

# Self-Supervised Distillation of Legacy Rule-Based Methods for Enhanced EEG-Based Decision-Making

Yipeng Zhang<sup>1</sup>[0000-0003-2869-4692], Yuanyi Ding<sup>1</sup>[0009-0006-5566-675X],  
Chenda Duan<sup>1</sup>[0009-0003-8652-3960], Atsuro Daida<sup>2</sup>[0000-0003-1350-7700], Hiroki  
Nariai<sup>2</sup>[0000-0002-8318-2924], and Vwani Roychowdhury<sup>1</sup>[0000-0003-0832-6489]

<sup>1</sup> Samueli School of Engineering, UCLA, USA

<sup>2</sup> Mattel Children's Hospital, David Geffen School of Medicine, UCLA, USA  
vwani@g.ucla.edu

**Abstract.** High-frequency oscillations (HFOs) in intracranial Electroencephalography (iEEG) are critical biomarkers for localizing the epileptogenic zone in epilepsy treatment. However, traditional rule-based detectors for HFOs suffer from unsatisfactory precision, producing false positives that require time-consuming manual review. Supervised machine learning approaches have been used to classify the detection results, yet they typically depend on labeled datasets, which are difficult to acquire due to the need for specialized expertise. Moreover, accurate labeling of HFOs is challenging due to low inter-rater reliability and inconsistent annotation practices across institutions. The lack of a clear consensus on what constitutes a pathological HFO further challenges supervised refinement approaches. To address this, we leverage the insight that legacy detectors reliably capture clinically relevant signals despite their relatively high false positive rates. We thus propose the Self-Supervised to Label Discovery (SS2LD) framework to refine the large set of candidate events generated by legacy detectors into a precise set of pathological HFOs. SS2LD employs a variational autoencoder (VAE) for morphological pre-training to learn meaningful latent representation of the detected events. These representations are clustered to derive weak supervision for pathological events. A classifier then uses this supervision to refine detection boundaries, trained on real and VAE-augmented data. Evaluated on large multi-institutional interictal iEEG datasets, SS2LD outperforms state-of-the-art methods. SS2LD offers a scalable, label-efficient, and clinically effective strategy to identify pathological HFOs using legacy detectors. The code is available at <https://github.com/roychowdhuryresearch/SS2LD>.

**Keywords:** Self-supervised learning · EEG · Variational Autoencoder

## 1 Introduction

Human-defined biomarkers encapsulate the expertise of medical professionals and play a pivotal role in medical diagnosis. In EEG-based clinical practice,

neurophysiological biomarkers such as spikes, spindles, and high-frequency oscillations (HFOs) are commonly used. To detect these biomarkers computationally, numerous statistical methods have been developed, including the IDE spike detector [15], the YASA spindle detector [16], and HFO detection such as STE [13] or MNI [20]. However, these rule-based approaches often suffer from low precision, generating false positives [25]. Clinicians must manually review a large volume of detections, which is both time-consuming and prone to human error.

Machine learning approaches, especially supervised methods, have been employed to refine detection results [4,8,22] or to classify EEG segments produced by threshold-based “pseudo-detectors” [18]. While these methods can show good performance on their defined test set, they are typically derived from relatively small cohorts and heavily depend on accurately labeled training data, limiting their scalability and generalizability. Annotating such events requires specialized medical knowledge, and clinical studies have shown low inter-rater reliability for these biomarkers [10,12,11]; standardization of annotation practices across institutions is also lacking [5,7]. Weakly supervised methods [23,9] reduce annotation demands, but require complex loss tuning, preventing clinical use.

Nonetheless, earlier studies have predominantly overlooked a critical insight: rule-based biomarker detectors, developed over many years of clinical research, have already been successfully deployed in clinical practice—even after manual inspection and refinement. These detectors typically maintain a very high recall rate [19,17,1]. In other words, although the detected events may include many false positives, these legacy detectors reliably capture a large share of signals of clinical interest. Furthermore, since the verification of these detected events is often performed through visual inspection of EEG tracing, the definition of clinically meaningful characterization is reflected in the EEG morphology. This insight suggests that clinically relevant biomarkers can be distilled from the output of the rule-based detectors in a self-supervised manner—particularly in scenarios where expert annotations are limited in large datasets or the precise definition of clinically meaningful events remains elusive.

Motivated by this insight, we leverage legacy detectors to identify clinically meaningful events (pathological) without requiring human labeling. Specifically, we focus on high-frequency oscillations (HFOs), a crucial intracranial EEG (iEEG) biomarker used to localize the epileptogenic zone from interictal (in-between seizures) recordings in epilepsy treatment [3]. Importantly, a clear consensus on the definition of pathological HFO is not yet established [24].

To address this challenge, we propose a Self-Supervised to Label Discovery (SS2LD) framework that automatically distills pathological HFOs from the large volume of events detected by legacy HFO detectors. Specifically, we first perform morphological pre-training using a variational autoencoder (VAE) to learn morphological representations of detector-identified events. Leveraging the model’s ability to encode distinct morphological features, we then use the VAE’s latent representations to cluster these events, extracting weak supervision corresponding to pathological events. Finally, we train a classifier on the pre-trained VAE for both feature extraction and data augmentation. This classifier takes latent

codes from the VAE encoder and is trained using both real and VAE-augmented data derived from the discovered weak supervision, thus refining classification boundaries. SS2LD distills the events generated by legacy detectors and improves clinical decision-making. Our contributions are summarized as follows:

- We propose the SS2LD framework that uses VAE-based morphological pre-training to automatically discover pathological HFO events without human labels and exploits these discoveries as weak supervision for classification.
- We employ the VAE’s generative capability to augment training data, enhancing the classifier’s robustness in distilling pathological HFOs.
- SS2LD is evaluated on large multi-institutional iEEG datasets, demonstrating superior clinical performance compared to existing state-of-the-art methods and underscoring its potential on improving clinical decision-making.

## 2 Methods

### 2.1 High-Frequency Oscillations in Intracranial EEG Recordings

For each subject  $k$  in an iEEG cohort, the recording  $\mathcal{X}^k$  comprises multiple channels, indexed by  $c$ , selected clinically to target probable epileptogenic regions. An HFO event  $h_i$  in channel  $c$  of  $\mathcal{X}^k$  is defined by its start time  $s_i$  and end time  $e_i$ , such that  $h_i = \mathcal{X}_c^k[s_i : e_i]$ . These segments may reflect pathological (epileptogenic) or non-pathological (normal) activity. Since HFO durations  $e_i - s_i$  vary, we construct a fixed-length window  $w_i = \mathcal{X}_c^k[\frac{s_i+e_i}{2} - L : \frac{s_i+e_i}{2} + L]$ , centered at the event’s midpoint, to ensure consistent input sizes for neural networks.

### 2.2 Self-supervised Pre-Training

**Variational Autoencoder:** We train a VAE to capture HFOs’ morphological features. The VAE includes a *Time-Frequency Morphological Encoder*  $\mathcal{E}_\theta$  and a *Morphological Decoder*  $\mathcal{D}_\gamma$ . The encoder  $\mathcal{E}_\theta$  transforms an HFO event  $h_i$ , represented as a fixed-length signal  $w_i$ , into a time-frequency representation  $\mathcal{W}_i = \text{MorletTransform}(w_i)$ , then projects it into a lower-dimensional latent space, yielding a distribution parameterized by  $\mu_\theta(w_i)$  and  $\Sigma_\theta(w_i)$ . A latent code  $\mathbf{z}_i \sim \mathcal{N}(\mu_\theta(w_i), \Sigma_\theta(w_i))$  is sampled, and the decoder  $\mathcal{D}_\gamma$  reconstructs  $\hat{\mathcal{W}}_i = \mathcal{D}_\gamma(\mathbf{z}_i)$ , reconstructing the HFO’s morphological feature.

**Encouraging Morphological Understanding:** To measure discrepancies between the real and reconstructed representations, we employ the perceptual loss [6]  $\mathcal{L}_{\text{perceptual}}$ , which uses the  $L_2$  distance between feature maps extracted from different layers of a pre-trained VGG16 network  $\phi(\cdot)$ . Specifically,  $\mathcal{L}_{\text{perceptual}} = \sum_l \|\phi_l(\mathcal{W}_i) - \phi_l(\hat{\mathcal{W}}_i)\|_2^2$ . To further disentangle the latent space and enhance the morphological expressiveness, we adopt a learnable  $\beta$ -VAE strategy. We introduce learnable  $\beta$  on the original VAE loss function and we dynamically adjust

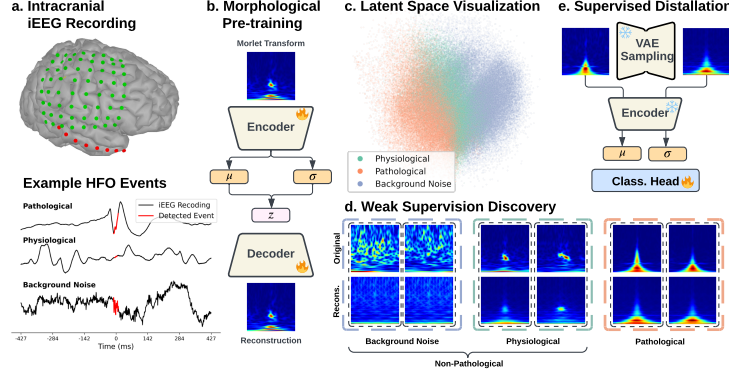


Fig. 1: **Overview of the Proposed SS2LD Framework.** (a) HFO events are detected in interictal (in-between seizures) intracranial EEG recordings using rule-based detectors, including pathological, physiological, and background noise events. (b) A VAE is pre-trained to reconstruct HFO morphological features, derived via Morlet Transform, with the encoder mapping inputs to latent distributions and the decoder reconstructing sampled latent codes. (c) Latent representations are projected by PCA, with a color-coded 2D plot representing clustered event types. (d) Hierarchical clustering identifies distinct morphological groups, enabling weak supervision for classifying pathological events. (e) A classifier is trained on both original latent codes ( $\mu$ ) and VAE-generated samples ( $\hat{\mu}$ ) to further refine the decision boundary.

$\beta$  to balance between reconstruction and disentanglement of latent representations throughout training. Specifically, in each minibatch, the  $\beta$  is updated by:  $\beta_{\text{new}} = \beta + \beta_{\text{lr}} \cdot (\mathbb{E}[\mathcal{L}_{\text{KL}}] - \mathbb{E}[\mathcal{L}_{\text{perceptual}}])$ , where  $\beta_{\text{lr}}$  is a learning rate for  $\beta$ ; and  $\mathcal{L}_{\text{KL}} = \text{KL}(q(z | x) \| p(z))$ , where  $q(z | x)$  is the approximate posterior and  $p(z)$  is the standard normal prior. Finally, the pre-train objective is:  $\mathcal{L}_{\text{pretrain}} = (1 - \beta)\mathcal{L}_{\text{perceptual}} + \beta\mathcal{L}_{\text{KL}}$  ( $\beta \in [0, 1]$ ). Minimizing this loss encourages precise reconstructions and a disentangled, regularized latent representation.

### 2.3 Weak Label Discovery from Structured Latent Space

After pre-training, we aim to gather meaningful insight into the latent representation by visualizing the input signals, their corresponding reconstructed signals, and the associated latent codes within the latent space. In Figure 1, we observe that the entire latent space could be broadly clustered into three morphologies, each with its own clinical significance: (1) pathological, (2) physiological, and (3) background noise. Furthermore, the model’s reconstruction ability for background noise is limited (Figure 1d), likely due to the diverse morphologies of such events. To provide weak supervision, we conduct hierarchical clustering in the latent space. We first fit a k-means cluster with  $k = 2$  across the entire latent space to separate background noise from HFO events, selecting the cluster with the higher reconstruction loss as the background class. Next, we cluster the remaining HFO events (again with  $k = 2$ ) to distinguish pathological from

physiological events. We label the cluster predominantly located within the resected region as pathological. Finally, we assign a weak supervision label  $l = 1$  to pathological events and  $l = 0$  otherwise (physiological and background noise).

## 2.4 VAE-Based Augmentation for Biomarker Distillation

The latent-space visualization suggests that distinct morphological characteristics of the biomarkers are separated within this space. However, since k-means does not account for complex latent geometries, the class boundaries may not conform well to the actual distribution of biomarkers. We, thus, freeze the VAE parameters and train a classification head  $\mathcal{F}_\psi$  on top of the encoder to predict pathological events. By leveraging and distilling the weak supervision introduced by k-means, this training step refines the decision boundary. Furthermore, to improve generalization, we augment the training with VAE-generated surrogates: for each  $\mathcal{W}_i$ , we sample  $\hat{\mathcal{W}}_i$ . We then encode both  $\mathcal{W}_i$  and  $\hat{\mathcal{W}}_i$  to latent vectors  $\mu_i$  and  $\hat{\mu}_i$ . Assuming similar morphology implies the same category, the classifier loss is  $\mathcal{L}_{\text{classifier}} = \text{BCE}(\mathcal{F}_\psi(\mu_i), l_i) + \text{BCE}(\mathcal{F}_\psi(\hat{\mu}_i), l_i)$ , ensuring consistent predictions across original and surrogate samples.

# 3 Results

## 3.1 Dataset

We conduct experiments using the Open iEEG dataset [21], which comprises recordings from 185 epilepsy patients from two institutions (UCLA & Detroit). Surgical outcomes (post-resection seizure freedom) are available for 162 patients (113 seizure-free). Each electrode (channel) is clinically annotated as resected or non-resected, with a total of 686,410 detected HFOs (from two kinds of HFO detector, STE and MNI). We randomly shuffle the subjects and conduct subject-wise fivefold cross-validation. We also extend our experiment using Zurich iEEG dataset [14], which includes 20 epilepsy patients. Following [2], we apply a bipolar montage to the EEG recordings, yielding 15 patients (10 seizure-free) with valid resection annotations (excluding five due to annotation errors of resection margin). We detect 97,511 HFOs (STE & MNI) using the HFO detection parameters in [22]. Since each subject in the Zurich iEEG dataset has multiple runs, we treat all detected events across runs as a single set.

## 3.2 Implementation Details

We implement the VAE using a convolutional neural network with residual layers as the backbone, setting the latent dimension to 16. Each HFO event is represented in a fixed window, producing a time-frequency plot spanning 10–290 Hz over 570 ms, then resized to  $64 \times 64$ . For the Open iEEG dataset, we divide each fold into training ( $n = 119$ ), validation ( $n = 30$ ), and test sets ( $n = 36$ ), ensuring balanced institutional representation (UCLA/Detroit). The VAE is

trained with stratified sampling capping samples per subject at 2,500 per epoch. Pre-training occurs from scratch using the Adam optimizer ( $lr = 10^{-3}$ , weight decay  $\lambda = 10^{-5}$ ) for 100 epochs with a batch size of 512, initializing  $\beta_{\text{init}} = 1$  and  $lr_{\beta} = 10^{-4}$ . For classifier training, we use all training set events per epoch, training with Adam ( $lr = 3 \times 10^{-4}$ ,  $\lambda = 10^{-5}$ ) and batch size of 4096 for nine epochs. All experiments are performed on NVIDIA A6000 GPUs.

### 3.3 Evaluation Metrics

Since there are no explicit ground-truth label for individual HFO, we evaluated our biomarker classifier using clinical evidence. We introduce two metrics to evaluate the effectiveness of the predicted pathological HFOs: ability to predict post-operative seizure outcome and classification specificity.

**Surgical Outcome Prediction:** Predicting surgical outcomes is one of the most promising ways to validate the clinical effectiveness of a biomarker, as it provides a real-world assessment of its prognostic value in guiding epilepsy treatment. A key feature used in this validation is the Resection Ratio. The Resection Ratio ( $RR^k$ ) measures the proportion of pathological HFOs located in surgically resected channels for patient  $k$ , which can be defined as  $RR^k = \frac{\text{Number of pathological HFOs in resected channels}}{\text{Total number of pathological HFOs}}$ , a value of  $RR^k = 1$  indicates complete resection of pathological HFOs, while  $RR^k = 0$  implies no overlap. To predict surgical outcomes (success or failure), we train a logistic regression model with balanced label weights using  $RR^k$  on the training and validation sets, then evaluate it on the test set and Zurich iEEG dataset. Given the class imbalance ( $\sim 2/3$  success), we report both accuracy (ACC) and F1 scores across five folds.

**Classification Specificity:** We leverage another evidence to validate the effectiveness of the predicted pathological HFO. If a patient remains seizure-free after resection, the preserved brain region should contain few or no pathological HFO. We compute the specificity for each surgical success patient  $k$  as  $\text{Specificity}^k = \frac{\text{Number of non-pathological HFO in the preserved region}}{\text{Total number of HFO detected in the preserved region}}$ , where a higher specificity indicates better classification performance. The overall specificity (SPEC) is then defined by averaging across all surgical success patients (across five folds), given by  $\text{SPEC} = \frac{1}{|S|} \sum_{k \in S} \text{Specificity}^k$ , where  $S$  is the set of successful cases.

### 3.4 Comparison with State-of-the-Art Methods

We compare SS2LD against two state-of-the-art pipelines for classifying pathological HFOs, leveraging their publicly available models. These pathological HFOs are defined as: (1) **Epileptogenic HFO (eHFO)** [9], identified via weakly supervised learning rooted in clinical evidence; and (2) **HFO with Spike Discharge (spkHFO)** [22], trained on human-annotated events informed by clinical expertise. Table 1 presents the performance of SS2LD alongside other pipelines, reporting the averaged ACC and F1 scores across five folds,

Table 1: Performance comparison in Open iEEG and Zurich iEEG datasets.

	Open iEEG Dataset			Zurich iEEG Dataset		
	ACC	F1	SPEC	ACC	F1	SPEC
eHFO [9]	$0.538 \pm 0.119$	$0.397 \pm 0.069$	0.543	$0.533 \pm 0.000$	$0.364 \pm 0.000$	0.745
spkHFO [22]	$0.594 \pm 0.129$	$0.434 \pm 0.070$	0.703	$0.600 \pm 0.000$	$0.500 \pm 0.000$	0.819
SS2LD	<b><math>0.612 \pm 0.131</math></b>	<b><math>0.464 \pm 0.069</math></b>	<b>0.749</b>	<b><math>0.640 \pm 0.037</math></b>	<b><math>0.583 \pm 0.050</math></b>	<b>0.909</b>

along with their standard deviations (STD). Specificity is averaged across all patients for the Open iEEG dataset. For the Zurich iEEG dataset, since SS2LD is trained using five-fold cross-validation, specificity is first averaged across patients and then across the five folds. For [9,22], the classifier is trained on distinct patient subsets from the Open iEEG dataset over five folds; however, decision boundary variations do not impact performance on the Zurich iEEG dataset, yielding an STD of zero for ACC and F1 scores. Nevertheless, SS2LD outperforms both in seizure outcome prediction and specificity across all datasets.

**Ablation Studies:** We present ablation studies on key components of the proposed SS2LD: (1) *Weak Supervision*, using weak supervision provided from k-means clustering; (2) *Supervised Distillation*, training the classifier with supervised distillation (SD) but without VAE augmentation (AUG). Additionally, we investigate the impact of latent space dimensions ( $\text{DIM} = 8, 32, \text{and } 64$ ). A latent dimension of 8 appears overly restrictive, limiting the model’s ability to capture critical details of pathological events. While SS2LD achieves higher specificity with latent dimensions above 16, the denser latent details can lead to overfitting and overly conservative predictions (favoring non-pathological labels), potentially reducing its clinical utility. Table 2 highlights the effectiveness of our proposed modules and the selection of latent dimensions.

Table 2: Ablation study results for SS2LD on the Open iEEG dataset.

	SD	AUG	DIM	ACC	F1	SPEC
Weak Supervision			16	$0.557 \pm 0.101$	$0.423 \pm 0.066$	0.739
Supervised Distillation	✓		16	$0.568 \pm 0.116$	$0.433 \pm 0.058$	0.715
SS2LD	✓	✓	8	$0.562 \pm 0.134$	$0.426 \pm 0.065$	0.736
	✓	✓	16	<b><math>0.612 \pm 0.108</math></b>	<b><math>0.464 \pm 0.063</math></b>	0.749
	✓	✓	32	$0.593 \pm 0.128$	$0.442 \pm 0.065$	0.761
	✓	✓	64	$0.574 \pm 0.126$	$0.431 \pm 0.061$	0.752

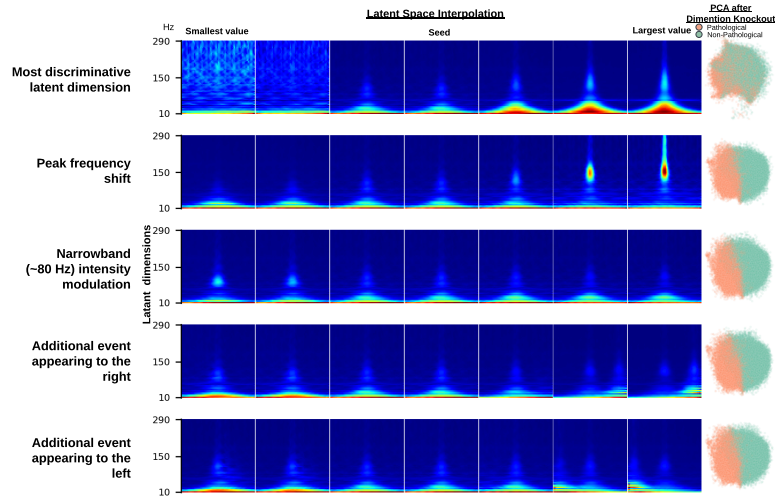


Fig. 2: **Latent Dimension Interpolation and Dimension Knockout:** We visualize the interpolation results focusing on the top 5 most expressive latent dimensions in one example fold. **(Left)** Each row corresponds to a specific latent dimension, annotated with its inferred role in encoding HFO morphology, including variations in spectral intensity across frequency bands (e.g., 80 Hz), peak frequency shifts, and the presence of additional events. The central spectrogram represents the original unperturbed "seed", derived from the mean latent code of predicted pathological events. Notably, the first dimension corresponds to background noise to pathological events. **(Right)** PCA is performed after knocking out each latent dimension to assess its influence on the classifier's decision boundary. Each scatter plot visualizes the distribution of latent embeddings post-knockout, with colors indicating classifier-predicted pathological (orange) or non-pathological (green) events. Overall, the first dimension contributes the most to the final decision boundary.

### 3.5 Neurophysiological Meaningful Latent Space

To investigate the neurophysiological role of each latent dimension, we interpolate the latent space using the mean latent code of predicted pathological events as a seed. Each dimension is adjusted from its smallest (0.001 percentile) to largest (0.999 percentile) value across all latent codes. The VAE decoder then transforms these interpolated latent codes into their corresponding time-frequency representations, enabling us to visualize how variations in each dimension affect the encoded neurophysiological features. Furthermore, to evaluate each dimension's influence on the decision boundary, we conduct latent knockout by setting a dimension to zero, recalculating the PCA, and color-coding results by classifier predictions. Increased mixing of pathological and non-pathological predictions suggests a stronger contribution to the decision boundary. Figure 2 illustrates how distinct latent dimensions encode unique HFO morphological features and affect the decision boundary, demonstrating the meaningful latent space that VAE learned.



## 4 Conclusion

SS2LD introduces a self-supervised approach that outperforms supervised methods while offering explainable decision-making. The VAE pretraining yields a computational latent space that is interpretable via its decoder and could be applicable to various iEEG applications. However, the current model excludes frequency bands outside 10–290 Hz, potentially overlooking informative signals. Future work will extend the evaluation to broader frequency ranges, additional biomarkers beyond HFOs, and more diverse datasets. With these advancements, SS2LD has the potential to be seamlessly integrated into clinical workflows, providing objective, interpretable, and time-efficient pathological HFO predictions to support more effective treatment planning.

**Acknowledgments.** Hiroki Nariai is supported by the National Institute of Neurological Disorders and Stroke (NINDS) K23NS128318, the Sudha Neelakantan Venky Harinarayan Charitable Fund, the Elsie and Isaac Fogelman Endowment, and the UCLA Children’s Discovery and Innovation Institute (CDI) Junior Faculty Career Development Grant (#CDI-TTCF-07012021). Atsuro Daida is supported to research abroad by the Uehara Memorial Foundation and SENSHIN Medical Research Foundation.

**Disclosure of Interests.** The authors have no competing interests in the paper.

## References

1. von Ellenrieder, N., Andrade-Valença, L.P., Dubeau, F., Gotman, J.: Automatic detection of fast oscillations (40–200 Hz) in scalp EEG recordings. *Clinical Neurophysiology* **123**(4), 670–680 (2012)
2. Fedele, T., Burnos, S., Boran, E., Krayenbühl, N., Hilfiker, P., Grunwald, T., Sarnthein, J.: Resection of high frequency oscillations predicts seizure outcome in the individual patient. *Scientific Reports* **7**(1), 13836 (2017)
3. Jacobs, J., Zijlmans, M., Zelman, R., Chatillon, C.É., Hall, J., Olivier, A., Dubeau, F., Gotman, J.: High-frequency electroencephalographic oscillations correlate with outcome of epilepsy surgery. *Annals of Neurology: Official Journal of the American Neurological Association and the Child Neurology Society* **67**(2), 209–220 (2010)
4. Jing, J., Sun, H., Kim, J.A., Herlopian, A., Karakis, I., Ng, M., Halford, J.J., Maus, D., Chan, F., Dolatshahi, M., et al.: Development of expert-level automated detection of epileptiform discharges during electroencephalogram interpretation. *JAMA neurology* **77**(1), 103–108 (2020)
5. Johnson, G.W., Doss, D.J., Morgan, V.L., Paulo, D.L., Cai, L.Y., Shless, J.S., Negi, A.S., Gummadaelli, A., Kang, H., Reddy, S.B., et al.: The interictal suppression hypothesis in focal epilepsy: network-level supporting evidence. *Brain* **146**(7), 2828–2845 (2023)
6. Johnson, J., Alahi, A., Fei-Fei, L.: Perceptual losses for real-time style transfer and super-resolution. In: *Computer Vision—ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part II* 14. pp. 694–711. Springer (2016)
7. Kini, L.G., Bernabei, J.M., Mikhail, F., Hadar, P., Shah, P., Khambhati, A.N., Oechsel, K., Archer, R., Boccanfuso, J., Conrad, E., et al.: Virtual resection predicts surgical outcome for drug-resistant epilepsy. *Brain* **142**(12), 3892–3905 (2019)

8. Ma, K., Lai, D., Chen, Z., Zeng, Z., Zhang, X., Chen, W., Zhang, H.: Automatic detection of high frequency oscillations (80-500hz) based on convolutional neural network in human intracerebral electroencephalogram. In: 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). pp. 5133–5136. IEEE (2019)
9. Monsoor, T., Zhang, Y., Daida, A., Oana, S., Lu, Q., Hussain, S.A., Fallah, A., Sankar, R., Staba, R.J., Speier, W., Roychowdhury, V., Nariai, H.: Optimizing detection and deep learning-based classification of pathological high-frequency oscillations in epilepsy. *Clinical Neurophysiology* pp. 129–140 (10 2023). <https://doi.org/https://doi.org/10.1016/j.clinph.2023.07.012>
10. Nariai, H., Wu, J.Y., Bernardo, D., Fallah, A., Sankar, R., Hussain, S.A.: Inter-rater reliability in visual identification of interictal high-frequency oscillations on electrocorticography and scalp eeg. *Epilepsia open* **3**, 127–132 (2018)
11. Scheuer, M.L., Bagic, A., Wilson, S.B.: Spike detection: Inter-reader agreement and a statistical turing test on a large data set. *Clinical Neurophysiology* **128**(1), 243–250 (2017)
12. Spring, A.M., Pittman, D.J., Aghakhani, Y., Jirsch, J., Pillay, N., Bello-Espinosa, L.E., Josephson, C., Federico, P.: Interrater reliability of visually evaluated high frequency oscillations. *Clinical Neurophysiology* **128**(3), 433–441 (2017)
13. Staba, R.J., Wilson, C.L., Bragin, A., Jhung, D., Fried, I., Engel Jr, J.: High-frequency oscillations recorded in human medial temporal lobe during sleep. *Annals of neurology* **56**(1), 108–115 (2004)
14. T, F., N, K., P, H., Li, A., J., S.: "interictal ieeg during slow-wave sleep with hfo markings" (2023). <https://doi.org/doi:10.18112/openneuro.ds003498.v1.1.1>
15. Uda, H., Kuroda, N., Firestone, E., Ueda, R., Sakakura, K., Kitazawa, Y., Choromanski, D., Cools, M., Luat, A.F., Asano, E.: Normative atlases of high-frequency oscillation and spike rates under sevoflurane anesthesia. *Clinical Neurophysiology* **167**, 117–130 (2024)
16. Vallat, R., Walker, M.P.: An open-source, high-performance tool for automated sleep staging. *Elife* **10**, e70092 (2021)
17. Warby, S.C., Wendt, S.L., Welinder, P., Munk, E.G., Carrillo, O., Sorensen, H.B., Jennum, P., Peppard, P.E., Perona, P., Mignot, E.: Sleep-spindle detection: crowd-sourcing and evaluating performance of experts, non-experts and automated methods. *Nature methods* **11**(4), 385–392 (2014)
18. Xu, Z., Wang, T., Cao, J., Bao, Z., Jiang, T., Gao, F.: Bect spike detection based on novel eeg sequence features and lstm algorithms. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* **29**, 1734–1743 (2021)
19. Zelmann, R., Mari, F., Jacobs, J., Zijlmans, M., Dubeau, F., Gotman, J.: A comparison between detectors of high frequency oscillations. *Clinical Neurophysiology* **123**(1), 106–116 (2012)
20. Zelmann, R., Mari, F., Jacobs, J., Zijlmans, M., Chander, R., Gotman, J.: Automatic detector of high frequency oscillations for human recordings with macro-electrodes. In: 2010 Annual International Conference of the IEEE Engineering in Medicine and Biology. pp. 2329–2333. IEEE (2010)
21. Zhang, Y., Daida, A., Liu, L., Kuroda, N., Ding, Y., Oana, S., Monsoor, T., Duan, C., Hussain, S.A., Qiao, J.X., Salamon, N., Fallah, A., Sim, M.S., Sankar, R., Staba, R.J., Jr., J.E., Asano, E., Roychowdhury, V., Nariai, H.: "open ieeg dataset" (2024). <https://doi.org/doi:10.18112/openneuro.ds005398.v1.0.1>
22. Zhang, Y., Liu, L., Ding, Y., Chen, X., Monsoor, T., Daida, A., Oana, S., Hussain, S., Sankar, R., Fallah, A., et al.: Pyhfo: lightweight deep learning-powered end-to-

- end high-frequency oscillations analysis application. *Journal of Neural Engineering* **21**(3), 036023 (2024)
23. Zhang, Y., Lu, Q., Monsoor, T., Hussain, S.A., Qiao, J.X., Salamon, N., Fallah, A., Sim, M.S., Asano, E., Sankar, R., et al.: Refining epileptogenic high-frequency oscillations using deep learning: a reverse engineering approach. *Brain communications* **4**(1), fcab267 (2022)
  24. Zijlmans, M., Jiruska, P., Zelmann, R., Leijten, F.S., Jefferys, J.G., Gotman, J.: High-frequency oscillations as a new biomarker in epilepsy. *Annals of neurology* **71**(2), 169–178 (2012)
  25. Zweiphenning, W., van’t Klooster, M.A., van Klink, N.E., Leijten, F.S., Ferrier, C.H., Gebbink, T., Huiskamp, G., van Zandvoort, M.J., van Schooneveld, M.M., Bourez, M., et al.: Intraoperative electrocorticography using high-frequency oscillations or spikes to tailor epilepsy surgery in the netherlands (the hfo trial): a randomised, single-blind, adaptive non-inferiority trial. *The Lancet Neurology* **21**(11), 982–993 (2022)