ICLR 2026

Omni-iEEG: A Large-Scale, Comprehensive iEEG Dataset and Benchmark for Epilepsy Research

Chenda Duan, Yipeng Zhang, Sotaro Kanai, Yuanyi Ding, Atsuro Daida, Pengyue Yu, Tiancheng Zheng, Naoto Kuroda, Shaun A. Hussain, Eishi Asano, Hiroki Nariai, Vwani Roychowdhury

UCLA Samueli School of Engineering | UCLA Mattel Children's Hospital, David Geffen School of Medicine | Children's Hospital of Michigan, Wayne State University School of Medicine

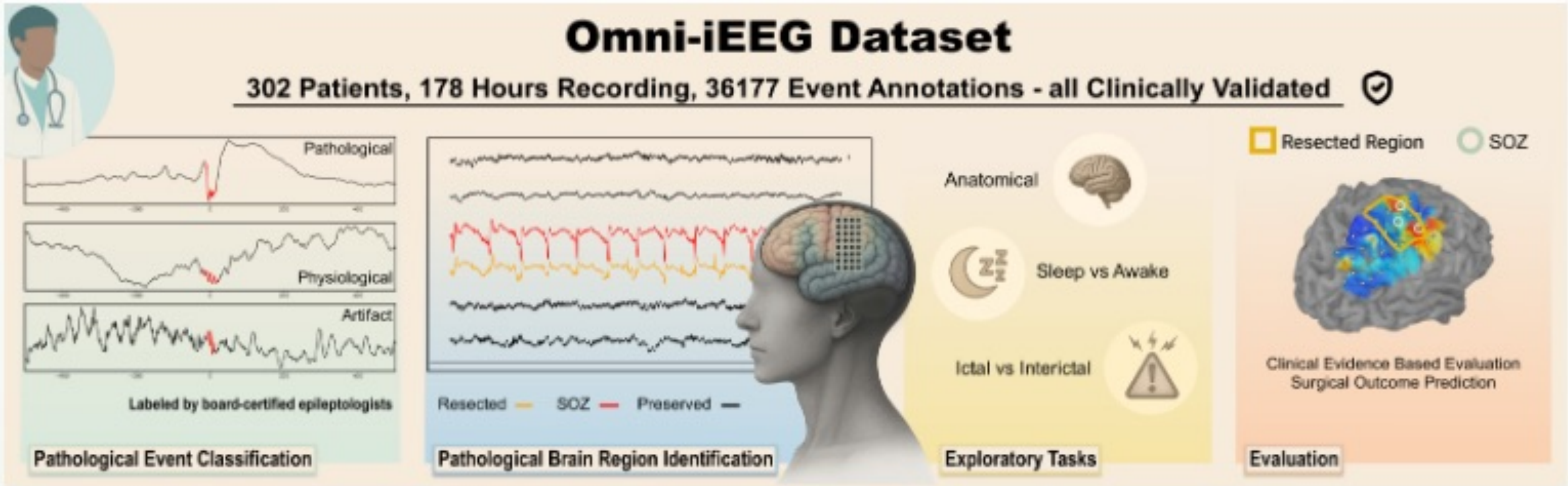
Large-Scale, Expert-Curated Dataset

Expert-Labeled HFO Events

Clinically Grounded Benchmark and Baselines

Translational Insights

302	**178h**	**36,177**
PATIENTS	PRE-SURGICAL IEEG	EXPERT HFO ANNOTATIONS



PROJECT PAGE

omni-ieeg.github.io/omni-ieeg

Data + code

The Bottleneck in Data-Driven Epilepsy Research

Why seizure localization matters

Epilepsy affects more than 50 million people worldwide, and about one-third of patients are drug-resistant. For these patients, seizure freedom often depends on accurately localizing the Epileptogenic Zone (EZ) from intracranial EEG to guide resection or neurostimulation.

50M+
PEOPLE AFFECTED WORLDWIDE

~1/3
DRUG-RESISTANT EPILEPSY

Seizure freedom
DEPENDS ON ACCURATE EZ LOCALIZATION

Why ML translation stalls

- Most public iEEG studies remain single-institution and difficult to validate across centers.
- Released cohorts are heterogeneous in file format, metadata schema, channel naming, SOZ labels, and surgical outcome annotation.
- Expert-labeled pathological biomarker events are rarely shared, limiting supervision for clinically meaningful models.

What is still missing

- A fair cross-center benchmark for event-level and channel-level pathology identification.
- Evaluation protocols tied to clinical evidence and postoperative outcome, not just generic signal classification.
- A public testbed for end-to-end long-context models and cross-domain transfer learning.

Core need: a harmonized, multi-center iEEG resource with expert biomarker labels, unified clinical metadata, and evaluation grounded in real surgical decision-making.

MANUAL REVIEW + FRAGMENTED PUBLIC DATA -> LIMITED CLINICAL TRANSLATION

RESOURCE

Omni-iEEG: A Foundation for Machine Learning in Epilepsy

Core resource contribution

- Large-scale pre-surgical iEEG resource aggregated across eight epilepsy centers.
- Clinically harmonized metadata and expert-labeled biomarker events released in one public benchmark.

COHORT

302

patients aggregated from multiple public releases across 8 leading epilepsy centers.

RECORDING SCALE

178h

of pre-surgical high-resolution iEEG curated into a standardized, reusable resource.

CLINICAL METADATA

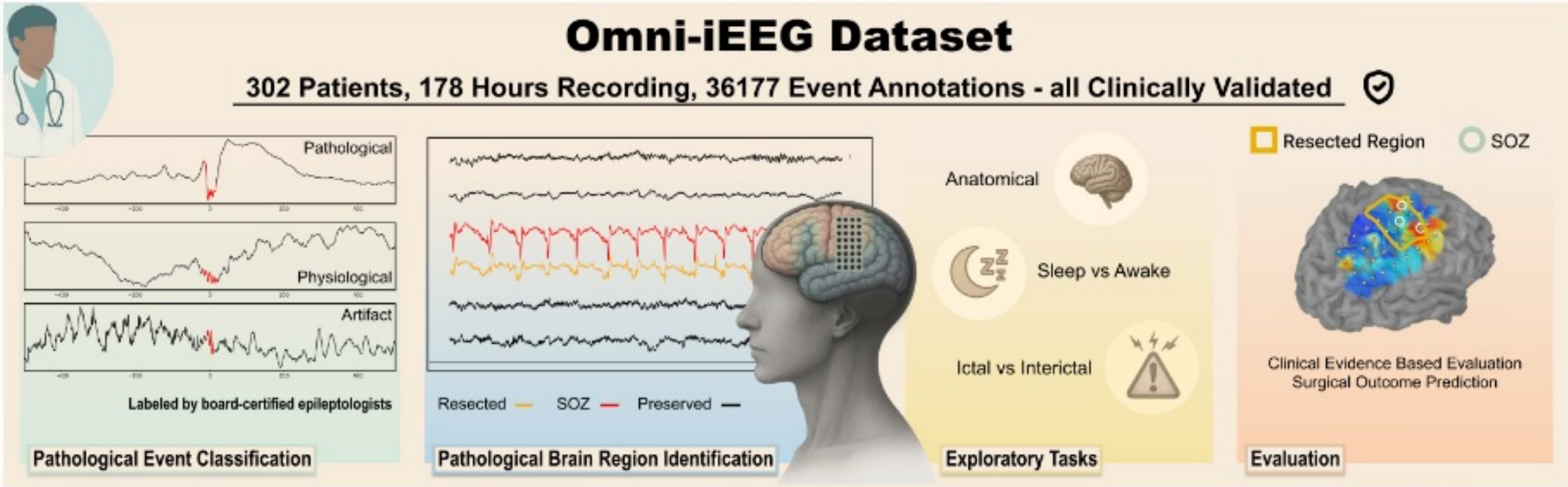
SOZ, Resection, Outcome

validated channel- and patient-level labels verified by board-certified epileptologists.

EXPERT LABELS

36,177 HFO Events

expert-labeled biomarker events spanning artifact, non-spkHFO, and spkHFO classes.



Overview figure from the paper/poster showing the downstream event-level, channel-level, and exploratory benchmarks enabled by the Omni-iEEG resource.

Clinically Grounded Benchmark Tasks

TASK 1

Pathological Event Classification

3-way classification of HFO events into artifact, physiological/non-spkHFO, or pathological/spkHFO.

Metrics: macro Precision, Recall, F1, and one-vs-rest macro AUC.

TASK 2

Pathological Channel Identification

Positive channels = SOZ. Negative channels = preserved channels from seizure-free patients.

Metrics: Precision, Recall, Specificity, and AUC.

CLINICAL ENDPOINT

Outcome Prediction via Resection Ratio

Given channel-level pathological scores, we compute a patient-level RR and test whether higher RR correlates with postoperative seizure freedom.

Metric: ROC-AUC on seizure-free vs non-seizure-free patients.

Clinical endpoint definition

RR is the patient-level quantity used for outcome evaluation: it measures how much predicted pathology falls inside the resected region.

$$RR = \frac{\sum_{c \in \text{resected}} s_c}{\sum_{c \in \text{all}} s_c}$$

Higher RR should correlate with seizure freedom; ROC-AUC is then computed on seizure-free vs non-seizure-free outcome labels.

Exploratory tasks

Beyond the two primary tasks, Omni-iEEG also includes three exploratory tasks to broaden the benchmark scope:

Anatomical Region Classification
Predict frontal, temporal, parietal, limbic, or occipital location from interictal segments.

Ictal vs Interictal Detection
Distinguish seizure-period recordings from non-seizure periods for automated seizure identification.

Sleep vs Awake Classification
Model vigilance state to separate physiological context from pathological activity.

EVALUATION IS TIED TO CLINICAL EVIDENCE, NOT JUST GENERIC SIGNAL CLASSIFICATION

Architectural Baselines: Events vs Segments

EVENT-BASED MODELS

Clinical priors over discrete HFO events

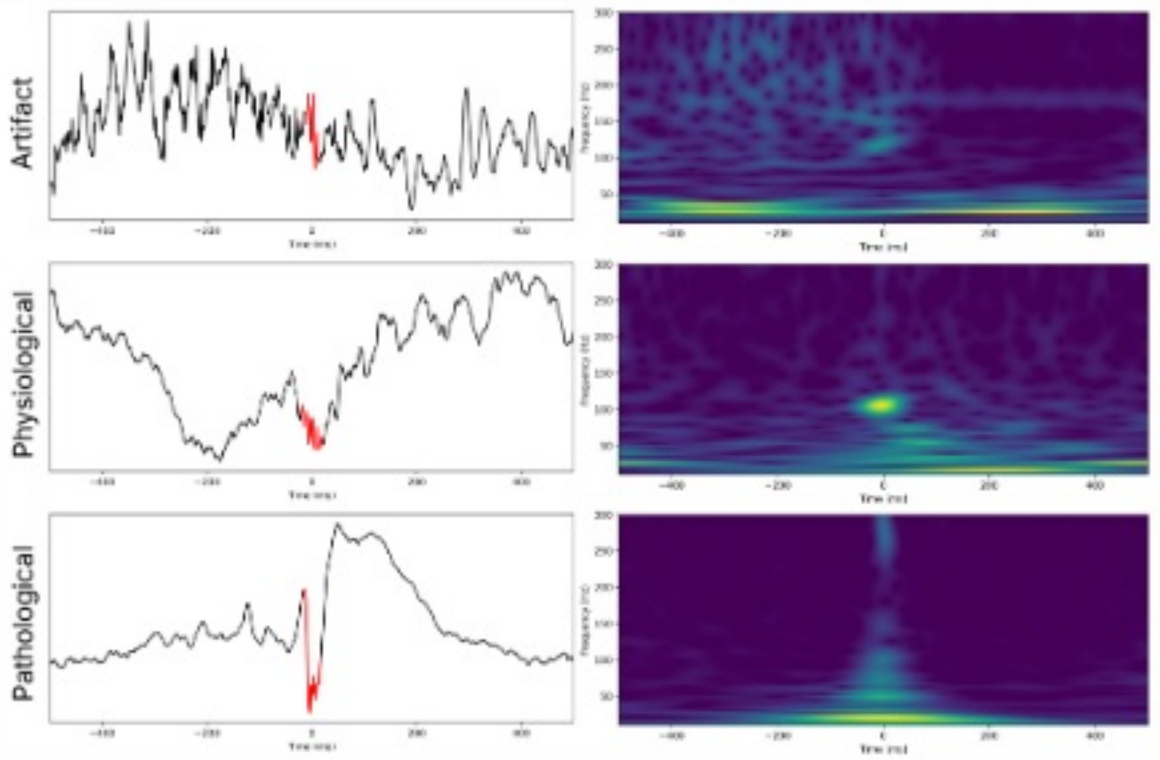
First detect candidate HFOs, then classify pathological events, and finally aggregate event rates into a channel-level pathological score.

INPUT
Short HFO-centered event windows with waveform and time-frequency structure.

REPRESENTATIVE BASELINE
PYHFO-OMNI, extending a clinically motivated HFO pipeline.

PIPELINE
Detection -> event classification -> channel aggregation.

KEY ASSUMPTION
Pathological channels should exhibit elevated rates of pathological HFO events.



Event-level representation used for pathological event classification: short HFO windows with morphology and frequency signatures.

SEGMENT-BASED MODELS

End-to-end learning from minute-long iEEG

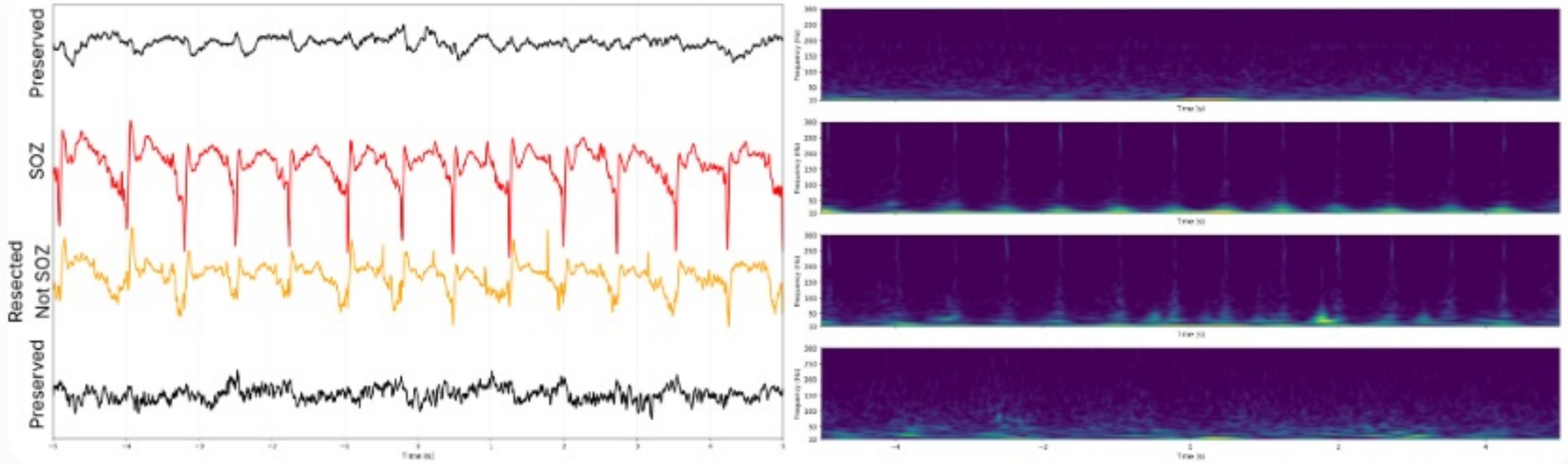
Directly model 1-minute, 1000 Hz interictal segments without explicit event detection, using broader temporal and spectral context for channel pathology prediction.

INPUT
Minute-long spectrotemporal representations capturing long-range interictal dynamics.

REPRESENTATIVE MODELS
TIMECONV-CNN and **CLAP** for end-to-end and cross-domain transfer learning.

CHANNEL LABELS
Positive = SOZ channels; Negative = preserved channels from seizure-free patients.

WHY IT MATTERS
Tests whether long-context signals contain predictive pathology beyond hand-crafted event counting.



Segment-level representation for pathological channel identification: larger context, richer dynamics, and compatibility with TimeConv-CNN and CLAP.

The benchmark is designed to compare two distinct modeling regimes: clinically grounded biomarker pipelines versus end-to-end long-context models that may discover additional pathological structure.

Benchmark Results

Pathological Event Classification

Macro-averaged Precision, Recall, F1, and one-vs-rest AUC on the 3-way HFO event task.

MODEL	PRECISION	RECALL	F1	AUC
LSTM+Attn	0.735	0.736	0.734	0.911
PatchTST	0.776	0.769	0.773	0.931
TimesNet	0.759	0.773	0.765	0.922
PyHFO-Omni	**0.803**	**0.811**	**0.806**	**0.939**

Best numbers are highlighted. PyHFO-Omni is the strongest event-level baseline.

Pathological Channel ID + Outcome Prediction

Channel metrics use SOZ vs preserved seizure-free channels; outcome AUC evaluates whether resection-ratio-based predictions align with seizure freedom.

MODEL	PRECISION	RECALL	F1	SPECIFICITY	CHANNEL AUC	OUTCOME AUC
EVENT-BASED MODELS						
eHFO	0.605	0.647	0.620	0.410	0.661	0.452
PyHFO_spkHFO	0.600	0.643	0.614	0.409	0.656	0.497
PyHFO-Omni_spkHFO	0.580	0.699	0.564	0.695	0.735	**0.744**
SEGMENT-BASED MODELS						
SEEG-NET	0.579	0.717	0.526	0.605	0.785	0.595
CLAP	0.594	0.700	0.601	0.782	0.768	0.677
TimeConv-CNN	**0.626**	**0.745**	**0.647**	**0.823**	**0.806**	0.738

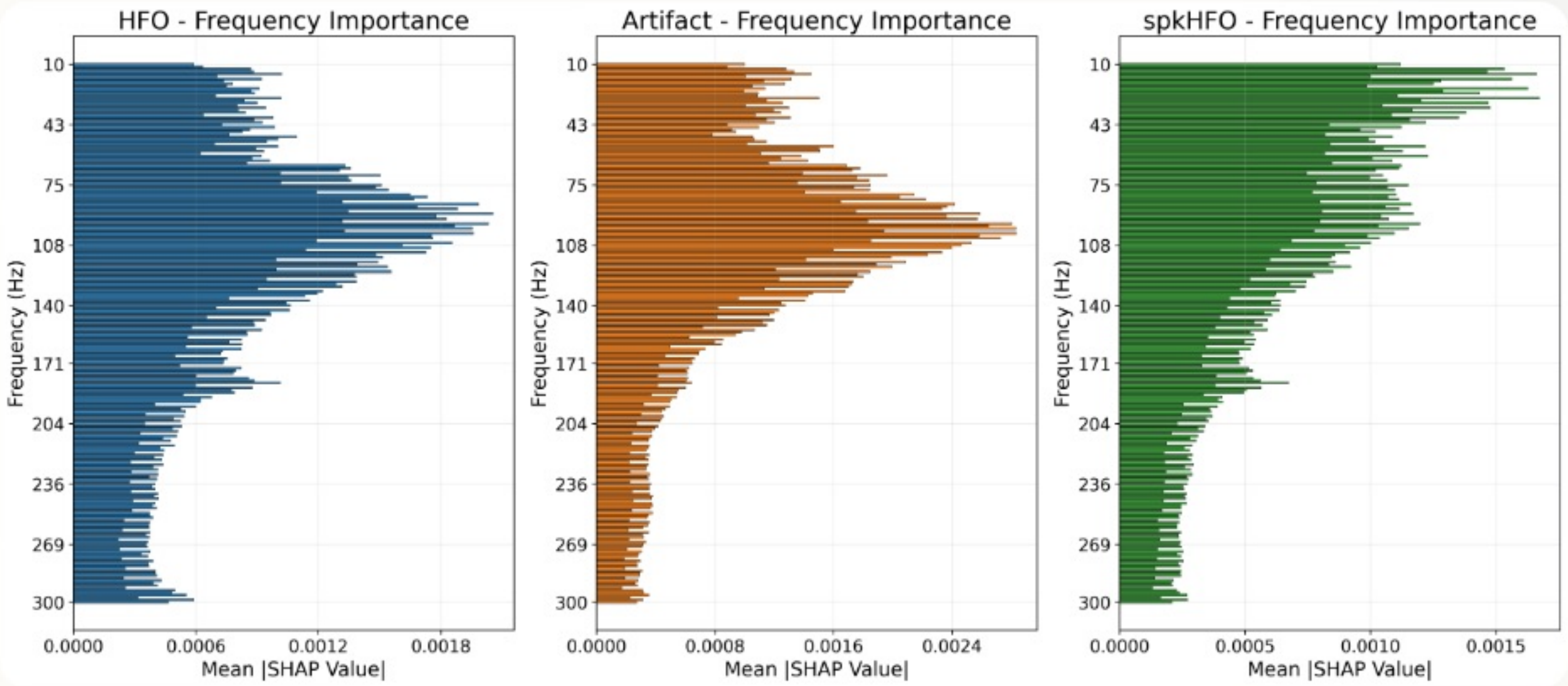
FINDINGS

Key Findings and Translational Insights

Main takeaways

- **PyHFO-Omni** is the strongest event-level baseline, showing that clinically grounded HFO pipelines remain highly competitive.
- **TimeConv-CNN** achieves the best channel identification performance and nearly matches the best outcome AUC.
- **CLAP** reaches competitive channel AUC, suggesting that audio-pretrained representations transfer effectively to iEEG.
- An exploratory **YAMNet** case study found 87 SOZ segments labeled as "helicopter" and 0 preserved segments in the same patient.

Interpretability matters: SHAP emphasizes the 10-30 Hz band, aligning with clinically recognized spike morphology instead of suggesting spurious high-frequency shortcuts.



Frequency importance summary from the paper: low-frequency spike-related structure remains highly informative, especially for spkHFO-related predictions.