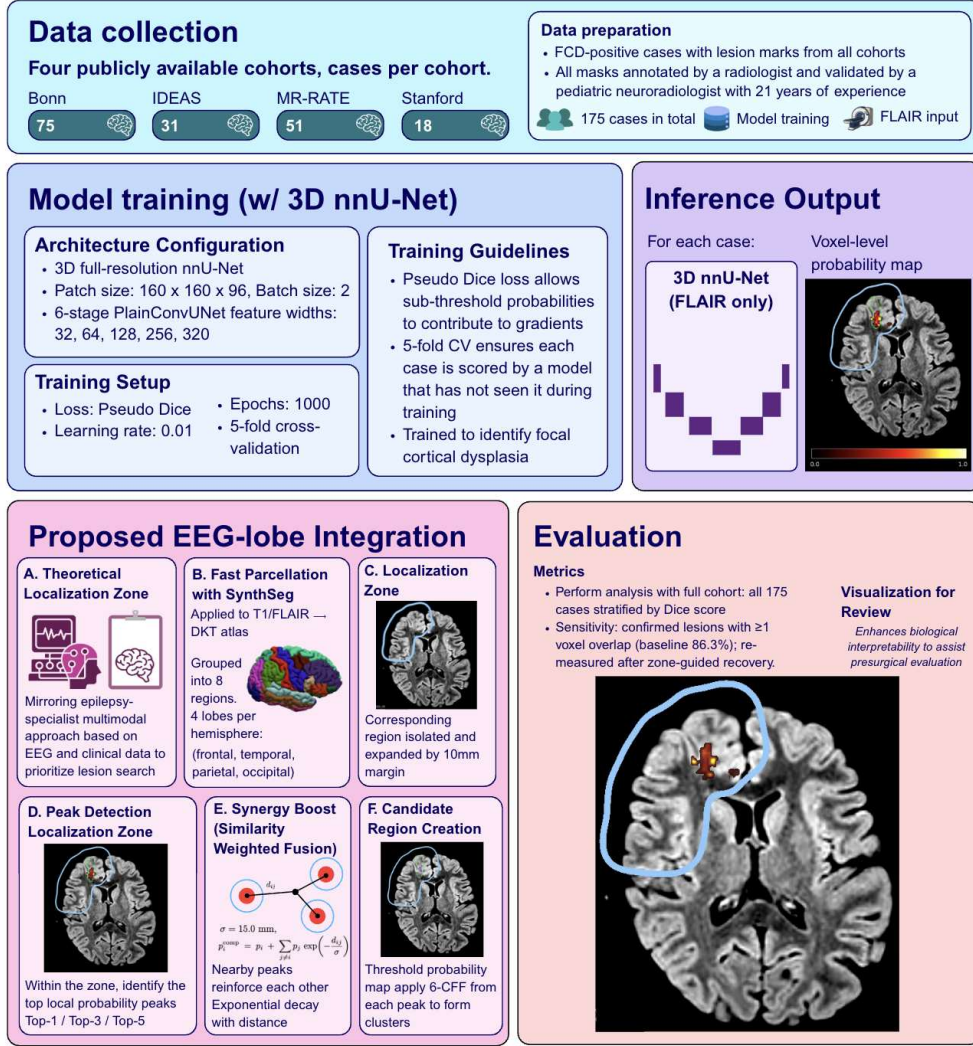


**Fig. 1:** The localization-guided recovery pipeline. A baseline 3D nnU-Net trained on T2-FLAIR produces a voxel-level FCD probability map. A lobar localization zone, taken here from each case’s confirmed lesion location as an error-free stand-in for an EEG-based lobar finding, defines the search region. Inside the zone, an adaptive peak-detection and clustering step surfaces sub-threshold signal from missed lesions; outside it, a stricter threshold limits false positives.

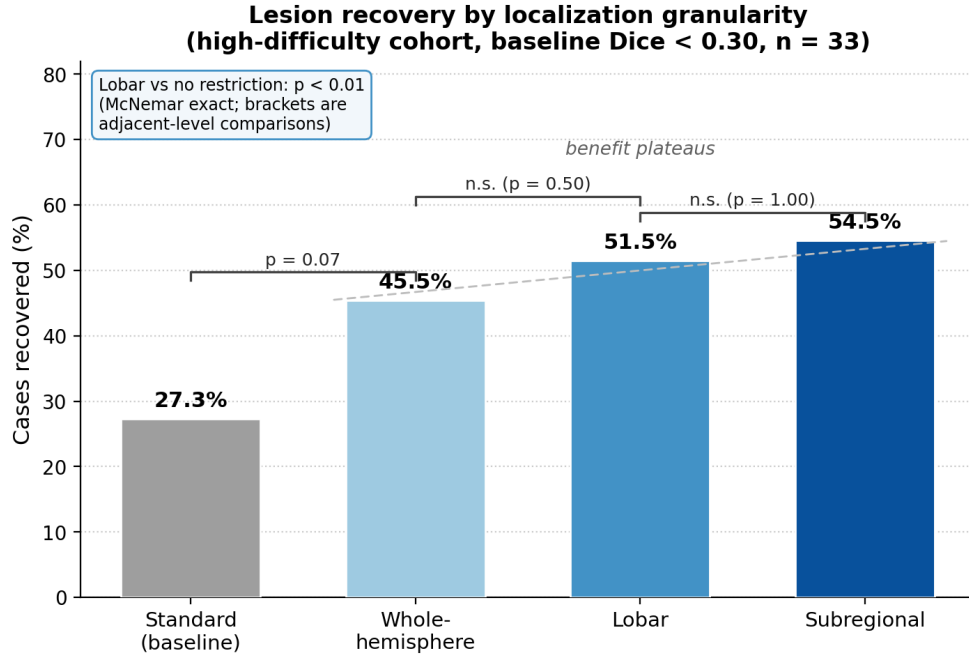


**Fig. 2:** Cases recovered (Dice above 0) by localization granularity across the 33-case high-difficulty cohort (baseline Dice below 0.30). Recovery rises when anatomical restriction is applied (27.3% to 45.5% at the hemisphere) and then plateaus: lobar (51.5%) and subregional (54.5%) add little. Brackets give McNemar’s exact test between adjacent levels; restriction helps most at the moderate hemisphere-to-lobe range (no restriction to hemisphere  $p = 0.07$ ), while finer restriction is not significant (hemisphere to lobar  $p = 0.50$ ; lobar to subregional  $p = 1.00$ ). As an anchor, lobar restriction beats no restriction at  $p < 0.01$ . Mean Dice per condition: standard 9/33 (0.0456), whole-hemisphere 15/33 (0.0607), lobar 17/33 (0.0588), subregional 18/33 (0.0635). This test uses a stricter recovery definition than the multi-peak lenient-overlap analysis in the text.

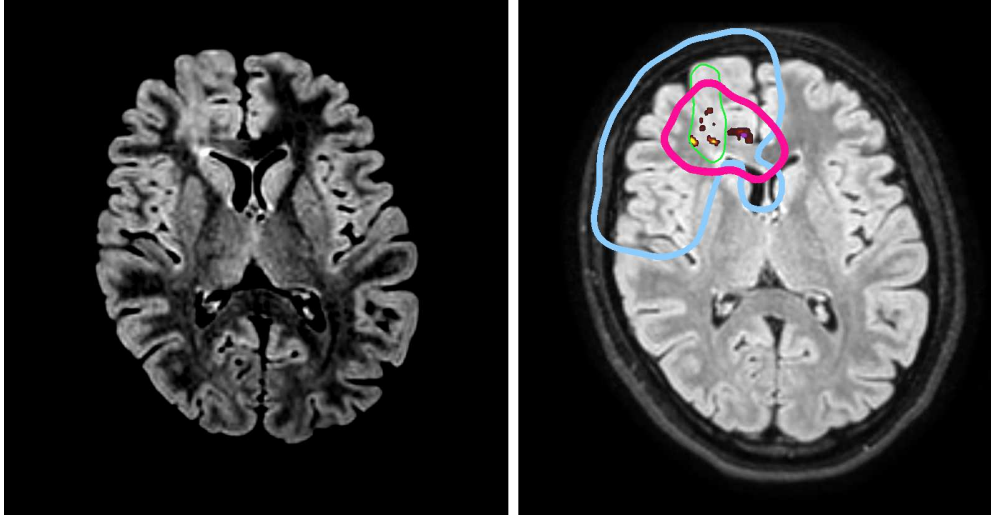


**Fig. 3:** A recovered silent miss (MR-RATE\_subtle-00266) on native T2-FLAIR, with no suprathreshold model output at baseline (Dice 0.0). Left, the unannotated case; right, the same slice with the localization zone, probability overlay, and recovered peaks. Color legend (right panel): pale blue outline, localization zone; green outline, ground-truth lesion; hot (thermal) colormap discs, model probability around each detected sub-threshold peak (warmer is higher); orange contour, an isolated peak core; purple contour, peaks fused by the similarity step of Equation 1; pink contour, the envelope grouping nearby peaks into one candidate; each peak is labeled with its rank and composite probability. The zone-restricted multi-peak search raised the Dice from 0.0 to 0.447 (best-overlapping cluster among the top five peaks).

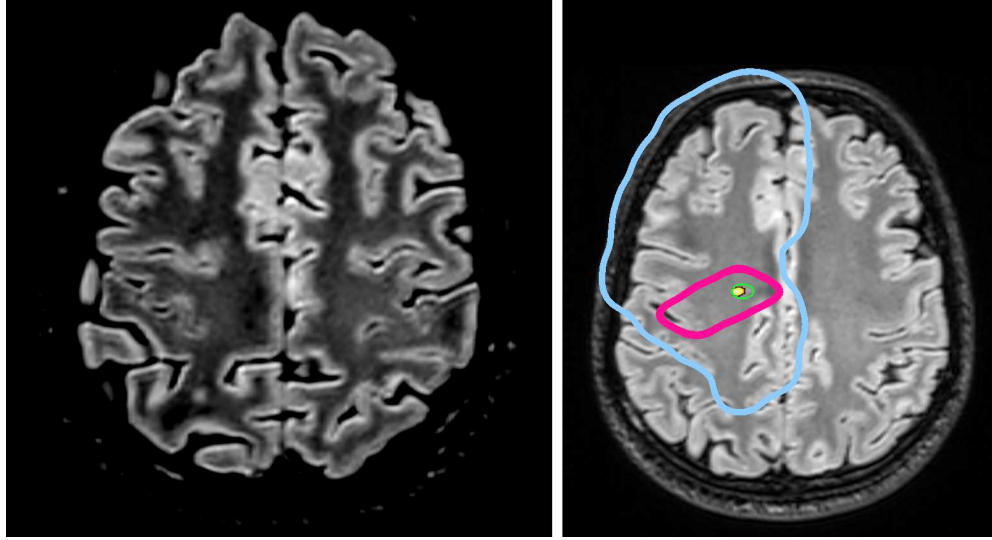


**Fig. 4:** A second recovered silent miss (MR-RATE\_subtle-00024, frontal right, baseline Dice 0.0). Left, the unannotated case; right, the same slice with the overlay, using the same color legend as Figure 3: pale blue is the localization zone, green is the ground-truth lesion, the hot (thermal) colormap is model probability, orange is an isolated peak core, purple is a fused peak core, and pink is the envelope grouping nearby peaks. The zone-restricted multi-peak search raised the Dice from 0.0 to 0.261.

**Table S2:** Chance-baseline test for each of the 24 misses. “Real Dice” is the best-overlapping cluster among the top five peaks. “Perm. null” and “Rand. null” are the fractions of 300 shuffled-probability and 300 random-cluster draws that recovered the case. A recovery is credible when its null values are low; the three large-lesion recoveries with high null values (Bonn\_sub-00040, Bonn\_sub-00133, IDEAS\_subtle\_122) are not separable from chance.

| Case                 | Failure mode | Real Dice | Perm. null | Rand. null |
|----------------------|--------------|-----------|------------|------------|
| Bonn_sub-00034       | silent       | 0.000     | 0.003      | 0.020      |
| Bonn_sub-00038       | silent       | 0.015     | 0.017      | 0.033      |
| Bonn_sub-00040       | silent       | 0.284     | 0.280      | 0.860      |
| Bonn_sub-00048       | silent       | 0.000     | 0.010      | 0.010      |
| Bonn_sub-00062       | silent       | 0.000     | 0.023      | 0.027      |
| Bonn_sub-00065       | mislocalized | 0.010     | 0.020      | 0.043      |
| Bonn_sub-00066       | mislocalized | 0.209     | 0.013      | 0.040      |
| Bonn_sub-00080       | silent       | 0.431     | 0.010      | 0.020      |
| Bonn_sub-00095       | mislocalized | 0.186     | 0.013      | 0.030      |
| Bonn_sub-00103       | silent       | 0.262     | 0.030      | 0.123      |
| Bonn_sub-00108       | silent       | 0.000     | 0.020      | 0.033      |
| Bonn_sub-00116       | mislocalized | 0.025     | 0.013      | 0.043      |
| Bonn_sub-00122       | silent       | 0.104     | 0.080      | 0.383      |
| Bonn_sub-00133       | silent       | 0.096     | 0.280      | 0.610      |
| IDEAS_subtle_122     | silent       | 0.019     | 0.163      | 0.657      |
| IDEAS_subtle_182     | silent       | 0.000     | 0.053      | 0.093      |
| IDEAS_subtle_192     | silent       | 0.000     | 0.023      | 0.033      |
| MR-RATE_subtle-00024 | silent       | 0.261     | 0.027      | 0.067      |
| MR-RATE_subtle-00040 | mislocalized | 0.000     | 0.120      | 0.330      |
| MR-RATE_subtle-00266 | silent       | 0.447     | 0.060      | 0.173      |
| MR-RATE_subtle-00317 | mislocalized | 0.000     | 0.010      | 0.030      |
| MR-RATE_subtle-00326 | silent       | 0.291     | 0.053      | 0.183      |
| Inst_001             | mislocalized | 0.000     | 0.037      | 0.073      |
| Inst_002             | silent       | 0.000     | 0.017      | 0.060      |

Inst\_001 and Inst\_002 are de-identified labels for the two institutional cases; they carry no medical record information.

**Table S3:** Recovery of the 24 complete misses by localization zone under lobe-restricted multi-peak search. Counts are small and are reported for completeness.

| Zone                                        | Recovered / total |
|---------------------------------------------|-------------------|
| Frontal left                                | 4 / 7             |
| Frontal right                               | 4 / 6             |
| Temporal right                              | 3 / 4             |
| Temporal left                               | 1 / 4             |
| Occipital right                             | 1 / 2             |
| Insula (reassigned to frontal and temporal) | 1 / 1             |
| All                                         | 14 / 24           |

**Table S4:** Wrong-prior robustness. Top-5 lobar recovery of the 24 misses when the localization prior points to the correct lobe, the contralateral-homotopic lobe, a random other lobe (averaged over all incorrect lobe-and-side zones), and no restriction (whole-brain search). A wrong prior recovers fewer true lesions than no restriction at all.

| Localization prior           | Recovered / 24 | Silent / 17 | Mislocalized / 7 |
|------------------------------|----------------|-------------|------------------|
| Correct lobe                 | 14             | 10          | 4                |
| Contralateral-homotopic lobe | 1              | 1           | 0                |
| Random wrong lobe (expected) | 1.5            | 1.3         | 0.2              |
| No restriction (whole brain) | 9              | 8           | 1                |

Recovery uses the same top-5 any-voxel-overlap criterion as the main analysis. The random-wrong row averages recovery over every (lobe, side) zone that is not the correct one.

**Table S5:** Hyperparameter sensitivity of the recovery count. Each parameter was swept with the others held at their default; the top-5 lobar recovery is flat across margin and hot-fraction, and the out-of-zone threshold leaves the soft-rule recovery unchanged. Only the number of candidate peaks changes the count, and it saturates by the fourth peak.

| Parameter             | Values tested                | Recovered / 24               |
|-----------------------|------------------------------|------------------------------|
| Zone margin           | 5, 10, 15, 20 mm             | 14 at every value            |
| Cluster hot-fraction  | 0.15, 0.20, 0.25, 0.30, 0.35 | 14 at every value            |
| Candidate peaks       | top-1 / 2 / 3 / 4 / 5        | 8 / 11 / 13 / 14 / 14        |
| Out-of-zone threshold | 0.70, 0.80, 0.90, 0.95       | 8 at every value (soft rule) |

The first three rows use the top-5 any-voxel-overlap recovery (default 14 of 24). The out-of-zone threshold governs only the stricter soft-penalty single-cluster recovery (Dice above 0), which is flat at 8 of 24; it does not enter the top-5 count.

The 20 cases without a recorded clinical localization, whose zone was taken from the ground-truth intersection with the parcellation, comprise the 18 institutional cases and two public-cohort cases (IDEAS\_subtle\_306 and MR-RATE\_subtle-00180).

## Acknowledgements

This work was supported by the California Institute of Technology Summer Undergraduate Research Fellowship (SURF) program.

## Declarations

### Funding

This work was supported in part by the California Institute of Technology Summer Undergraduate Research Fellowships (SURF) program.

### Conflict of interest

The authors declare no competing interests.

### Ethics approval and consent to participate

The three public datasets (Bonn, IDEAS, MR-RATE) are de-identified and publicly available. The 18 institutional cases were obtained at a single tertiary epilepsy center under Institutional Review Board approval, with a waiver of informed consent for retrospective analysis of de-identified imaging.

### Data availability

The three public datasets (Bonn, IDEAS, MR-RATE) are available as cited in the References. The 18 institutional cases are not publicly available owing to patient-privacy restrictions and institutional policy.

## Code availability

Available from the corresponding author upon reasonable request.

## References

- [1] Jiménez-Murillo D, Castro-Ospina AE, Duque-Muñoz L et al (2023) Automatic detection of focal cortical dysplasia using MRI: a systematic review. *Sensors* 23:7072. <https://doi.org/10.3390/s23167072>
- [2] Ripart M et al (2025) Detection of epileptogenic focal cortical dysplasia using graph neural networks: a MELD study. *JAMA Neurol* 82(4):397. <https://doi.org/10.1001/jamaneurol.2024.5406>
- [3] Esmeraldo MA, Chambers S, Kravutske Y et al (2026) Pipeline evaluation of a state-of-the-art AI algorithm for detection of focal cortical dysplasia: insights into potential failure sources. *Brain Inf* 13:13. <https://doi.org/10.1186/s40708-026-00299-w>
- [4] Pastore LV, Sudhakar SV, Mankad K, De Vita E, Biswas A, Tisdall MM, Chari A, Fignini M, Tahir MZ, Adler S, Moeller F, Cross JH, Pujar S, Wagstyl K, Ripart M, Löbel U, Cirillo L, D’Arco F. Integrating standard epilepsy protocol, ASL-perfusion, MP2RAGE/EDGE and the MELD-FCD classifier in the detection of subtle epileptogenic lesions: a 3 Tesla MRI pilot study. *Neuroradiology*. 2025 Mar;67(3):665-675. doi: 10.1007/s00234-024-03488-8. Epub 2024 Oct 23. PMID: 39441414.
- [5] Kersting LN et al (2025) Detection of focal cortical dysplasia: development and multicentric evaluation of artificial intelligence models. *Epilepsia* 66(4):1165–1176. <https://doi.org/10.1111/epi.18240>
- [6] Joshi S et al (2025) A nnU-Net-based automatic segmentation of FCD type II lesions in 3D T2-FLAIR MRI images. *Front Artif Intell* 8:1601815. <https://doi.org/10.3389/frai.2025.1601815>
- [7] Isensee F, Jaeger PF, Kohl SAA, Petersen J, Maier-Hein KH (2021) nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. *Nat Methods* 18(2):203–211. <https://doi.org/10.1038/s41592-020-01008-z>
- [8] Reilly C, Atkinson P, Das KB, Chin RFMC, Aylett SE, Burch V, Gillberg C, Scott RC, Neville BGR (2014) Neurobehavioral comorbidities in children with active epilepsy: a population-based study. *Pediatrics* 133:e1586–e1593
- [9] Schuch F et al (2023) An open presurgery MRI dataset of people with epilepsy and focal cortical dysplasia type II. *Sci Data* 10(1):475. <https://doi.org/10.1038/s41597-023-02386-7>

- [10] Taylor PN et al (2025) The Imaging Database for Epilepsy And Surgery (IDEAS). *Epilepsia* 66(2):471–481. <https://doi.org/10.1111/epi.18192>
- [11] Forithmus, University of Zurich, NVIDIA, Istanbul Medipol University (2023) MR-RATE Open Source Dataset. Hugging Face. <https://huggingface.co/datasets/Forithmus/MR-RATE>
- [12] Klein A, Tourville J (2012) 101 labeled brain images and a consistent human cortical labeling protocol. *Front Neurosci* 6:171. <https://doi.org/10.3389/fnins.2012.00171>
- [13] Sun R et al (2023) Deep learning based source imaging provides strong sublobar localization of epileptogenic zone from MEG interictal spikes. *NeuroImage*. <https://pmc.ncbi.nlm.nih.gov/articles/PMC10771628/>