

ABSTRACT BOOK

P01
Applying Machine Learning to the Hemispheric Surgery Outcome Prediction Scale (HOPS)

A. Weil[1], S.O. Laperriere[1], A. Hadjinicolaou[1], A. Fallah[2], H. Hops_collaborators[1]

[1]Université De Montréal - Montreal (Canada), [2]University Of California, Los Angeles - Los Angeles (Canada)

Introduction

Cerebral hemispheric surgeries (hemispherectomy and hemispherotomy) are established treatments for children with drug-resistant epilepsy, providing the possibility of seizure freedom in approximately two-thirds of patients. Despite their effectiveness, these surgeries carry significant risks and long-term neurological consequences, making accurate patient selection and counseling essential. The Hemispheric Surgery Outcome Prediction Scale (HOPS) was developed to meet this need, drawing on data from 1,267 pediatric hemispheric surgeries performed across 32 centers in 12 countries. The original HOPS study used multivariate logistic regression to identify independent predictors of seizure freedom and constructed a clinical prediction tool with an area under the curve (AUC) of 0.72. Key predictors included age at seizure onset, presence of generalized seizure semiology, presence of contralateral PET hypometabolism, underlying etiologic substrate, and history of prior non-hemispheric resective surgery. While the HOPS tool represented an important step forward in providing evidence-based prognostic information, its predictive performance leaves room for improvement. The objective of this study was to apply modern machine learning (ML) methods to the original HOPS dataset to improve predictive accuracy for seizure freedom following hemispherectomy.

Methods

The original HOPS dataset was split into training and test subsets. A series of machine learning algorithms were evaluated, including Random Forest, XGBoost, Multi-Layer Perceptron (MLP), and logistic regression (the baseline comparator). Models were trained using 10-fold cross-validation to ensure robustness and reduce overfitting. Predictor variables included the clinical and imaging features identified in the original HOPS study, and performance was assessed using AUC as the primary metric. The goal was to determine whether more flexible, nonlinear models could capture complex interactions between predictors that might not be adequately represented by logistic regression.

Results

Tree-based ensemble methods demonstrated superior performance compared to both logistic regression and neural network models. Random Forest and XGBoost achieved approximately a 10% improvement in AUC relative to logistic regression, indicating a markedly stronger ability to discriminate seizure-free from non–seizure-free patients at follow-up. The MLP, while more flexible in theory, performed less consistently, likely reflecting the modest dataset size relative to its parameter complexity. These findings suggest that tree-based ML models are particularly well suited to clinical prediction tasks in heterogeneous, multicenter datasets such as HOPS.

Conclusions

The original HOPS model was a valuable step toward outcome prediction after hemispheric surgery, but its moderate accuracy (AUC ≈ 0.72) reflected the limits of linear regression in capturing complex clinical interactions. By applying machine learning techniques—particularly Random Forest and XGBoost—we achieved significantly higher predictive accuracy while preserving clinical interpretability. This refinement of the HOPS framework has important clinical implications: it can improve preoperative counseling, guide surgical decision-making, and help identify patients less likely to benefit from surgery. Looking forward, external validation, integration into clinical platforms, and incorporation of additional predictors such as advanced neuroimaging will be essential. Ultimately, this study demonstrates the potential of machine learning to transform predictive modeling in epilepsy surgery and underscores the value of multicenter collaborations to advance patient care.

P02

Perioperative use of P-glycoprotein inhibitors has positive effect for seizure outcomes after epilepsy surgery

J. Shen[1]

[1]Cleveland Clinic - Cleveland (United States)

Introduction

One third of patients with epilepsy have drug-resistant epilepsy (DRE) and continue to have seizures despite medical therapy. Resective epilepsy surgeries offer the best chances for sustained seizure-freedom in patients with DRE, but surgical outcomes can vary. With seizures recurring in up to 50% of patients, there is an unmet need to better understand the drivers of postoperative seizure recurrence.

Recent data suggest that variation in the ABCB1 gene may influence seizure outcomes of resective epilepsy surgery, with overexpression of P-Glycoprotein (P-gp) encoded by ABCB1 potentially explaining higher rates of recurrence. This study aims to obtain real-world evidence to evaluate this hypothesis by testing the relationships between peri-operative use of P-gp inhibitors and seizure outcomes of epilepsy surgery.

Methods

This study included 717 adult patients with drug-resistant epilepsy who underwent resective brain surgery at the Cleveland Clinic from 2010 to 2022. Patients were divided into treatment and control groups based on their use of seven P-gp inhibitors within 14 days of surgeries. Patient information including demographics, clinical features, EEG, imaging tests, and seizure outcomes was collected from the Cleveland Clinic Epilepsy Center Outcomes Research Registry. Peri-operative P-gp inhibitor usage data and history of medications and comorbidities were extracted from Cleveland Clinic's electronic health records from the Epic® clinical information system.

We applied a machine learning enhanced causal inference approach to analyze the target P-gp inhibitors. For each drug, we generated balanced cohorts by excluding samples violating the positivity assumption, fitting a machine learning model to reweight the samples, and evaluating covariate balance after reweighting. We then calculated the hazard ratios for seizure recurrence per drug and conducted a meta-analysis to synthesize these results for overall effect evaluation. We also conducted a time-to-event analysis for seizure recurrence using the Kaplan Meier estimators. Model evaluation was done through calculating the absolute standardized mean difference (ASMD) and plotting the calibration, propensity distribution, and area under the receiver operating characteristics (AUROC) curves. Heterogeneity was assessed across analyses using the I^2 Statistic.

Results

We tuned our machine learning model to generate balanced sub-cohorts for the causal analyses. The ASMD difference after reweighting was reduced to below the threshold of 0.1 for all features.

The propensity distribution plot, calibration curve, and AUROC curve together showed that treatment assignments were close to random after the weighting.

From the causal inference and meta-analysis, we found that the random-effects model hazard ratio for seizure recurrence associated with the use of P-gp inhibitor drugs was 0.80 (95% CI [0.64-0.99], P=0.0399). The corresponding hazard ratio using a fixed-effect model was 0.82 (95% CI [0.73-0.92], P=0.0007). Both models demonstrated a statistically significant reduction in the risk of seizure recurrence with the use of P-gp inhibitors. The cross-study heterogeneity quantified by the inconsistency index I^2 is 40.3% (95% CI [0.0%-74.9 %], P=0.12), which indicated negative evidence of heterogeneity. Taken together, these results indicate that seizure recurrence risk is reduced by approximately 20% with the use of P-gp inhibitors, and this effect is based on a sufficiently homogeneous population.

Results from the time-to-event analysis further supported the association of P-gp inhibitors with reduced risk of seizure recurrence. Recurrence rate was significantly lower for patients who were taking P-gp inhibitors (HR=0.782, 95% CI [0.657-0.929], P=0.0052]. The covariate-adjusted survival curves with bootstrapped 95% confidence interval showed a clear separation between the treated and control groups, with the treated group survival rates consistently higher over time.

Conclusions

The peri-operative use of P-gp inhibitors in drug-resistant epilepsy may positively influence seizure outcomes. Additional validation in prospective cohorts is warranted to determine if P-gp inhibitors could provide a minimally invasive approach to improve surgical outcomes. In summary, the key points of this study are: (1) Abnormal P-gp expression in neuronal cells can contribute to epileptogenesis; (2) P-gp expression could represent a marker for recurrence of seizures postoperatively; (3) The use of P-gp inhibitors perioperatively may have additional benefits for patients who undergo epilepsy surgeries; (4) Machine learning and statistical analysis methods applied to large-scale electronic health records data may lead to the discovery of positive effects on seizure freedom from a wider range of P-gp inhibitors.

P03
Intracranial EEG Correlates of Cortical Thinning in Drug-Resistant Focal Epilepsy

L. Jürgensen[1, 2], P. Müller[1, 2], R. Terziev[1, 2], C. Meisel[1, 2]

[1]Charité-Universitätsmedizin Berlin - Berlin (Germany), [2]Berlin Institute For Health - Berlin (Germany)

Introduction

Drug-resistant focal epilepsy is characterized by both structural brain alterations and abnormal neuronal dynamics, contributing to cognitive impairment and poor clinical outcomes. Regional cortical thinning has been linked to disease burden, yet how such structural changes relate to intracranial electrophysiological markers of network function — and how both jointly inform cognitive outcome — remains unclear. Measures of critical brain dynamics, such as temporal correlations (TCs) in intracranial EEG (iEEG), reflect the temporal persistence of neural activity, serve as a hallmark of near-critical network dynamics, and index the brain's capacity to maintain information over time. We investigated whether regional cortical thinning is associated with systematic alterations in iEEG-derived markers of network dynamics in patients with drug-resistant focal epilepsy and whether combining structural and dynamic features improves prediction of cognitive performance.

Methods

We analyzed multiday intracranial EEG recordings from 40 patients undergoing invasive monitoring at a single center. Electrode locations were corrected for brain shift, transformed into MNI space, and assigned to cortical regions using the HCP-MMP1 (Glasser) atlas. Regional cortical thickness was extracted from corresponding MRI parcellations. From iEEG data, we computed temporal correlations alongside frequency-band power, interictal epileptiform discharge (IED) rates, automatically detected using an established deep learning–based algorithm, and slow-wave sleep (SWS) occurrence, and assessed their interactions.
To evaluate the predictive relevance of these features, we additionally trained supervised machine learning models to predict patient-level cognitive performance using (i) structural features alone (cortical thickness), (ii) electrophysiological features alone (iEEG-derived network dynamics), and (iii) combined structural–dynamic feature sets, enabling a direct comparison of their predictive contributions.

Results

Temporal correlations exhibited a hierarchical organization across cortical regions, consistent with established functional hierarchies of the human cortex, and were systematically modulated by sleep stage and antiseizure medication load. Focusing on structure–dynamics relationships, region-wise analyses across patients revealed a consistent positive association between cortical thickness and temporal correlations (mean slope = 0.29, $p = 0.006$). In contrast, within-patient analyses across regions did not show a consistent relationship, suggesting substantial spatial

heterogeneity in local structure–dynamics coupling and a dominant contribution of inter-individual disease burden.

Conclusions

Together, these findings link epilepsy-related cortical thinning to deviations from near-critical network dynamics and provide a mechanistic framework bridging structural imaging markers with electrophysiological dysfunction in drug-resistant focal epilepsy. Ongoing work evaluates whether integrating structural and dynamic features improves prediction of cognitive performance compared to either modality alone, aiming to inform multimodal biomarkers of disease burden and functional outcome.

P04
HFOs, VHFOs and beyond

M. Brazdil[1]

[1]Masaryk University, 1 St Department Of Neurology - Brno (Czechia)

Introduction

Human brain cell assemblies fire electrical impulses up to very high frequencies, which limit is not yet known.

Methods

Using advanced intracranial microEEG recordings in a cohort of epileptic patients, we newly identified short-lasting oscillations in frequencies between 2 and 8 kHz.

Results

These ultra fast oscillations were consistently and locally detected in epileptic hippocampi but were extremely rare in non-epileptic brain tissue.

Conclusions

The discovered electrophysiological phenomena thus may reflect neuronal hyperexcitability.

P05
Enhancing Infantile Spasms Diagnosis Through Remote EEG: A PPIE Approach

I. Peh[1], P. Natarajan[2], L. Smith[1], B. Chybowski[3, 4], A. Gonzalez Sulser[3, 5], J. Rowling[6], U. United Kingdom Infantile Spasms Trust[7], J. Escudero[4, 8], J. Shetty[1, 9]

[1]Child Life And Health, Cir, Deanery Of Clinical Sciences, School Of Medicine, University Of Edinburgh - Edinburgh (United Kingdom), [2]Medical School, University Of Edinburgh - Edinburgh (United Kingdom), [3]Muir Maxwell Epilepsy Centre, University Of Edinburgh - Edinburgh (United Kingdom), [4]Institute For Imaging, Data And Communications, School Of Engineering, University Of Edinburgh - Edinburgh (United Kingdom), [5]Centre For Discovery Brain Sciences, School Of Medicine, University Of Edinburgh - Edinburgh (United Kingdom), [6]United Kingdrom Infantile Spasms Trust - Bristol (United Kingdom), [7]United Kingdom Infantile Spasms Trust - Bristol (United Kingdom), [8]Muir Maxwell Epilepsy Centre, University Of Edinburgh - Edinburgh (United Kingdom), [9]Royal Hospital For Children And Young People, Nhs Lothian - Edinburgh (United Kingdom)

Introduction

Infantile spasms (IS), the most common severe form of epilepsy in infants, has an incidence of 0.5–3 per 1,000 live births. A delay in treating IS is associated with reduced developmental scores at age four, with a linear decline in developmental scores with increased treatment delays. Early intervention and prompt cessation of spasms minimises the duration of epileptic encephalopathy. Additionally, there appears to be no threshold effect; treatment urgency arises immediately, without a preceding buffer period and a treatment delay of more than one week is linked to poorer long-term neurodevelopmental outcomes.

The current pathway for diagnosis and monitoring of IS is often long, complicated and prone to delays. Additional barriers arise from access inequality for families living in remote areas and/or resource-limited settings. The timely availability of an EEG, hospital beds, and specialist clinicians required to review the data obtained, is variable.

The EPIC Project (‘Enabling the Early and Equitable Diagnosis of Epilepsy in Infants in the Community’) aims to address the question of whether signal analysis and causal artificial intelligence (AI) can identify seizures in infants and predict outcomes from EEGs, addressing barriers to early diagnosis. Initial feasibility testing was supported by the Harmonised IAA grant at the University of Edinburgh.

The team then aims to co-develop a remote EEG device, embedded with the automatic seizure detection algorithm, that would allow families to accelerate referral to a specialist through community-based screening of cases based on EEG. The EPIC Project has been awarded a multidisciplinary >£2m grant by the UKRI EPSRC for a duration of three years.

Importantly, the concept of a remote EEG device, designed to enable automatic detection of epilepsy and facilitate timely referral to specialist hospitals that underpins EPIC, originated from parents. Here we report our initial Patient Public Involvement and Engagement (PPIE) work that shaped our project and describe our future plans.

Methods

An initial survey was conducted with families who had lived experience of an IS diagnosis. The survey was distributed through the UK Infantile Spasms Trust (UKIST) and to parents attending the Edinburgh Children's hospital (February 2024 - April 2025). This survey included questions (structured and free text) regarding their opinions of the healthcare setting surrounding their child's diagnosis, in particular, the EEG testing. Parent's preferences for the digital solution to improve the diagnostic procedure were also asked. Free text information was analysed thematically.

Results

Of 107 respondents, a third had a diagnostic delay greater than 30 days. We obtained rich qualitative data that revealed recurring themes regarding the factors parents attribute to delayed diagnosis. These factors include the dismissal of parental concerns, misdiagnosis, limited availability of EEG facilities, and inadequate access to specialists or paediatric neurologists with the expertise required to identify the condition. Facilitators that contributed to a timely diagnosis frequently involved actions that imposed a burden on the families or are difficult to control for. These included, but were not limited to, explicitly identifying the suspected condition (infantile spasms), obtaining a clear video recording of a seizure episode, or having a close family member/friend working in the healthcare sector.

The desired features of a remote EEG device included a compact, quiet and lightweight device, with the ability for longer recording while maintaining daily routine and maximising the child's comfort. Some specific suggestions included adjustable electrodes or a wireless cap that would require a single fitting process, as well as a signal to inform the families the device has been fitted properly, and that EEG recording has commenced.

This work is ongoing and will continue to grow throughout the project. 47% of respondents expressed a desire to participate in a focus group to help shape our research, with seventeen parents emailing us their contact details directly. The team is currently co-designing a survey for healthcare professionals regarding their expectations and suggestions for a remote healthcare solution, as well as how it can be effectively implemented within the community. An additional family survey is being co-produced that will allow us to confirm whether there are specific subgroups of families that experience longer diagnostic delays and the burden families face as a result. In addition to these findings, we are planning workshops aimed at enhancing awareness of infantile spasms among families and healthcare professionals. Furthermore, we are designing engagement events that will provide children with opportunities to share their perspectives on the diagnostic device and suggest ways to improve the diagnostic process. These events will also serve to educate children about the stigma surrounding epilepsy and actions they can take to help if they see someone having an epileptic seizure.

Conclusions

PPIE is embedded throughout the EPIC Project. The EPIC team will work with families to co-create and deliver a remote EEG solution. The device will be used in the community and will provide a fair and equitable solution for personalised early diagnosis in infantile epilepsies. Our initial PPIE work confirms there is often a delay in diagnosing IS, placing a significant burden on the families. Parents also feel there is a clear need for a better solution that must be co-developed with individuals who have lived experience. There is a great willingness from families to be involved and turn their, potentially negative experiences, into a positive scientific outcome. The themes derived from the involvement of those directly affected by the conditions, will be taken forward for designing our EPIC study and EEG device, supported by the >£2m UKRI EPSRC grant (UKRI1659).

P06

Harnessing automated spike detections to tailor length of stay in the Epilepsy Monitoring Unit

J. Kim[1], C. Aguila[1], W. Ojemann[1], T. Adenekan[1], J. Asuncion[1], J. Jing[2], M. Westover[2], K. Davis[1], B. Litt[1], E. Conrad[1]

[1]University Of Pennsylvania - Philadelphia (United States), [2]Massachusetts General Hospital - Boston (United States)

Introduction

Temporal lobe epilepsy (TLE) is the most common localization of drug-resistant epilepsy in adults. Identifying whether TLE is localized to the left, right, or both temporal lobes is crucial for directing effective surgical treatment. Clinicians typically admit patients to the epilepsy monitoring unit (EMU) to capture multiple spontaneous seizures, but this process may not be sufficient to exclude bilateral TLE due to the variability in seizure onset over time. Interictal spikes may also lateralize TLE. When spikes in TLE are mostly unilateral, the side of the spikes usually aligns with the side of the clinician-defined epilepsy. However, interpreting cases with spikes on both sides of the brain is more challenging. One limitation of relying on spikes to lateralize TLE is that spikes are traditionally analyzed visually, which restricts the ability to estimate spike rates over prolonged recordings. New spike detection algorithms have enabled the automated identification of interictal spikes with high accuracy, often surpassing human reliability. Measuring interictal spike asymmetry from intracranial EEG quantitatively lateralizes TLE. Based on this, we hypothesize that the left-to-right asymmetry of spike rates in scalp EEG will also differ among patients with left-sided, right-sided, and bilateral TLE. Currently, there is no reliable automatic method to quantitatively count spikes on the left and right hemisphere on scalp EEG. Developing a method to estimate TLE laterality using quantitative spike detections from scalp EEG could improve surgical planning, shorten EMU stays, and reduce the need for iEEG.

Methods

We included 365 patients with clinically-diagnosed who underwent scalp EEG monitoring in the EMU at the Hospital of the University of Pennsylvania (HUP) between 2017 and 2024. For 2017–2022 (model development cohort), EEG data consisted of technologist-selected clips stored to reduce clinical storage size. For 2023–2024 (testing cohort), full recordings were available; we excluded the 30 minutes before and after clinically annotated seizures. Spike detector inputs were notch-filtered at 60Hz, band-pass filtered at 0.5-70 Hz, and resampled to 128 Hz. Spike rates were determined using SpikeNet2, a convolutional neural network based on ResNet. To assess laterality of spikes passing the detection threshold of 0.43 (taken from the original publication), we developed a method called Leave One Hemisphere Out (LOHO). This method works by setting all values in one hemisphere to zero and re-running the spike detector to derive a hemisphere-specific probability (Fig. 1A). The side with the highest probability was defined as the dominant side. Finally, the spike asymmetry index (SAI) was calculated using the formula (Fig. 1B): SAI=(SpikesLeft- SpikesRight)/(SpikesLeft+ SpikesRight), where SpikesLeft is the spike rate in the left hemisphere, and SpikesRight is the spike rate in the right hemisphere.

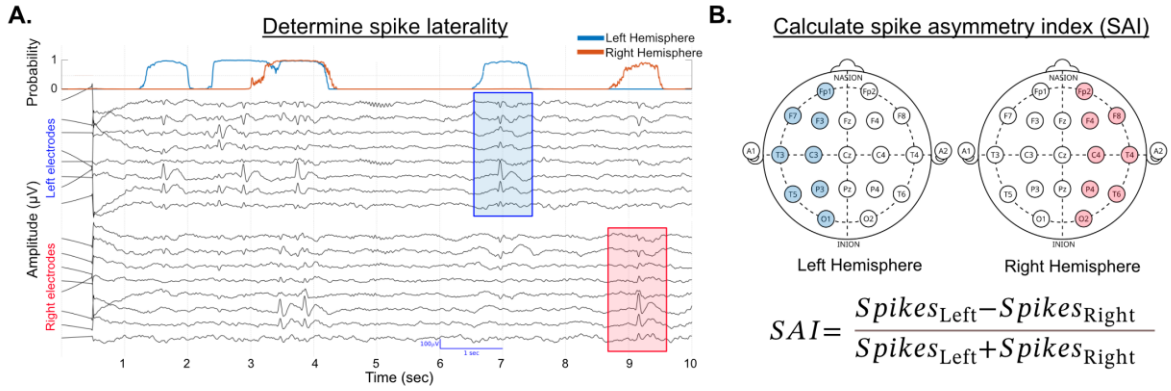
Results

We validated LOHO method to ensure the reliability of automatic spike laterality detection. For this, a random sample of 200 detected spike events was visually reviewed by a board-certified epileptologist (EC) to determine the number of true versus false detections. The accuracy of LOHO was 89%. We compared SAI between patients in the model development cohort with left-sided, right-sided, and bilateral seizure onset zones (SOZs). We found a significant difference in the SAI across groups, with left TLE patients having more left-sided spikes, and right TLE patients having more right-sided spikes (Kruskal–Wallis: $\chi^2(2) = 106.34$, $p < 0.0001$). Post-hoc pairwise Mann-Whitney U tests revealed significant differences between all group pairs (left vs. right: $U = 8646.50$, Bonferroni-corrected $p < 0.0001$, left vs. bilateral: $U = 5067.50$, Bonferroni-corrected $p < 0.0001$, and right vs. bilateral: $U = 870.00$, Bonferroni-corrected $p < 0.0001$). We also found a significant difference in the spike rates across groups (Kruskal–Wallis: $\chi^2(2) = 6.29$, $p = 0.04$). Post-hoc pairwise Mann-Whitney U tests revealed higher spike rates in bilateral TLE than left TLE ($U = 2704.00$, Bonferroni-corrected $p = 0.03$). In addition, we evaluated whether SAI could predict SOZ laterality in unseen patients. On the independent test set (n = 104), left vs. right/bilateral TLE model accuracy was 0.70 (AUC = 0.78, 95% CI = 0.69–0.87; AUPRC = 0.80, 95% CI = 0.69–0.91); right vs. left/bilateral TLE model accuracy was 0.75 (AUC = 0.80, 95% CI = 0.70–0.88; AUPRC = 0.71, 95% CI = 0.56–0.85) (Fig. 2A-B). Model performance improved with increasing recording durations, plateauing near 20 hours (Fig. 2C). Comparable accuracy within the first 20 hours of recording suggests further utility for tailoring EMU length of stay. Patients with higher prior probability of bitemporal epilepsy require more unilateral seizures to rule out bilateral disease. Real-time SAI estimates could help calibrate that requirement, offering a real-time, patient-specific guide to the optimal hospital length-of-stay, which typically averages 4-5 days. For example, using the Bayesian framework of Struck et al., in a patient with semiology suggestive of left TLE who has an SAI-modeled left TLE probability of 80% after 20 hours of recording, four left-sided seizures would be needed to achieve >90% post-test probability of unifocal disease. Conversely, if the model predicts only 20% likelihood of left TLE, then nine left-sided seizures would be needed, at which point it might be preferable to pursue an implanted device for more definitive lateralization. Furthermore, to facilitate external validation, we evaluated the performance on spikes detected using Persyst, commercial spike detection, using our models trained on LOHO method. Testing the model on spikes measured using Persyst demonstrated similar performance as with spikes detected using SpikeNet2 (left vs. right/bilateral AUC 0.83, 95% CI = 0.73–0.91; right vs. left/bilateral TLE AUC 0.85, 95% CI = 0.76–0.92).

Conclusions

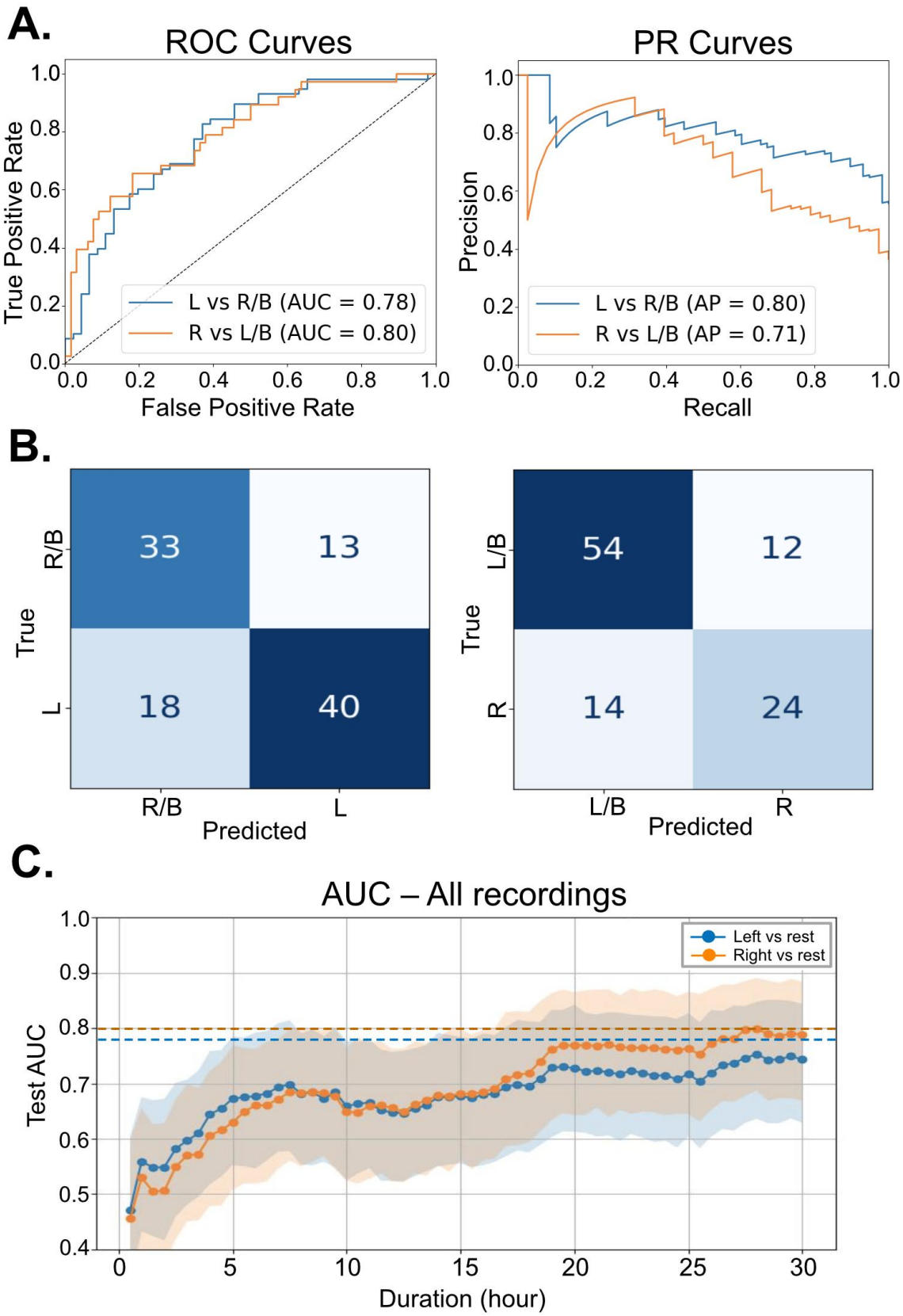
Quantitative estimates of spike rate asymmetry accurately predicted TLE laterality in unseen patients. This study supports greater use of automated spike detection from scalp EEG as a non-invasive tool for surgical planning in TLE. Using this method can decrease both the reduce the morbidity and cost of unnecessary invasive diagnostic and therapeutic interventions.

Spike laterality determination and calculation



https://secure.key4events.com/onsite/api.aspx?e=1960&img=Fig1.png&ai=74037&op=getabstractimg&dirN=0

SAI predicts TLE laterality in unseen patients



https://secure.key4events.com/onsite/api.aspx?e=1960&img=Fig2.png&ai=74037&op=getabstractimg&dirN=1

P07

A deep neural network approach to detect sharp wave ripples in human hippocampus

M. Szuromi[1], C. Chu[2, 3], M. Kramer[1, 2]

[1]Boston University - Boston (United States), [2]Johns Hopkins University - Baltimore (United States), [3]Kennedy Krieger Institute - Baltimore (United States)

Introduction

Sharp wave ripples (SWRs) are an electrophysiological biomarker of human memory. These electrographic events occur in the hippocampus and can be measured in human intracranial local field potential (LFP) recordings from patients with suspected mesial temporal lobe epilepsy (MTLE). Despite the promise of hippocampal SWRs as a biomarker of human memory, accurate detection of SWRs in noisy brain voltage recordings remains a challenge. Existing detection methods utilize time-consuming visual classification schemes or apply automated methods prone to false positive detections. Deep neural networks provide new opportunities for accurate, rapid classification of human SWRs without explicit feature selection. However, to train and validate these deep neural networks requires a large database of verified human SWRs. To address these challenges, we trained and validated a deep neural network using a large dataset of verified human hippocampal SWRs.

Methods

Our dataset consists of hippocampal recordings from 62 subjects (sampling rates ≥ 500 Hz, 960 total bipolar recordings, mean 16 and standard deviation 9 bipolar recordings per subject). We created a preliminary training set by manually identifying SWRs from human hippocampal recordings in 5 subjects. For each subject, we first applied a well-established SWR detector to identify candidate SWR detections during all N2/N3 sleep captured in a 24 hr period. We then visually inspected each candidate SWR detection (64044 total, minimum 2657, maximum 30556 per subject) and manually identified only those detections that satisfied established criteria for visual identification of human SWRs; at least 4 cycles at 80 Hz or above, not due to artifact. Doing so resulted in 1579 manually validated SWR detections (mean 320 detections per subject). For initial training, we also selected 1579 candidate SWRs not manually validated and 1579 background events (not containing a candidate SWR detection) from the 5 subjects, matching the number of manually validated SWRs in each subject. We used these data to train a long short-term memory (LSTM) deep neural network to operate directly on the sequences of real and imaginary parts of the Fourier transformed data (1-200 Hz, avoiding frequencies with common line noise artifact). The network consisted of five layers: (i) a sequence input layer, (ii) a bidirectional LSTM layer, (iii) a fully connected layer with two outputs, (iv) a softmax layer, and (v) a classification layer that computes the cross-entropy loss for classification with two mutually exclusive classes (“SWR” or “not-SWR”). We trained and tested the initial LSTM using leave-one-out cross-validation and then applied the initially trained detector (N=5 subjects) to continue training and validation following a human-in-the-loop protocol. This initial LSTM will then be used to identify candidate SWRs for validation in the full dataset (n=62 subjects) resulting in an anticipated > 20000 validated SWRs to train the final detector, which will be validated in a dataset with behavioral outcomes.

Results

Our initial LSTM detector performed well: at the maximum F1-score of 0.73, we found sensitivity of 0.82, positive predictive value (PPV) of 0.66, and accuracy of 0.80. Repeating this analysis using an alternative detector that only utilizes power at each frequency degraded performance; at the maximum F1-score of 0.49, we found sensitivity of 0.74, PPV of 0.46, and accuracy of 0.62.

Conclusions

These preliminary results support our approach to automate detection of high-quality, human hippocampal SWRs using a LSTM network with Fourier-based features. Our continuing work focuses on detector training using a human-in-the-loop procedure and validation using a naive dataset with offline memory assays.

P08
Detection of Infantile Spasms in Clinical EEG Using the PySeizure framework

B. Chybowski[1], I. Peh[2], B. Jordan[3], K. Fisher[3], A. Gonzalez-Sulser[2], J. Escudero[2], J. Shetty[2, 3]

[1]University Of Edinburgh - Edinburgh (United Kingdom) - Edinburgh (United Kingdom), [2]University Of Edinburgh - Edinburgh (United Kingdom), [3]Nhs Lothian - Edinburgh (United Kingdom)

Introduction

Early and accurate seizure detection is essential for diagnosing epilepsy, especially in infancy, when delays can have irreversible consequences. Infantile Spasms (IS) is a severe epileptic encephalopathy in which delayed diagnosis is associated with developmental regression, long-term disability, and poorer life outcomes. Despite its urgency, the diagnostic process is frequently slowed by limited access to paediatric neurophysiology expertise and the time-intensive nature of manual EEG interpretation.

Artificial intelligence (AI) has the potential to support clinicians in interpreting EEG, but many published approaches depend on dataset-specific optimisation or research-only conditions that limit their clinical use. To address this, we previously developed PySeizure, an open-source machine-learning framework designed for reproducibility, transparency, and general applicability (Chybowski *et al*., 2025, PySeizure: A single machine learning classifier framework to detect seizures in diverse datasets. arXiv:2508.07253). PySeizure was originally developed and evaluated using two established open datasets: the Children's Hospital Boston-Massachusetts Institute of Technology Scalp EEG Database (CHB-MIT) and the Temple University Hospital EEG Seizure Corpus (TUSZ). The strong results across both datasets demonstrated that generalisable seizure detection is possible without relying on dataset-specific tuning.

In this study, we extend our work to evaluate whether PySeizure can translate effectively to clinical EEG recordings from infants with suspected Infantile Spasms. Our objective was to assess whether an automated, dataset-agnostic framework could support early IS identification from routine EEG and potentially reduce reliance on specialist-only interpretation, especially in settings where access to expertise is limited.

Methods

Framework and previous evaluation.

PySeizure provides an automated pipeline for EEG analysis that includes artefact reduction, standardising electrode channels, segmentation, optional data augmentation, and both feature-based and deep-learning modelling. In earlier work, PySeizure achieved strong within-dataset performance on CHB-MIT (ROC AUC 0.904±0.059) and TUSZ (ROC AUC 0.864±0.060). Cross-dataset testing, where models trained on one dataset were directly evaluated on the other, showed robust generalisability (ROC AUC 0.615±0.039 and 0.762±0.175). Light post-processing with majority voting improved accuracy and sensitivity while maintaining stability. These findings established a technical foundation for translation into clinical settings.

Clinical dataset.

We evaluated the framework using EEG recordings from 98 infants who were assessed for suspected IS at the Royal Hospital for Children and Young People, Edinburgh. The median age was 7 months, and the average recording length was 30 minutes. EEGs were processed using the same automated PySeizure pipeline that had been validated on CHB-MIT and TUSZ. A structured augmentation module quadrupled the dataset, enhancing its variability to better capture differences in infant EEG. We trained and evaluated seven models within PySeizure, including classical machine-learning approaches based on temporal, spectral, correlation-based, and graph-theoretic features, as well as deep-learning models trained directly on raw EEG. A voting ensemble model and light post-processing were used to stabilise predictions across segments.

Patient and Public Involvement.

To ensure that our work addressed real diagnostic challenges, we engaged with families through the UK Infantile Spasms Trust (UKIST). We received feedback from more than 100 respondents, whose experiences helped shape the study and ensured that our priorities remained aligned with family-centred clinical needs related to early diagnosis.

Results

Open dataset benchmarks.

As previously reported, PySeizure performed strongly across both CHB-MIT and TUSZ and maintained stability when transferred between datasets with different montages and patient characteristics. This confirmed that the framework does not depend on dataset-specific properties and can be applied reliably across different EEG sources.

Clinical performance.

When the infant EEGs were processed using the same automated PySeizure pipeline and the models were trained directly on this clinical IS dataset, the individual models achieved robust performance, with an average ROC AUC of 0.77 ± 0.078. Combining predictions across models in a voting ensemble, however, substantially improved accuracy, achieving a ROC AUC of 0.93, with a sensitivity of 0.64 and a specificity of 0.81. This highlights the added strength of integrating multiple models to stabilise predictions and capture complementary information from the EEG features.

These results are clinically meaningful given the high variability of infant EEG, the short duration of routine recordings in the context of seizures, with a mean duration of approximately two seconds, and the well-documented inter-rater differences in diagnosing IS. The findings demonstrate that models trained on real-world infant recordings, within a fully automated and reproducible pipeline, can detect IS-related EEG abnormalities with high discrimination and practical clinical utility.

Conclusions

This study shows that a reproducible and generalisable AI framework, originally validated on open seizure datasets, can be effectively applied to routine infant EEG for early identification of Infantile Spasms. PySeizure achieved strong discriminatory performance without IS-specific optimisation, supporting its potential role in time-critical diagnostic pathways and in reducing dependence on specialist-only EEG interpretation.

This work forms a foundation for the broader goals of the EPIC project (Enabling early and equitable diagnosis of epilepsy in infants in the community). EPIC aims to develop a simple and

clinically usable AI-assisted brain monitoring approach for seizure detection in early childhood, including severe syndromes such as IS. While the project covers a broader set of clinical, technical, and translational goals, these findings demonstrate that reliable automated detection from routine clinical EEG is achievable and provides a foundation for the project.

The project continues to be guided by strong patient and public involvement, including ongoing contributions from the Epilepsy Research Institute's Shape Network and extensive engagement with families affected by IS through the UK Infantile Spasms Trust. Their lived experience underscores the urgent need for earlier, fairer, and more personalised diagnostic support in paediatric epilepsy.

By combining clinically grounded AI methodology with direct community involvement, EPIC aims to co-develop a digital tool that can be used with families and clinicians, for families and clinicians. This study represents a significant step toward that goal and toward improving access to early diagnosis in paediatric epilepsy care.

This work was supported by a seed funding grant from the Bayes Centre and by a Harmonised IAA grant at the University of Edinburgh. We acknowledge collaboration with UKIST and BrainsView.

P09
Multimodal Integration of SEEG Seizure Dynamics and Cortical Geometric Eigenmodes

A. Romeo[1], P. Ajay[1], D. Oliver[1], O. Juan[1], H. Bolus[1]

[1]University Of South Alabama - Mobile, Al (United States)

Introduction

The relationship between intrinsic cortical geometric eigenmodes and the propagation of epileptic seizures remains poorly characterized. Spatial eigenmodes of the cortex—derived from the Laplace–Beltrami operator on subject-specific surface reconstructions—represent fundamental spatial patterns of cortical activity and may provide a mechanistic framework for understanding seizure networks. In this study, we investigated how stereoelectroencephalography (SEEG)–defined seizure propagation patterns align with cortical eigenmodes and whether these modes can inform or optimize intracranial electrode placement.

Methods

We retrospectively analyzed SEEG recordings and clinically annotated seizure networks from 10 patients undergoing invasive evaluation for drug-resistant epilepsy. For each patient, cortical surface reconstructions were generated using the FreeSurfer pipeline, and geometric eigenmodes were computed using the Laplace–Beltrami operator. SEEG contact locations were co-registered to the cortical surface, and seizure network dynamics were compared to spatial eigenmodes across multiple time points, including seizure onset, early propagation, and peak ictal spread. Qualitative visual overlays were used to assess spatial correspondence.

Results

Across patients, distinct eigenmodes exhibited variable degrees of spatial overlap with seizure onset zones and propagation pathways. Visualization of eigenmodes alongside SEEG heat maps revealed time-dependent relationships between geometric modes and ictal evolution.

Conclusions

Our preliminary findings suggest that cortical eigenmodes may provide a useful structural prior for interpreting seizure networks and could help guide more optimal SEEG electrode placement. Incorporating eigenmode analysis into pre-implant planning may enhance coverage of ictogenic networks, improve diagnostic yield, and ultimately inform patient-specific strategies for surgical intervention.

P10
Automatic detection of interictal epileptiform discharges in In-ear EEG

E. Dessevres[1], J. Pyrzowski[2, 3], J.N. Reyes[4], G. Huberfeld[2, 3], M. Kopytek-Beuzen[5], J. Moreno[6], M. Pielas[7], A.M. Uscátegui[6], J. Pazuelo[8], G. Arkoun[8], F. Segura-Quijano[4], M.A. Venegas[6], P. Kunert[5], M. Valderrama[9], M. Le Van Quyen[1]

[1]Laboratoire D'imagerie Biomédicale (lib), Inserm U1146, Sorbonne Université, Umr7371 Cnrs - Paris (France), [2]Université Paris Cité, Institute Of Psychiatry And Neuroscience Of Paris (ipnp), Inserm U1266, Neuronal Signaling In Epilepsy And Glioma - Paris (France), [3]Hôpital Fondation Adolphe De Rothschild, Service De Neurologie - Paris (France), [4]Department Of Electrical And Electronic Engineering, Universidad De Los Andes - Bogotá (Colombia), [5]Department Of Neurosurgery, Csk Uck Wum - Warsaw (Poland), [6]Uniepilepsias Unidad De Estudio Y Tratamiento Integral De Las Epilepsias - Bogotá (Colombia), [7]Department Of Neurology And Epileptology, Children's Memorial Health Institute - Warsaw (Poland), [8]Naox Technologies - Paris (France), [9]Department Of Biomedical Engineering, Universidad De Los Andes - Bogotá (Colombia)

Introduction

In routine clinical practice, scalp electroencephalography (EEG) typically provides only short snapshots of brain activity, while current ambulatory systems remain cumbersome and can often be worn for only a few hours. In contrast, in-ear EEG offers a discrete, non-invasive alternative suitable for long-term ambulatory monitoring. Previous work has shown that interictal epileptiform discharges (IEDs) visible on scalp EEG also appear in in-ear recordings and can be visually identified in most patients due to the strong correlation between in-ear signals and the T7–T8 scalp electrodes [1]. However, clinicians are not accustomed to annotating IEDs directly from single-channel in-ear signals without the spatial context provided by multichannel scalp EEG. Moreover, because in-ear devices enable continuous monitoring over several days, the visual review of IEDs becomes time-consuming and challenging. This study evaluated the feasibility of automatically detecting IEDs in in-ear EEG signals using deep learning.

Methods

We investigated two independent cohorts of patients with drug-resistant focal epilepsies undergoing video-EEG monitoring at the Medical University of Warsaw, Poland (N=8) and Bogotá, Colombia (N=9). Co-acquisition of scalp and in-ear EEG was performed as in our previous work [2]. Four experienced, board-certified neurophysiologists (GH, MK, JM, JP) independently reviewed the scalp recordings for IEDs following the International Federation of Clinical Neurophysiology (IFCN) criteria. Additionally, whole-night sleep dual-device recordings were obtained in healthy participants at ONIROS SAS, Bogotá, Colombia – Comprehensive Sleep Care Center (N=13). Signals were bandpass filtered (0.3–35 Hz) before analysis. We extracted 1-s in-ear windows and represented each as a 256×256 px image combining the raw waveform (top) and a time–frequency representation (bottom). Windows containing IEDs from epileptic patients formed the positive class; background windows from healthy subjects (the sleep cohort) formed the negative class, ensuring the absence of unannotated IEDs. We discarded windows with excessive or insufficient amplitude (>300 µV or <15 µV), spiky artifacts (prominence >100 µV, width <30 ms), zero-crossing rate <2, entropy <0.1, and Hjorth mobility <0.05—criteria designed to

exclude windows that could not represent IEDs and to allow the model to train on more ambiguous cases. Given the natural class imbalance (>100 background windows per IED), we augmented IED windows by shifting their temporal position by 0.25 s, 0.5 s, and 0.75 s. A random sampler ensured equal proportions of IED and background windows in each batch. To avoid confounding effects related to sleep, the sampler also balanced sleep stages across batches. Images were fed into a Swin-S v2 transformer backbone pretrained on ImageNet-1k, followed by an argmax output for binary classification.

Results

Across n=2,857 annotated IED windows from 17 epileptic patients and n=279,421 background windows from 13 healthy subjects, we evaluated the model using 5-fold cross-validation with patient-independent splits. The model achieved a sensitivity of ~50%, a specificity >95%, and an area under the ROC curve of ~0.90. Most temporal and frontal spikes annotated on the scalp were correctly identified in the in-ear channel—consistent with the strong correspondence between in-ear EEG and T7–T8 scalp electrodes—, whereas centroparietal spikes were largely missed.

Conclusions

Our results demonstrate that automatic detection of IEDs from a single in-ear EEG channel is feasible. Acceptable performance was achieved for temporal and frontal epilepsies, whereas IEDs in centrotemporal epilepsies were detected less reliably. Overall, these findings highlight the potential utility of machine learning–based approaches for long-term in-ear EEG monitoring, including the characterization of ultralong epileptiform activity patterns—such as multi-day IED modulation—and their relationship to seizure risk and treatment response.

References:

[1] Moumane, H., Pazuelo, J., Nassar, M., Juez, J. Y., Valderrama, M., & Le Van Quyen, M. (2024). Signal quality evaluation of an in-ear EEG device in comparison to a conventional cap system. Frontiers in Neuroscience, 18, 1441897.

[2] Pyrzowski J., et al. In-ear EEG detection of individual focal interictal epileptiform discharges. 5th International Congress on Mobile Health and Digital Technology in Epilepsy, Copenhagen Denmark, September 4-6 2025, na.eventscloud.com/website/74027/home-558/.

P11

Pose-Based Deep Learning Model Distinguishes Infantile Spasms from Mimics in Home Videos

G. Rivera[1], L. Rhode[2], M. Cronk[1], E. Kossoff[1], R. Chellappa[3]

[1]Department Of Neurology, Johns Hopkins University School Of Medicine - Baltimore (United States), [2]Department Of Electrical And Computer Engineering, Johns Hopkins University - Baltimore (United States), [3]Department Of Electrical And Computer Engineering, Department Of Biomedical Engineering, Johns Hopkins University - Baltimore (United States)

Introduction

Infantile Epileptic Spasms Syndrome (IESS) is a severe epilepsy characterized by epileptic spasms. Some patients with IESS have developmental plateau or regression and a disorganized EEG pattern known as hypsarrhythmia. Delayed or missed diagnosis is common and is associated with worse neurodevelopmental outcomes (Hussain et al., 2017). Although caregivers increasingly present with smartphone videos of suspected episodes—which has been shown to reduce time to diagnosis (Rao et al., 2023)—accurate interpretation requires specialized expertise. Clinical presentations of epileptic spasms are heterogeneous, ranging from subtle to dramatic, and often resemble non-epileptic conditions such as Sandifer syndrome.

Recent computer vision approaches show promise in addressing this diagnostic challenge. A pilot study by Eguchi et al. (2023) evaluated the ability of a deep neural network to identify epileptic spasms in a single patient with IESS hospitalized for video EEG monitoring. More recently, a foundational vision model fine-tuned on social media-derived videos of 141 children with epileptic spasms and 127 healthy controls achieved an AUC of 0.96 with 82% sensitivity and 90% specificity; notably, spasm mimics were underrepresented in the healthy cohort (Miron et al., 2025).

To address the critical challenge of distinguishing epileptic spasms from spasm mimics, we are developing a dedicated vision model trained on a curated dataset enriched with confirmed spasm mimics. Our goal is to improve early identification of infantile spasms and support timely referrals for definitive diagnosis and treatment.

Methods

The study was approved by the Johns Hopkins IRB. Eligible participants were <24 months old and had home videos with clear facial visibility; most also showed two to four limbs. Videos of insufficient quality to permit reliable assessment were excluded.

We curated a dataset of epileptic spasms (256 clips from 29 children) and spasm mimics (178 clips from 32 children) with written parental consent or under IRB-approved implicit-consent procedures. Mean clip duration was 2.4 seconds for epileptic spasms and 4.6 seconds for mimics. Spasm subtypes include flexor, extensor, mixed, subtle, and focal. Spasm mimics include head bobbing, Sandifer syndrome, shuddering spells, sleep myoclonus, paroxysmal tonic upgaze, and

other stereotypies. Labels were based on clinical interpretations by each patient's primary pediatric neurologist and independently verified by study investigators.

The estimated 2D pose skeleton keypoints of the infants were extracted using the MMAction2 pose extraction pipeline. For each frame, the 17 skeleton keypoints and detection confidence score per joint from each frame were saved. We then trained a skeleton-based action recognition model using the PoseC3D architecture as implemented in MMAction2. The PoseC3d architecture (Duan et al., 2021) relies on 3D heatmap stack of the joints which represents the spatio-temporal sequence of joint location. The network receives input tensors of shape (C, T, V, M) were C=3 (x,y,confidence), T= frames, V= 17 joints, and M=number of persons. A ResNet-50 backbone and SlowOnly temporal modelling with pretraining on Kinetic 500 dataset was used as provided in MMAction2's PoseC3D configuration. The final feature representation is passed through a global spatiotemporal pooling layer and a fully connected classification head. The model relies purely on pose information and does not use raw pixels or optical flow.

Results

We performed five-fold cross-validation with splits defined at the patient-level to prevent data leakage. Clips from the same patient never appeared in both the training and test sets within a fold. For each fold, we trained the model on the training folds and reported performance on the held-out test fold. Out-of-fold predictions were pooled across all five test folds for analysis.

Because multiple clips were obtained from the same infant, we reported patient-level performance. For patient-level analysis, we aggregated clip-level predictions per patient and averaged clip-level predicted probabilities to obtain a patient-level prediction score. Performance metrics (AUC, sensitivity, specificity) and 95% confidence intervals were computed at the patient level using nonparametric bootstrap resampling of patients.

Our model discriminates epileptic spasms from spasm mimics with an AUC of 0.86 (CI 0.78-0.92, five-fold cross-validation). Using a threshold of 0.5, a sensitivity of 0.81 and a specificity of 0.73 were computed.

Conclusions

Our pilot results show that a pose-based model using only 2D skeletal keypoints can discriminate epileptic spasms from common mimics in caregiver-recorded, real-world videos. Skeleton-based action recognition is a promising approach for automated detection of infantile spasms, and mimic-rich datasets are essential for developing reliable tools to support early diagnosis.

P12

SEEG-Transformer: A Transformer-Based Architecture for Integrating Interictal Multimodal Features for Epileptogenic Zone Localization

P. Nejedly[1], M. Pail[2], V. Hrtonova[1], P. Daniel[2], B. Matouskova[1], I. Dolezalova[2], P. Jurak[1], P. Klimes[1], M. Brazdil[2]

[1]Institute Of Scientific Instruments Of The Cas - Brno (Czechia), [2]Brno Epilepsy Center, Department Of Neurology, Member Of Ern-Epicare, St. Anne's University Hospital, Faculty Of Medicine, Masaryk University - Brno (Czechia)

Introduction

Precise localization of the epileptogenic zone (EZ) is essential for planning effective surgical interventions in drug-resistant epilepsy, yet interpreting SEEG recordings remains complex and subjective. Building on our previous graph neural network (GNN) approach (Nejedly et al., 2025), we introduce a Transformer-based deep-learning model designed to integrate heterogeneous SEEG biomarkers with spatial, anatomical, and connectivity information. Unlike GNNs, which rely on local neighborhood propagation, the Transformer architecture leverages global attention, enabling every electrode to contextualize its activity relative to all others. This allows the model to capture long-range dependencies in epileptic networks that would be ignored in the GNN framework.

Methods

For each electrode, we extracted a multimodal feature set including SEEG spectral band features, interictal spike rates, and phase–amplitude coupling, along with MNI coordinates, topological spatial-density measures derived from inter-electrode geometry, and functional connectivity metrics. These features were embedded into a shared latent space and processed through self-attention layers. Positional encodings incorporating anatomical and spatial information further guided the attention mechanism to respect implantation geometry while allowing to capture global functional interactions. The model outputs an EZ probability for each electrode, providing an interpretable ranking of candidate targets.

Results

We retrospectively evaluated the proposed framework on 30-minute resting state recordings from 90 patients from St. Anne's University Hospital Brno who underwent SEEG implantation followed by resective surgery or thermocoagulation. Among these, 38 achieved Engel I outcomes and 52 had Engel II–IV outcomes. Using five-fold patient cross-validation, the SEEG-Transformer achieved an AUPRC of 0.74 and AUROC of 0.92 in Engel I patients, and AUPRC of 0.21 and AUROC of 0.79 in Engel II–IV patients. These results surpass state-of-the-art models, including our prior GNN. For reference, the spike rate biomarker yielded AUPRC/AUROC of 0.41/0.75 (Engel I) and 0.29/0.72 (Engel II–IV). Interestingly, in patients with poor outcomes, the model frequently assigned high importance to regions that were not targeted surgically. While this divergence from clinical decisions cannot be interpreted as evidence that these regions should

have been resected or ablated, it suggests that the framework may help generate alternative hypotheses for potential targets, which will require further validation.

Conclusions

In summary, this Transformer-based architecture provides a powerful framework for automated EZ prediction from SEEG data. By integrating diverse electrophysiological, anatomical, and connectivity features within a global attention-based model, it generalizes our earlier topology-aware GNN and achieves superior predictive performance. Our future work will examine the contribution of individual feature modalities. Proposed findings uncover the potential of topology aware architectures to automatically suggest targets for epilepsy surgery.

P13

Machine Learning–Based Prediction of Anti-Seizure Medication Withdrawal Outcomes in Pediatric-Onset Epilepsy

J. Cho[1], I. Yune[1], H.J. Yun[1], M.H. Kook[1], H.R. Jang[1], H. Kim[1]

[1]Seoul National University Bundang Hospital - Seungnam (Korea, Republic of)

Introduction

Withdrawal of anti-seizure medications (ASMs) is a critical decision in the management of epilepsy, as it directly impacts the long-term patient outcomes. Although several prognostic models have been proposed, these approaches have largely relied on conventional statistical methods. In this study, we aimed to develop and evaluate machine learning models to provide a more accurate and clinically useful prediction of ASM withdrawal outcomes.

Methods

We retrospectively analyzed pediatric-onset epilepsy patients who visited Seoul National University Bundang Hospital between June 2003 and August 2025. Patients who attempted ASM withdrawal were included. Successful withdrawal was defined as remaining seizure-free for at least one year after discontinuation. Clinical data were collected, and evaluated four supervised machine-learning algorithms, Support Vector Machine, Logistic Regression, Gradient Boosted Trees, and Random Forest, using a standardized training pipeline. Each model underwent hyperparameter optimization via grid search with five-fold cross-validation. After model fitting, we quantified feature importance using permutation-based methods to obtain model-agnostic estimates of variable contribution. In addition, we applied Principal Component Analysis and re-evaluated model performance on the transformed feature space to assess the stability of predictive accuracy under dimensionality reduction.

Results

We analyzed 294 patients who underwent ASM withdrawal. The median age at seizure onset was 8.5 years. At diagnosis, MRI findings were normal in 193 patients (70.6%). Developmental delay was present in 32 patients (11.8%), a family history of epilepsy in 14 (5.0%), and a history of febrile seizures in 53 (19.3%). Epilepsy type was focal in 164 patients (60.2%), generalized in 46 (16.9%), and unclassified in 62 (22.9%). A history of status epilepticus was identified in 28 patients (10.2%). Pre-withdrawal EEG revealed abnormalities in 63 patients (23.2%), whereas post-withdrawal EEG abnormalities persisted in 127 (46.6%). After at least one year of follow-up, 130 patients (47.5%) experienced recurrence. In predictive modeling, the random forest model achieved the highest AUC values of 0.85. The top three predictive features were presence of epilepsy duration before withdrawal, epilepsy type and presence of developmental delay.

Conclusions

Machine learning–based models demonstrated high predictive performance for ASM withdrawal outcomes. These findings suggest that ML approaches may provide a clinically useful tool to guide ASM withdrawal decisions.

P14

Leveraging Machine Learning to Identify Nested Oscillations as Electroencephalographic Biomarkers in a Model of Developmental Epileptic Encephalopathy

Q.Q. Sun[1], T. Cao[1], C. Zhang[1]

[1]University Of Wyoming - Laramie (United States)

Introduction

Introduction: The epileptic encephalopathies represent a group of severe neurological disorders characterized by progressive cognitive and behavioral dysfunction directly attributable to epileptic activity. Among the most challenging subtypes are conditions featuring electrical status epilepticus during sleep (ESES) and continuous spike-wave during sleep (CSWS), where abundant sleep-activated epileptiform activity leads to profound neurodevelopmental regression. A paramount concern in these and other refractory epilepsies is the risk of Sudden Unexpected Death in Epilepsy (SUDEP), the mechanisms of which remain poorly understood. Critically, the diagnostic and therapeutic management of these conditions is hampered by a reliance on subjective, time-consuming manual analysis of electroencephalography (EEG), which often fails to identify subtle but pathophysiologically critical electrophysiological biomarkers. One such class of biomarkers is nested oscillations, specifically phase-amplitude coupling (PAC), a phenomenon where the phase of a low-frequency rhythm (e.g., delta or theta) modulates the amplitude of a high-frequency oscillation (e.g., gamma, ripples, or fast ripples). PAC is increasingly recognized as a mechanism for neuronal communication and synaptic plasticity, but its pathological exaggeration is implicated in ictogenesis and seizure maintenance. This study investigates the role of pathological PAC as a biomarker for epileptic network dysfunction, ictal transitions, and SUDEP vulnerability, utilizing a genetically accurate rodent model. We focus on characterizing these coupling dynamics in detail to establish a foundational basis for future automated diagnostic technologies.

Methods

Methods: We employed the ank3-1b mutant mouse model (n=15, aged 8-12 weeks), which recapitulates key features of human epileptic encephalopathies, including spontaneous seizures, CSWS/ESES-like sleep-activated patterns, and a significant incidence of SUDEP. Wild-type littermates (n=10) served as controls. High-density intracranial EEG recordings were performed using multi-channel microelectrode arrays chronically implanted in key structures of the limbic and thalamocortical networks, including the hippocampal CA1 region, prefrontal cortex, somatosensory cortex, and thalamic nuclei. Data were acquired continuously over 24-hour periods at a sampling rate of 2.5 kHz to adequately capture high-frequency oscillations (HFOs) up to fast ripples. This setup allowed for the capture of both ictal (seizure) and interictal periods, including sleep-wake cycles. The primary analytical focus was on quantifying phase-amplitude coupling. Signals were preprocessed with artifact removal and band-pass filtering into distinct frequency bands: delta (0.5-4 Hz), theta (4-8 Hz), gamma (30-80 Hz), ripples (80-200 Hz), and fast ripples (200-600 Hz). The Hilbert transform was applied to extract the phase time-series of low-frequency bands and the amplitude envelope of high-frequency bands. Coupling was quantified using the Modulation Index (MI), with statistical validation against phase-scrambled surrogate data. Additional analyses included time-frequency decomposition and correlation of electrophysiological events with

simultaneous video-recorded behavioral and physiological parameters (e.g., respiration) to assess SUDEP risk.

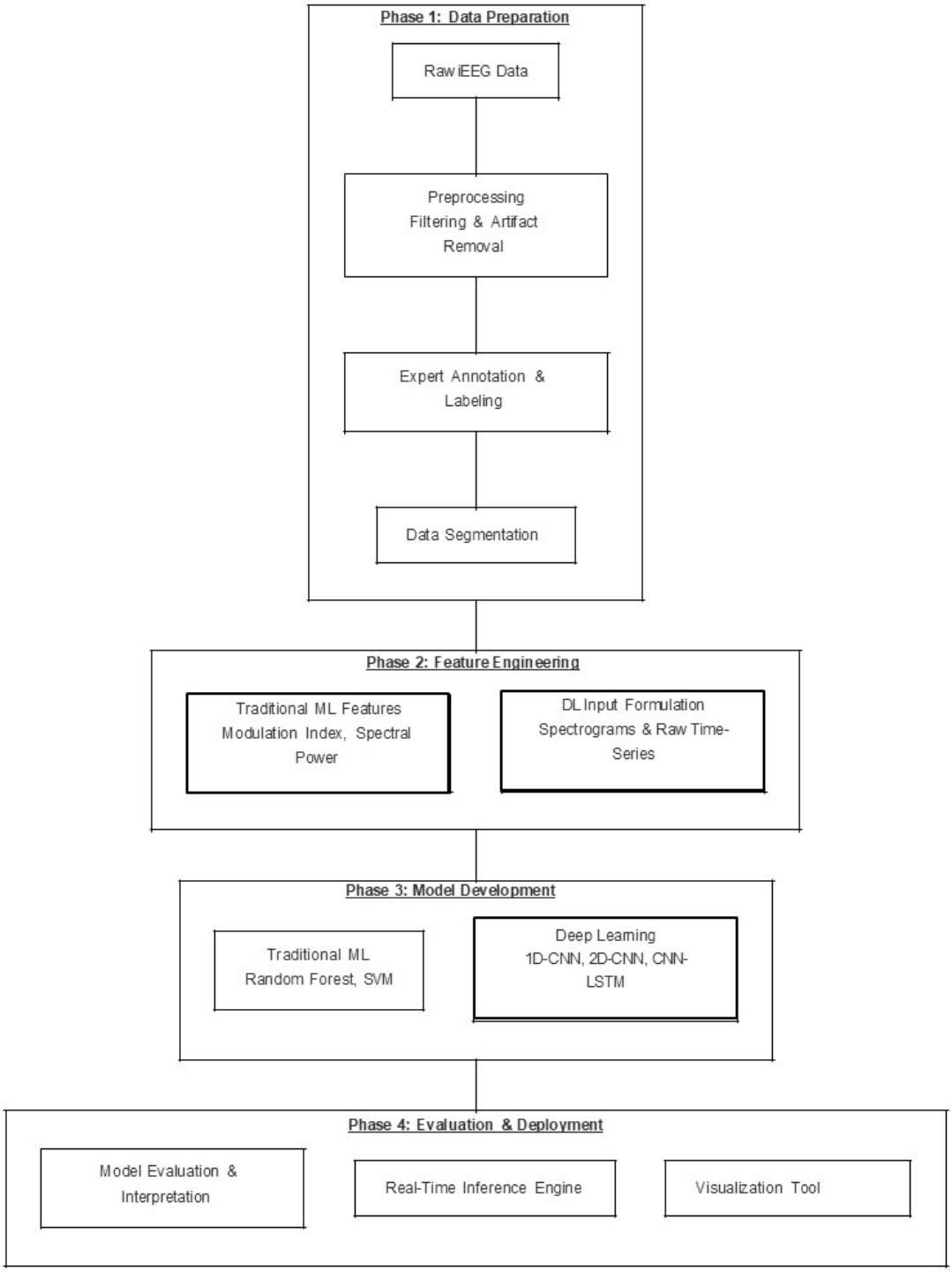
Results

Results: Our investigation revealed a profound dysregulation of cross-frequency coupling in the ank3-1b epileptic network. During interictal periods, mutants exhibited a significant global elevation in PAC compared to wild-type controls (Mean MI: 0.42 ± 0.08 vs. 0.15 ± 0.04, $p<0.001$). This enhanced coupling was not random but was intricately linked to discrete epileptiform events. Specifically, we found that theta-phase modulation of gamma and ripple amplitudes was robustly associated with interictal spikes, with the strongest coupling occurring in hippocampal channels immediately following the spike peak, suggesting a role in post-spike network synchronization and potential memory circuit disruption. A major novel finding was the identification of a distinct, pro-ictal coupling signature preceding seizure onset. We discovered that early delta-phase modulation of gamma-band activity served as a primary driver for the initiation of spontaneous spike-and-wave discharges (SWDs). This delta-gamma nesting emerged 5-8 seconds before the electrographic onset of SWDs and was spatially widespread across neocortical regions, indicating a mechanism of slow, cortical synchronization that recruits local microcircuits through gamma activity to pave the way for the ictal rhythm. During fully established ictal events (n=45 across mutants), the coupling dynamics intensified, with the nesting of ripples within theta phases increasing threefold. These high-ripple events were a hallmark of CSWS/ESES-like episodes and were strongly correlated with physiological biomarkers of SUDEP risk; in several recorded SUDEP cases, terminal seizures were characterized by an extreme, sustained surge in hippocampal theta-ripple PAC coinciding with abrupt respiratory suppression. The spatial and temporal patterns of these different PAC signatures (interictal spike-related, pre-ictal, and ictal) were consistently and reliably characterized through our detailed analytical pipeline, providing a rich and quantitative dataset.

Conclusions

Conclusions: This study provides a comprehensive characterization of pathological phase-amplitude coupling as a critical electrophysiological biomarker in a model of epileptic encephalopathy. We demonstrate that distinct PAC profiles are associated with specific network states: theta-gamma/ripple coupling with interictal spikes, and initiating delta-gamma coupling with SWD onset. These findings position nested oscillations not merely as epiphenomena but as active mechanisms in network hyperexcitability, ictogenesis, and the propagation of seizure activity. The strong correlation between specific PAC patterns, particularly high-amplitude theta-ripple coupling, and SUDEP risk indicators offers a potential electrophysiological predictor for this devastating outcome. By meticulously defining and quantifying these relationships, our work establishes a crucial foundational dataset. The logical and necessary progression of this research is the development of advanced computational tools to translate these findings into clinical practice. Therefore, future work will be dedicated to building and validating robust machine learning and deep learning algorithms for the automated, real-time detection of these PAC biomarkers from prolonged EEG recordings. The ultimate translational goal is to integrate such tools into diagnostic frameworks to improve early identification of epileptic encephalopathies, refine seizure forecasting, and guide targeted interventions to mitigate disease progression and SUDEP risk.

A Machine & Deep Learning Workflow



https://secure.key4events.com/onsite/api.aspx?e=1960&img=FLOW-CHART.jpg&ai=74083&op=getabstractimg&dirN=0

Interictal spikes (IS) in Ank3-1b mutant mice.

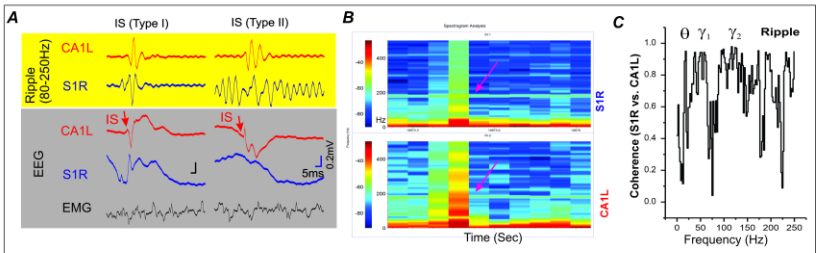


Fig. 2. Interictal spikes (IS) in Ank3-1b mutant mice. A) Two types of IS spikes were detected in CA1; all CA1 IS exhibited single spikes in EEG and divided into multiple spikes in the ripple band (80-250Hz). Type I IS had an initial negative voltage followed by slow depolarization. Type II IS had fast depolarization followed by rapid hyperpolarizations. **B)** Ca1 and S1 IS shows an increase in power in broad frequency ranges (delta-500Hz). **Note that nested-oscillations (**γ-ripple-fast-ripple**) underlies each IS. C)** Coherence between S1 and contralateral CA1 is high across many frequency bands.

https://secure.key4events.com/onsite/api.aspx?e=1960&img=fig-2.jpg&ai=74083&op=getabstractimg&dirN=1

P15

Optimized Unsupervised Clustering of Neuronal Activity in Human Epileptic Hippocampal Slices

M. Kolajova[1, 2], P. Nejedly[1, 2], J. Cimbalnik[3], P. Jurak[1], J. Halamek[1], O. Svecova[4], M. Bebarova[4], E. Brichtova[5], M. Hendrych[6], H. Hribkova[7], E. Zatloukalova[8], M. Brazdil[8, 9], P. Klimes[1, 3]

[1]Institute Of Scientific Instruments, The Czech Academy Of Sciences - Brno (Czechia), [2]First Department Of Neurology, Faculty Of Medicine, Masaryk University - Brno (Czechia), [3]International Clinical Research Center, St. Anne's University Hospital - Brno (Czechia), [4]Department Of Physiology, Faculty Of Medicine, Masaryk University - Brno (Czechia), [5]Department Of Neurosurgery, St. Anne's University Hospital And Faculty Of Medicine, Masaryk University - Brno (Czechia), [6]First Department Of Pathology, St. Anne's University Hospital And Faculty Of Medicine, Masaryk University - Brno (Czechia), [7]Department Of Histology And Embryology, Faculty Of Medicine, Masaryk University - Brno (Czechia), [8]Brno Epilepsy Center, Department Of Neurology, St. Anne's University Hospital, Faculty Of Medicine, Masaryk University - Brno (Czechia), [9]Behavioral And Social Neuroscience Research Group, Ceitec Central European Institute Of Technology, Masaryk University - Brno (Czechia)

Introduction

Understanding how neuronal populations interact within epileptic tissue requires methods capable of capturing both spatial and temporal structure in extracellular activity. Multielectrode arrays (MEAs) enable high-resolution recordings across neuronal networks, making them a powerful platform for identification of functional neuronal subgroups. The objective of this study was to apply optimized unsupervised spatiotemporal clustering of extracellular action potentials (EAPs) to characterize distinct neuronal groups in human epileptic hippocampal slices.

Methods

We analyzed 60-minute, 60-channel (59 + 1 reference) MEA recordings (25 kHz sampling) collected at the Department of Physiology, Faculty of Medicine, Masaryk University, in collaboration with the Brno Epilepsy Center (St. Anne's University Hospital). Hippocampal tissue was obtained from four patients undergoing resection for drug-resistant mesial temporal lobe epilepsy with hippocampal sclerosis (MTLE/HS). EAPs were detected using an automated spike detector, and for each event, signal segments from all channels were extracted within a predefined time window to construct a structured spatiotemporal dataset (Fig. 1A). Uniform Manifold Approximation and Projection (UMAP) was used for dimensionality reduction into a 2D embedding space, followed by clustering with hierarchical density-based spatial clustering of applications with noise (HDBSCAN).

To achieve optimal performance, Ray Tune, a hyperparameter-tuning library, was employed. A defined tuning pipeline evaluated 150 hyperparameter configurations within a specified search space, including UMAP parameters such as n_neighbors and min_dist, with OptunaSearch exploring candidate settings while maximizing the trustworthiness metric. For HDBSCAN, a grid-

search strategy varied parameters such as min_samples, min_cluster_size, metric, and cluster_selection_method, with the DBCV metric used to identify the most stable clustering configuration. Following optimization, clustered spikes were grouped by label, centered around their peaks, detrended, and averaged across channels to obtain representative spatiotemporal waveforms and amplitude–delay patterns (Fig. 1B).

Results

The optimized UMAP–HDBSCAN pipeline produced stable clustering outcomes, achieving a mean DBCV score of 0.79 and identifying an average of 22 clusters per dataset, with 96.88% data retention. Distinct spatial amplitudes and timing delays in each cluster indicated the presence of separable neuronal subgroups (Fig.2).

Conclusions

Unsupervised spatiotemporal clustering of MEA-recorded EAPs revealed distinct neuronal subgroups with unique spatial and temporal signatures. This high-resolution approach uncovered activity patterns in human hippocampal slices that cannot be detected with macroelectrodes and offers a useful tool for studying epileptogenesis in ex vivo epileptic tissue.

Dataset and example clustered waveforms.

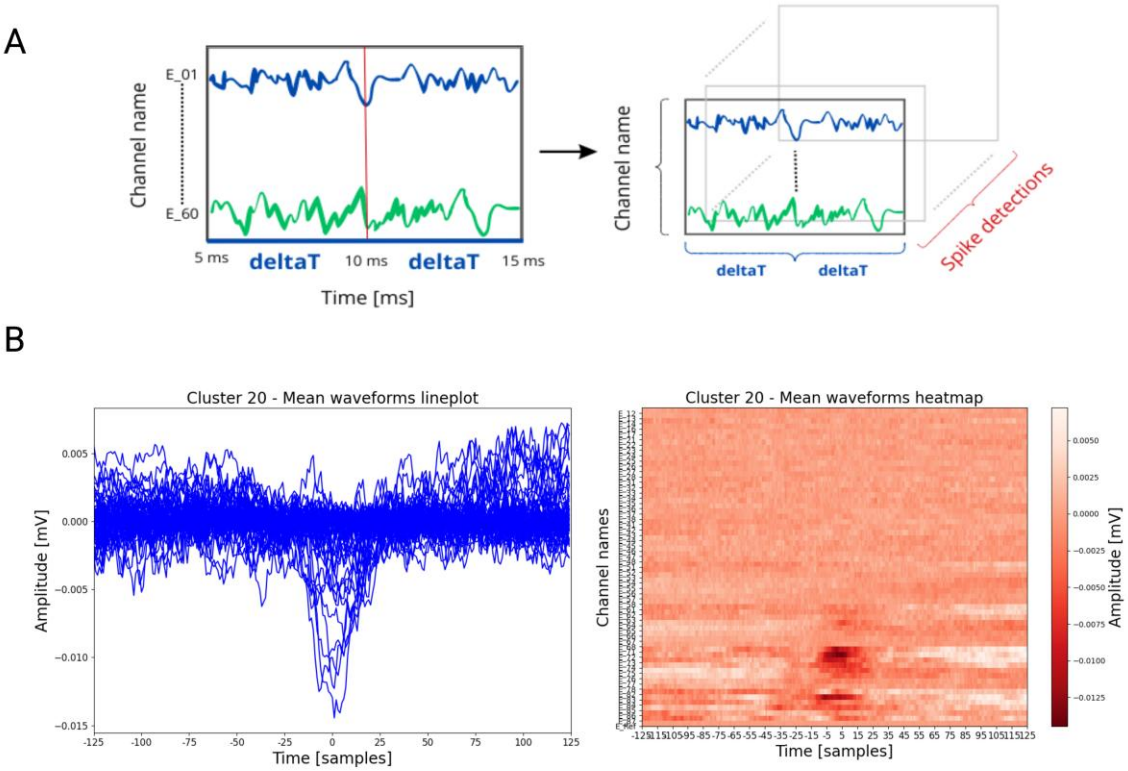


Figure 1: (A) Construction of the spatiotemporal dataset by extracting multichannel signal segments around each detected spike. (B) Mean waveforms from a selected cluster in Patient 16 shown as line plots and heatmaps, illustrating spatial amplitude differences and inter-channel timing delays.

https://secure.key4events.com/onsite/api.aspx?e=1960&img=AI_in_E_1.png&ai=74084&op=getabstractimg&dirN=0
Identification of distinct neuronal subgroups.

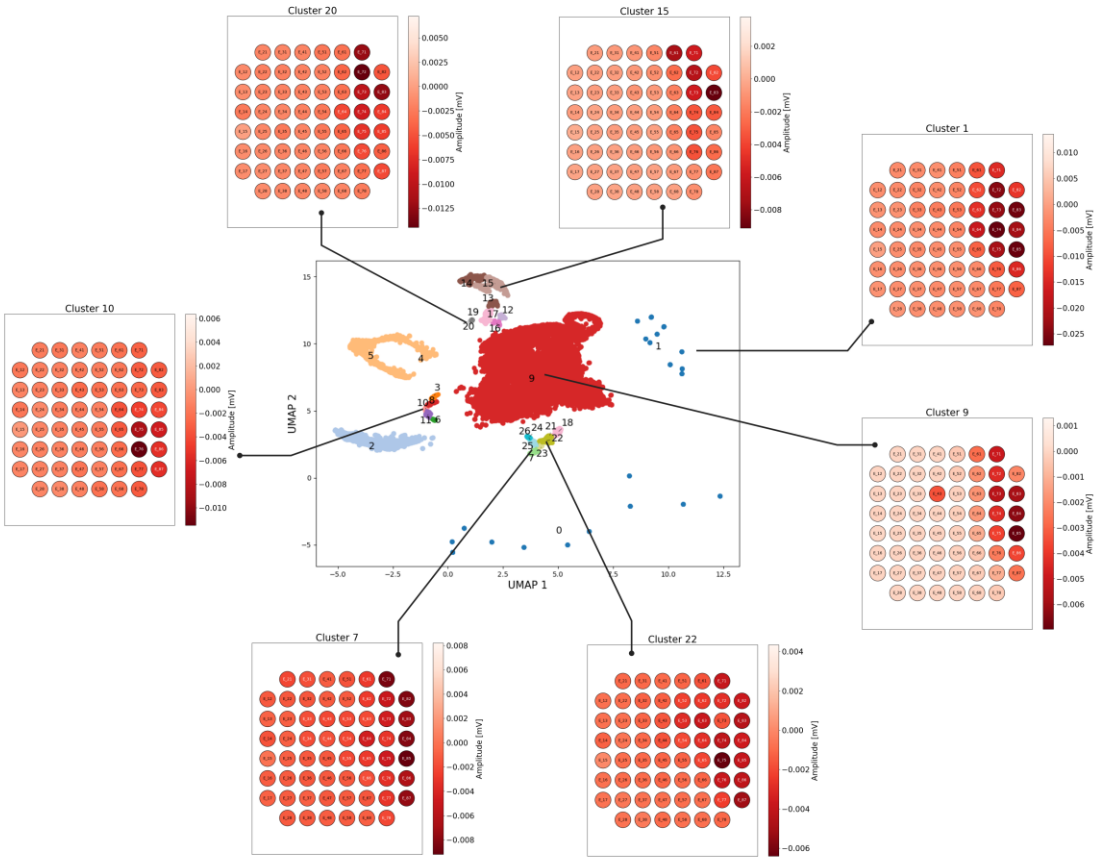


Figure 2: Patient 16, visualization of neural activity across selected clusters on a microelectrode array (MEA) layout, with each cluster displayed around a central UMAP plot. Each cluster is represented by a unique color, linked to specific MEA layout that illustrate signal amplitudes at individual electrodes at the time of the event. On the MEA grid, clusters - particularly Clusters 10 and 20, are located in different positions.

https://secure.key4events.com/onsite/api.aspx?e=1960&img=AI_in_ep_2.png&ai=74084&op=getabstractimg&dirN=1

P16
Cycle-based Control of Circadian Brain Network Dynamics for Seizure Reduction in Epilepsy

K. Saboo[1], J. Fan[1], E. Tadayon[1], E. Chang[1], V. Rao[1], A. Khambhati[1]

[1]University Of California San Francisco - San Francisco (United States)

Introduction

State-dependent effects of responsive neurostimulation (RNS) on brain network dynamics in epilepsy are not well understood. As a result, there is a lack of models that predict the effect of stimulation on treatment efficacy in RNS. We aim to discern the effect of stimulation on brain network dynamics and develop a computational model that captures this relationship. Such a model can be clinically useful in optimizing stimulation parameters.

There are three main challenges. (i) *A biomarker of ongoing seizure reduction is unavailable*. Previous studies discovered early biomarkers that predict long-term seizure reduction from RNS, such as circadian change in the aperiodic component of ECoG [1] and ECoG functional connectivity [2]. However, seizure reduction changes non-monotonically over time and heterogeneously across patients. ECoG biomarkers that reflect ongoing seizure reduction are needed to discern the effect of stimulation. (ii) *State-dependent effects of stimulation are not well understood*. Stimulation in RNS is closed loop and driven by seizure-like activity, which may fluctuate at circadian and multidien timescales [3]. Previous studies have discovered that the effectiveness of stimulation depends on seizure risk state that varies over multiple days [4]. However, the effect of timing of stimulation over a 24-hour period on seizure reduction is underexplored. (iii) *The effect of stimulation on brain networks is complex and non-linear*. Yet the data available to fit a non-linear model that captures this relationship is limited.

In this study, we investigated inter-ictal ECoG circadian dynamics of hippocampal networks and discovered that they were predictive of ongoing seizure reduction. The circadian dynamics of the network changed over months and the changes were predicted by the timing of stimulation over a 24-hour period. A non-linear Kuramoto model of oscillators mathematically described the effect of stimulation on circadian dynamics. We developed a data efficient model fitting strategy and found that the Kuramoto model explained circadian dynamics better than baseline models and enabled *in silico* optimization of stimulation.

Methods

Data: We studied participants with drug-resistant epilepsy who were treated with RNS through bilateral hippocampal leads. Patient-reported seizure counts were collected longitudinally at follow-up clinic visits and compared to baseline to quantify **seizure reduction** over time.

ECoG features: We investigated chronic interictal ECoG intermittently recorded by the RNS device. Functional connectivity (FC) within the left and right hippocampi was extracted for different bands. We separately tracked the **circadian cycles** of band-specific FC for each hippocampus. Circadian variation in the hour at which FC was strongest (*peak hour*) was measured using Lomb-Scargle analysis for irregularly sampled timeseries. For each band, we computed **circadian**

concordance, the extent to which the left and right hippocampi were cycling in synchrony (Figure 1).

Stimulation data: The RNS device tracks the hourly counts of stimulation events. Stimulation counts were averaged and normalized to obtain normalized mean stimulation at each hour of the day across days. **Non-uniformity of stimulation** (NUS) over 24 hours was computed by summing the normalized mean stimulation values across hours. NUS was also computed between and outside multi-hour intervals.

Kuramoto model: We modeled FC circadian peak dynamics over 6-month window using the non-linear Kuramoto model, incorporating left/right hippocampus FC and hourly stimulation events as oscillators. Coupling constants of the model were fitted to observed circadian cycle and stimulation peaks independently for each 6-month window. We used the fitted model to forecast longitudinal change in circadian peak. Forecasted circadian peaks and observed peaks were compared to estimate prediction error of the model.

In silico stimulation optimization: The Kuramoto model was used to optimize the temporal distribution of stimulation events in a day *in silico*. For a given peak hour of stimulation and configuration of circadian cycles, the Kuramoto model was used to forecast the dynamics of circadian cycle peaks over 6 months. The final peaks of the circadian cycles were used to measure circadian concordance. This process was repeated independently for each hour to find the peak hour of stimulation that maximized circadian concordance for each 6-month window.

Results

Data: We analyzed 100,137 ECoG clips collected over 64,820 days from 38 people who reported their seizure counts a total of 145 times.

Circadian concordance ECoG biomarker: We first built a multivariate model that predicted longitudinal seizure reduction using time-matched band-wise circadian features. Prediction error of the model was 15% better than baseline ($r=0.33$, $P=0.008$). High circadian concordance in the alpha band was associated with a better outcome and was a biomarker of ongoing seizure reduction ($r=0.41$, $P<0.001$, Figure 2).

Stimulation biomarker: We tested whether the distribution of stimulation over the course of the day was predictive of outcome. NUS outside 7AM – 11AM was most predictive of the outcome ($r=-0.50$, $P=0.001$) suggesting time-of-day effects of stimulation. Stimulation peak hour and NUS measures predicted change in circadian cycle peak in the alpha band ($r=0.19$, $P=0.001$). Thus, stimulation modulated circadian cycles and, consequently, circadian concordance.

Kuramoto model: The Kuramoto model explained the modulation of circadian peak in the alpha band over 6-month windows better than baseline models ($P<0.001$, median prediction error=0.02 hours). The model perfectly explained 52% of the samples, i.e., achieved a prediction error of less than one hour across hippocampi, and generalized across treatment responders and non-responders.

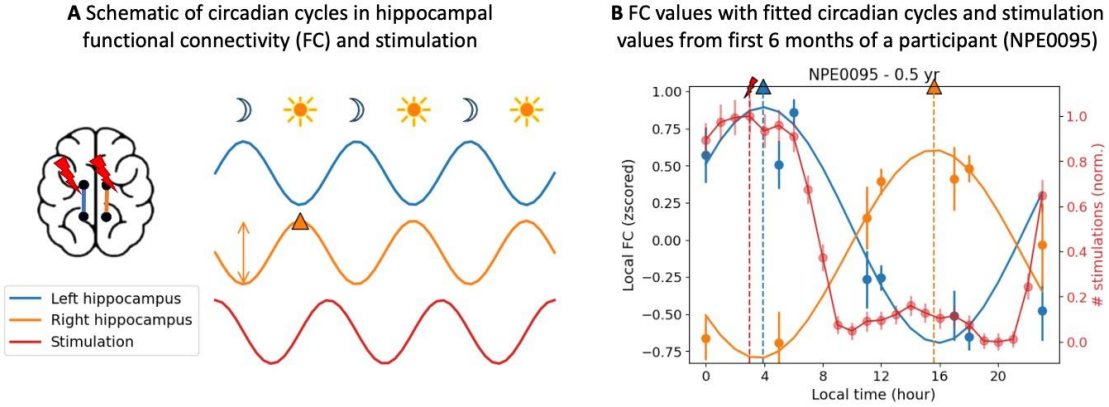
In silico optimization: Varying the peak hour of stimulation in the Kuramoto model altered circadian concordance (Figure 2C). First, we validated the *in silico* analysis. Lower NUS between 7AM – 11AM predicted higher circadian concordance *in silico* ($r=-0.26$, $P<0.001$). This supported the validity of the *in silico* analysis since it matched the empirical observation that lower stimulation predicts higher seizure reduction. Finally, Kuramoto model-driven optimal stimulation achieved high circadian concordance in 85 samples as compared to 39 samples for observed stimulation.

Conclusions

Alignment of circadian cycles in hippocampal network and timing of stimulation can be clinically useful to predict seizure reduction. While further prospective validation is needed, the Kuramoto model provides a testbed for optimizing the timing of stimulation over the circadian rhythm.

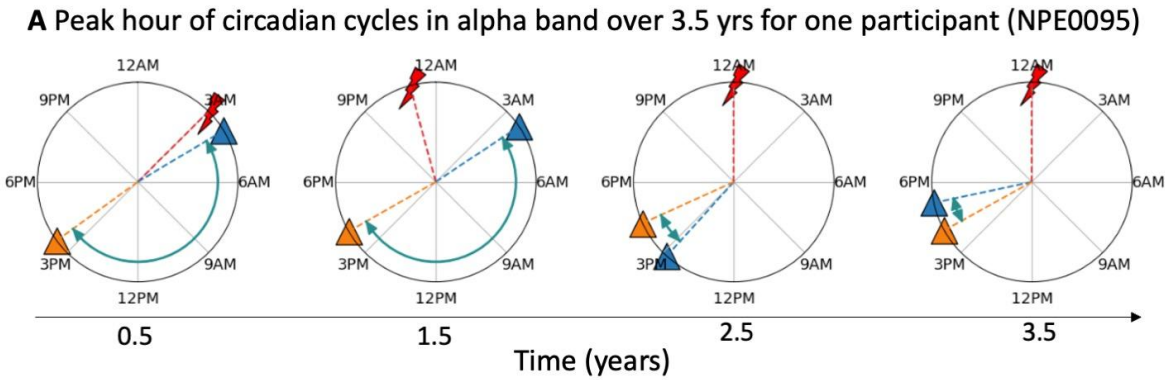
References:
[1] Charlebois, et al., *Epilepsia* 2024.
[2] Khambhati, et al., *Sci Transl Med* 2021.
[3] Baud, et al., *Nat Comm* 2018.
[4] Chiang, et al., *Brain Stimul* 2021.

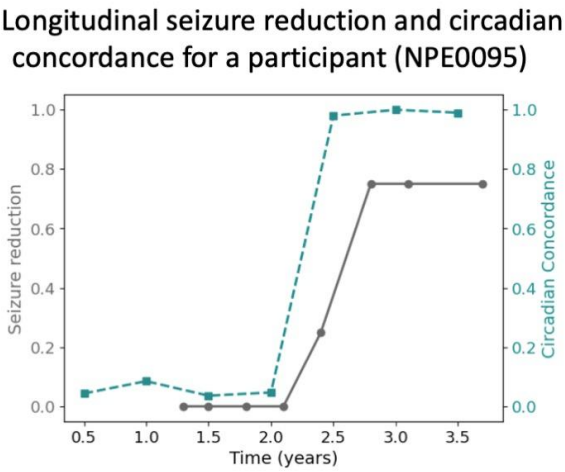
Network circadian cycles and stimulation

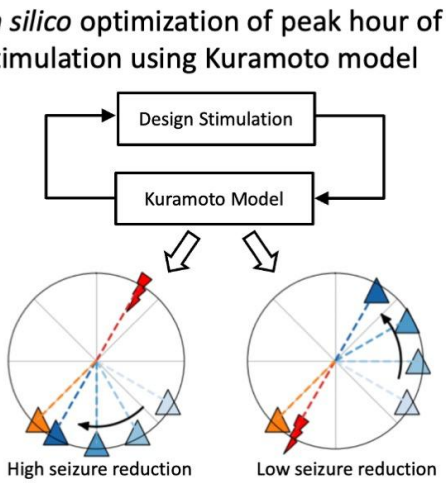


https://secure.key4events.com/onsite/api.aspx?e=1960&img=fig1.jpeg&ai=74086&op=getabstractimg&dirN=0

ECoG biomarker and stimulation optimization







 https://secure.key4events.com/onsite/api.aspx?e=1960&img=fig2.jpeg&ai=74086&op=getabstractimg&dirN=1

P17

Optimizing Machine Learning Models and Interictal Biomarkers for Localization of Epileptogenic Tissue

J. Cimbalnik[1], P. Klimes[2, 1], M. Pail[3, 4], B. Milan[3, 4]

[1]International Clinical Research Center, St Anne's University Hospital, Brno, Czech Republic - Brno (Czechia), [2]Institute Of Scientific Instruments, Czech Academy Of Sciences, Brno, Czech Republic - Brno (Czechia), [3]Brno Epilepsy Center, Medical Faculty Of Masaryk University And St. Anne's University Hospital, Brno, Czech Republic - Brno (Czechia), [4]Central European Institute Of Technology (ceitec), Masaryk University, Brno, Czech Republic - Brno

Introduction

Accurate localization of epileptogenic tissue using interictal intracranial EEG (iEEG) could substantially reduce the need for prolonged monitoring and improve presurgical evaluation. Machine-learning studies in this domain face several methodological challenges, including managing patient-level variability, selecting appropriate validation schemes, and handling severe class imbalance. We aimed to identify the most effective ML model and interictal feature set for epileptogenic tissue localization under rigorously patient-independent conditions and extreme real-world imbalance. We further evaluated how model performance varies across patients, including those with non-lesional or frontal epilepsy.

Methods

We analyzed iEEG recordings from 14 patients with excellent postoperative outcomes (Engel IA). Eleven patients formed the development cohort and three patients were held out completely as an external test set. Contacts were labeled as SOZ-resected (positive) or non-SOZ (negative), resulting in a highly imbalanced ground truth (~1–3% positives per patient).

All features were z-scored within each patient to eliminate patient-level confounding. Extracted interictal features included spike rate, entropy-based metrics, spectral power, power ratios, Hjorth parameters, and higher-order statistical descriptors. A broad set of classifiers (linear and non-linear SVMs, logistic regression, decision trees, random forests, and boosting methods) and hyperparameters were evaluated using nested leave-one-patient-out (LOGO) cross-validation optimized for average precision (AUPRC), an appropriate metric for extreme imbalance.

Feature importance was quantified globally and per patient using permutation importance.

Results

Across the 11-patient development cohort, the best-performing model was a linear SVM with balanced class weighting, achieving a mean F1 score of 0.95 +/- 0.35, median 0.94 and a mean AUPRC of 0.71 +/- 0.29, median 0.81.

On the external three-patient test cohort, individual AUPRC values ranged from 0.27 to 0.58, with a micro-averaged AUPRC of 0.40. Two patients showed moderate-to-high performance (AUPRC ≈ 0.55–0.58, F1 ≈ 0.4-0.44), while one patient achieved markedly lower performance (AUPRC ≈ 0.27, F1 ≈ 0.4). This lowest-scoring case corresponded to a non-lesional frontal epilepsy patient, in whom classical interictal biomarkers were weakly expressed.

Permutation importance (PI) revealed that interictal spike rate (mean PI = 0.393 +/- 0.022) and relative entropy (mean PI = 0.227 +/- 0.025) overwhelmingly dominated predictive performance across patients . Most other spectral and statistical features contributed minimally, indicating that only a small set of interictal biomarkers carries meaningful SOZ discriminability. Performance degraded noticeably in recordings with even mild peri-ictal contamination or reduced signal quality.

Conclusions

Under rigorous patient-independent evaluation and extreme imbalance, interpretable interictal biomarkers—primarily spike rate and entropy—enable meaningful SOZ localization. However, reduced performance in the frontal non-lesional epilepsy patient underscores that classical interictal biomarkers are not universally reliable. These findings highlight the critical importance of clean, high-quality interictal recordings and suggest that challenging surgical cases such as frontal, non-lesional, or low-SNR cases, may require phenotype-specific machine learning models and interictal biomarkers for localization of epileptogenic tissue. Tailoring model architecture or feature engineering to these phenotypes may offer a path toward broader clinical applicability of automated SOZ localization tools.

P18

Seizure Suppression Network: An Intrinsic Mechanism for Seizure Termination and Novel Target for Neurostimulation

A.H. Daraie[1], L.S. Sanchez[2], J.A. González-Martínez[3], A.S. Charles[2], J.Y. Kang[4], S.V. Sarma[2]

[1]Institute For Computational Medicine, Johns Hopkins University - Baltimore, Md (United States), [2]Department Of Biomedical Engineering, Johns Hopkins University - Baltimore, Md (United States), [3]Department Of Neurological Surgery, University Of Pittsburgh School Of Medicine - Pittsburgh, Pa (United States), [4]Department Of Neurology, Johns Hopkins University School Of Medicine - Baltimore, Md (United States) - Baltimore, Md (United States)

Introduction

Over 15 million people worldwide suffer from drug-resistant epilepsy (DRE), with surgical resection and responsive neurostimulation (RNS) offering a final option for seizure control. However, surgical success rates vary between 30-70%, while RNS achieves seizure freedom in fewer than 20% of patients after nine years. Both approaches target the seizure onset zone (SOZ), yet seizures are network phenomena beyond SOZ that involve multiple brain regions that terminate spontaneously through coordinated mechanisms. Prior studies demonstrate that seizure termination involves increasing network synchronization and critical transitions across spatial scales rather than simple energy depletion or inhibitory overload. We hypothesize that a Seizure Suppression Network (SSN) actively emerges during seizures to terminate seizure activity by exerting directional influence over the SOZ. The SSN is functionally influential to the SOZ and thus can inhibit seizures once they start. Identifying and targeting the SSN may revolutionize therapeutic approaches by leveraging the brain's natural seizure control circuitry rather than solely targeting seizure initiation sites directly.

Methods

We analyzed intracranial EEG (iEEG) recordings from three DRE patients across 2 epilepsy centers, and a total of 6 seizures. To characterize network dynamics, we employed state-space model representations fitted to 500ms sliding windows of iEEG data, treating the nonlinear brain as a linear time-varying system, shown in Fig 1A.

From estimated state-transition matrices, we computed directional influence metrics between all brain regions and quantified each node's network role using the Source-Sink Index (SSI), ranging from -1 (pure source: maximum outgoing influence, minimum incoming) to +1 (pure sink: maximum incoming influence, minimum outgoing). We tracked this metric alongside incoming network influence for both SOZ and non-SOZ nodes throughout seizure progression. To characterize each region's interaction with the SOZ, we calculated two complementary metrics: source strength (magnitude of influence to SOZ) and source selectivity (preferential targeting of SOZ over other regions). SSN nodes were defined as regions exhibiting either high source strength or high selectivity toward the SOZ during periods of peak SOZ incoming influence. Statistical validation was performed using logistic regression and Mann-Whitney U tests comparing SOZ versus non-SOZ channels across temporal segments (pre-ictal, ictal phases, post-ictal).

Results

Our three-patient preliminary analysis revealed consistent patterns supporting the existence of an emerging SSN during each seizure. Source-Sink Index analysis showed that SOZ channels exhibited significantly higher SSI compared to non-SOZ channels ($p<0.001$), indicating their role as network sinks (Fig 2A,B). Temporal analysis demonstrated that incoming network influence toward SOZ progressively increased during seizure evolution (Fig 2C). Across seizure progression, 88% of identified SSN nodes exhibited high selectivity or strength to SOZ nodes, with 10-20% showing both high source strength and high selectivity (Fig 2D-F). Notably, both SOZ and non-SOZ regions consistently activated as SSN nodes during mid-to-late seizure phases, demonstrating targeted influence toward SOZ regions before seizure termination (Fig 2E). Temporal analysis revealed SSN activation preceded seizure offset by 10-30 seconds, suggesting a therapeutic intervention window.

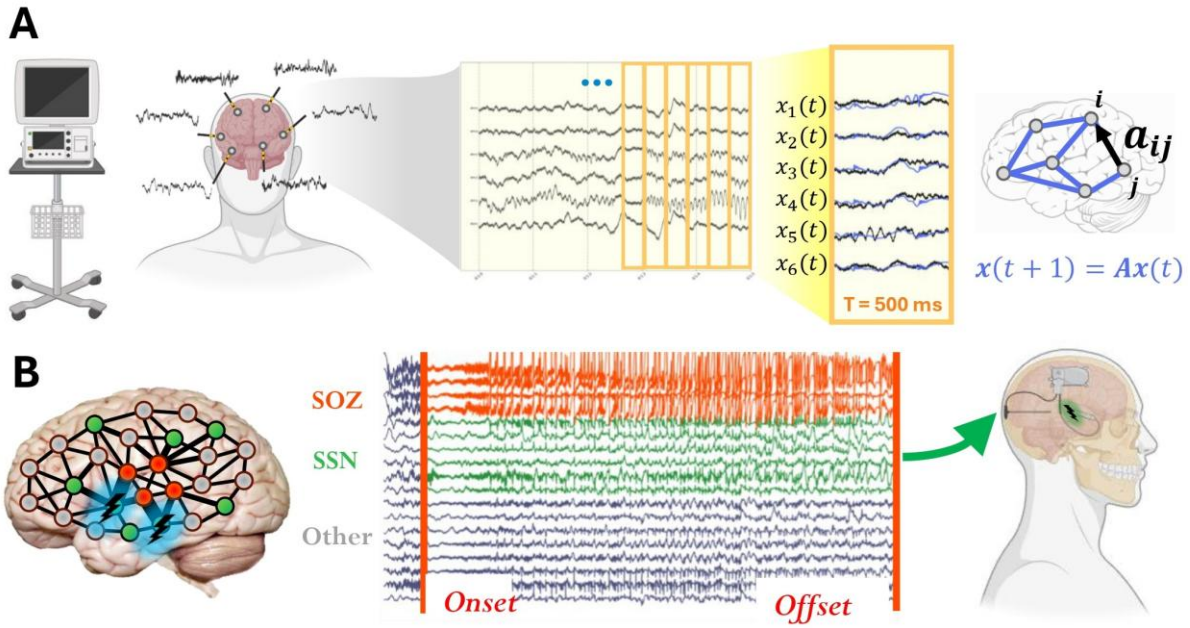
For the RNS case study, we identified thalamic nuclei among the top SSN nodes across multiple seizures, showing peak source strength to clinically annotated SOZ regions during mid-seizure periods; these nodes are highlighted in Fig 2E. These computationally identified locations precisely matched subsequently implanted RNS electrode sites, correlating with the patient's seizure freedom outcome at one-year follow-up. This concordance between computational SSN identification and clinical success provides preliminary validation of the SSN hypothesis.

Conclusions

Our work provides compelling preliminary evidence for a Seizure Suppression Network that actively terminates seizures through coordinated directional influence over epileptogenic regions. Using Source-Sink Index analysis, we demonstrated that SOZ nodes act as network sinks receiving progressively increased incoming influence during seizures, while SSN nodes emerge as selective sources targeting these regions. The SSN framework fundamentally shifts conceptualization from seizure generation to seizure control mechanisms, offering novel therapeutic targets beyond traditional SOZ-focused approaches. Successful RNS outcomes correlating with SSN node targeting suggest that stimulating the brain's intrinsic suppression mechanisms may prove more effective than targeting seizure initiation sites alone.

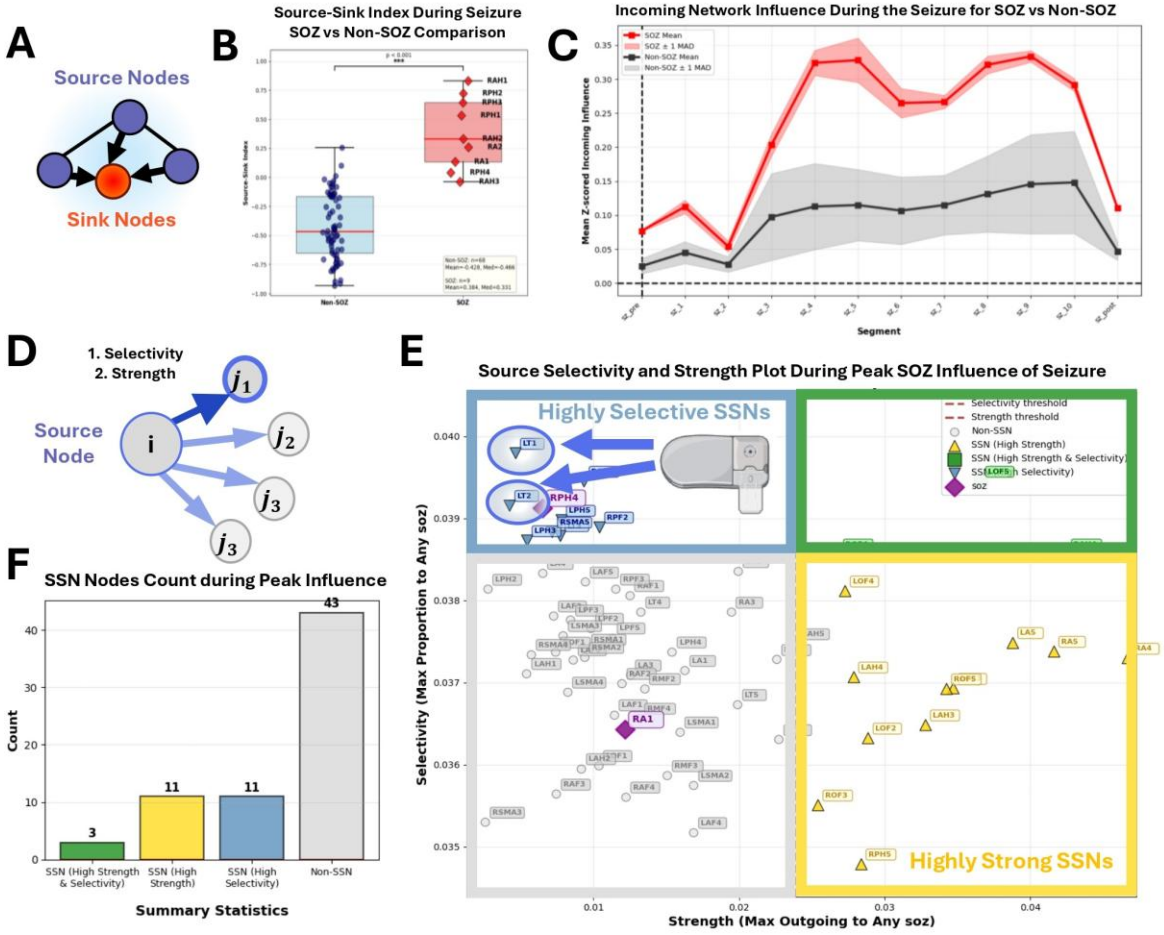
Future work will expand analysis to the complete multi-center cohort, correlating SSN characteristics with RNS outcomes across diverse patient populations and investigating whether thalamic nuclei and other deep brain structures serve as universal SSN components. Validating SSN-guided electrode placement through prospective clinical trials could establish new paradigms for precision targeting in epilepsy neuromodulation, potentially improving seizure freedom rates from current 18-20% toward surgical success rates of 60-70%. This research ultimately aims to translate computational network insights into clinical decision support tools that leverage rather than oppose the brain's natural seizure termination mechanisms.

DNM reveals Seizure Suppression Network target



https://secure.key4events.com/onsite/api.aspx?e=1960&img=figures+AIE2026_page-0001.jpg&ai=74094&op=getabstractimg&dirN=0

SOZ channels are dominant sinks during seizures



https://secure.key4events.com/onsite/api.aspx?e=1960&img=figures+AIE2026_page-0002.jpg&ai=74094&op=getabstractimg&dirN=1

P19
A Self-Supervised Learning Approach for EEG Representation Learning in Epilepsy

R. Saber[1, 2], E. Lemoine[3], D.K. Nguyen[1, 2], E. Bou Assi[1, 2]

[1]Department Of Neuroscience, Faculty Of Medicine, University Of Montreal - Montréal (Canada), [2]Research Center Of The University Of Montreal Hospital Center - Montréal (Canada), [3]New York University - New York (United States)

Introduction

Electroencephalography (EEG) is at the core of the diagnosis, assessment, and longitudinal management of epilepsy. Current machine learning approaches for EEG interpretation rely heavily on supervised learning, which requires large volumes of expert-labeled data. These resources are scarce in clinical epilepsy research. Supervised methods also face important limitations, including the need for extensive annotations, a risk of overfitting on small and imbalanced datasets, and limited generalizability across centers and recording conditions. These challenges are particularly acute in epilepsy diagnosis and prognosis, where annotated EEG datasets are often insufficient in size or sampling diversity. Self-supervised learning (SSL) offers a powerful alternative by leveraging large amounts of unlabeled EEG data to learn robust and transferable representations. In particular, contrastive self-supervised approaches such as Simple Framework for Contrastive Learning of Representations (SimCLR) can extract versatile EEG embedding spaces from raw or minimally processed signals. These representations can then be utilized across diverse downstream tasks, effectively serving as the basis for an EEG foundation model.

Methods

We conducted a retrospective study on a consecutive cohort of patients undergoing routine EEG at a tertiary care center in Montreal, Canada. All patients who underwent routine 20–60-minute EEGs (with or without sleep deprivation) between January 2018 and September 2019 at the Centre hospitalier de l'Université de Montréal (CHUM) were included. The analytical pipeline consisted of two stages. First, a self-supervised pretraining stage was used to learn general-purpose EEG representations using a SimCLR-based contrastive learning framework. Second, the pretrained encoder was fine-tuned indirectly by training downstream classifiers on the learned feature embeddings to perform multiple prediction tasks: epilepsy diagnosis, seizure recurrence prediction, and demographic characteristics (sex and age).

All recordings were preprocessed by applying a 60 Hz notch filter and resampling to 200 Hz. EEGs were segmented into overlapping 60-second windows with a 1-second step size, based on previously established optimal window length for this dataset. Each segment was transformed into a spectrogram using a discrete wavelet transform (DWT). For the SSL stage, two augmented views of each spectrogram were generated using Gaussian noise, time masking, frequency masking, and amplitude scaling. A ResNet-18 encoder with a two-layer feed-forward projection head was trained using a normalized temperature-scaled cross entropy contrastive loss function (NT-Xent) in a SimCLR framework.

In the downstream stage, hidden representations from the pretrained encoder were used to train Random Forest, Extreme Gradient Boosting (XGBoost), and Light Gradient Boosting (LightGBM) classifiers. The cohort was temporally split into training/validation and testing sets, with September 2019 as the split date, resulting in a temporally shifted evaluation. Models were trained at the segment level, and EEG-level predictions were obtained by computing the median of the segment-level probabilities. Performance across all tasks was evaluated using the area under the receiver operating characteristic curve (AUROC). All outcomes were binarized; the age threshold was set at 50 years, and recurrence prediction was assessed on a logarithmic scale at 3, 6, 12, and 24 months. The overall pipeline is depicted in Figure 1. For comparison, we trained a ResNet-18 model in a fully supervised manner for epilepsy diagnosis and subsequently applied transfer learning using the learned representations to predict additional outcomes, including recurrence, sex, and age.

Results

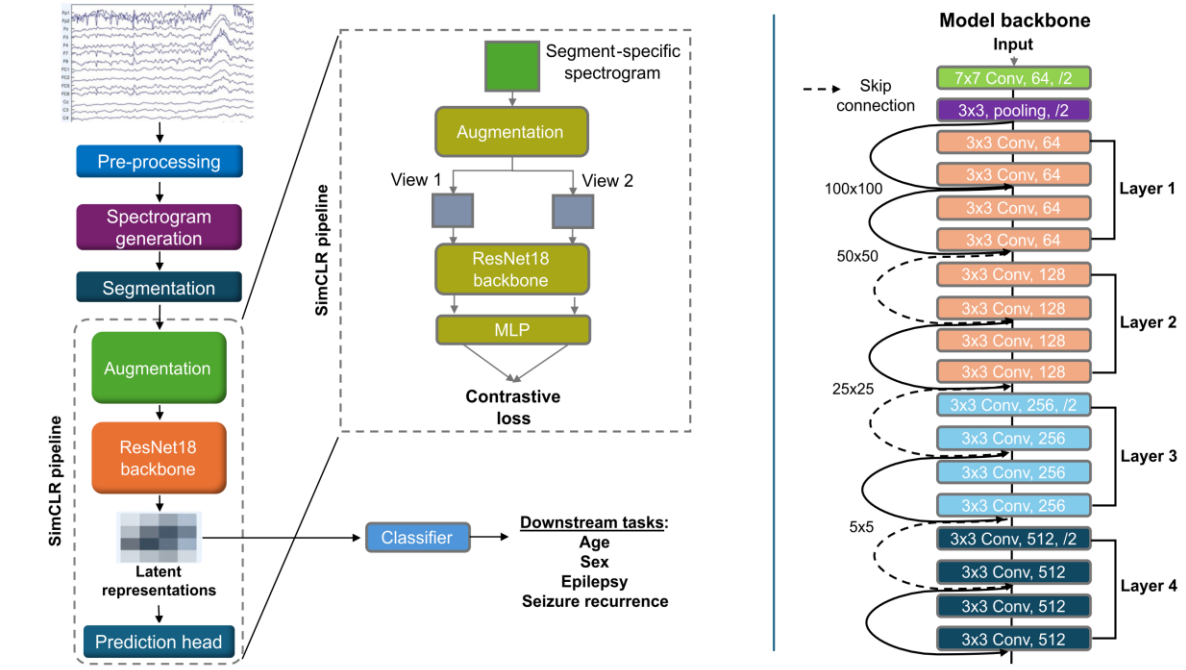
A total of 894 patients with 1,029 routine EEGs were included. The temporal split resulted in 760 patients (865 EEGs) in the training/validation set and 134 patients (164 EEGs) in the testing set. The highest EEG-level AUROC for epilepsy diagnosis was 0.71 using the Random Forest model (segment-level AUROC: 0.63). For seizure recurrence prediction, the best EEG-level AUROCs were: 0.66 at 3 months with LightGBM (segment-level: 0.62), 0.62 at 6 months with Random Forest (segment-level: 0.61), 0.65 at 12 months with LightGBM (segment-level: 0.62), and 0.64 at 24 months with Random Forest (segment-level: 0.64). For demographic prediction, sex was predicted with an EEG-level AUROC of 0.77 using LightGBM (segment-level: 0.74), and age (binary threshold: 50 years) achieved an AUROC of 0.71 using XGBoost (segment-level: 0.65). The EEG-level and segment-level receiver operating characteristic curves obtained with the SSL approach across all downstream tasks are presented in Figure 2. In comparison, a ResNet model trained for epilepsy diagnosis achieved a segment-level AUROC of 0.58, which is 8.6% lower than the performance obtained with the SSL approach. Furthermore, supervised transfer learning, performed using the representations learned by the ResNet model for epilepsy diagnosis to address additional tasks, consistently underperformed relative to the SSL representations. The best AUROC values obtained with supervised transfer learning were 0.58 for 24-month recurrence prediction, 0.61 for 12-month recurrence, 0.60 for 6-month recurrence, and 0.54 for 3-month recurrence. Similarly, the ResNet-based representations yielded only 0.69 for sex prediction and 0.69 for age prediction, in terms of segment-level AUROC.

Conclusions

This study demonstrates that a SimCLR-based SSL approach can learn general-purpose, transferable EEG representations that support multiple downstream clinical and demographic prediction tasks. The learned embeddings held meaningful information across epilepsy diagnosis, seizure recurrence prediction, and patient characteristic estimation, suggesting that SSL can serve as a foundation for developing a generalized EEG representation model. These results lay the groundwork for an EEG foundation model capable of supporting diverse clinical tasks including diagnosis, prognosis, and phenotyping. One limitation of this work is that comparisons were performed only against a ResNet-18 baseline. Future studies should evaluate SSL representations against a broader range of backbone architectures. Future steps should also leverage larger public and private unlabeled EEG datasets, exploring transformer-based backbone architectures for SimCLR pretraining, and expanding downstream task diversity. Clinically, such a unified representation model has the potential to streamline assistive EEG-based decision support by enabling a single foundation model to perform multiple predictive tasks.

Overall pipeline of the proposed SSL approach

Figure 1. Overall pipeline of the proposed SimCLR-based approach for EEG Representation Learning in Epilepsy

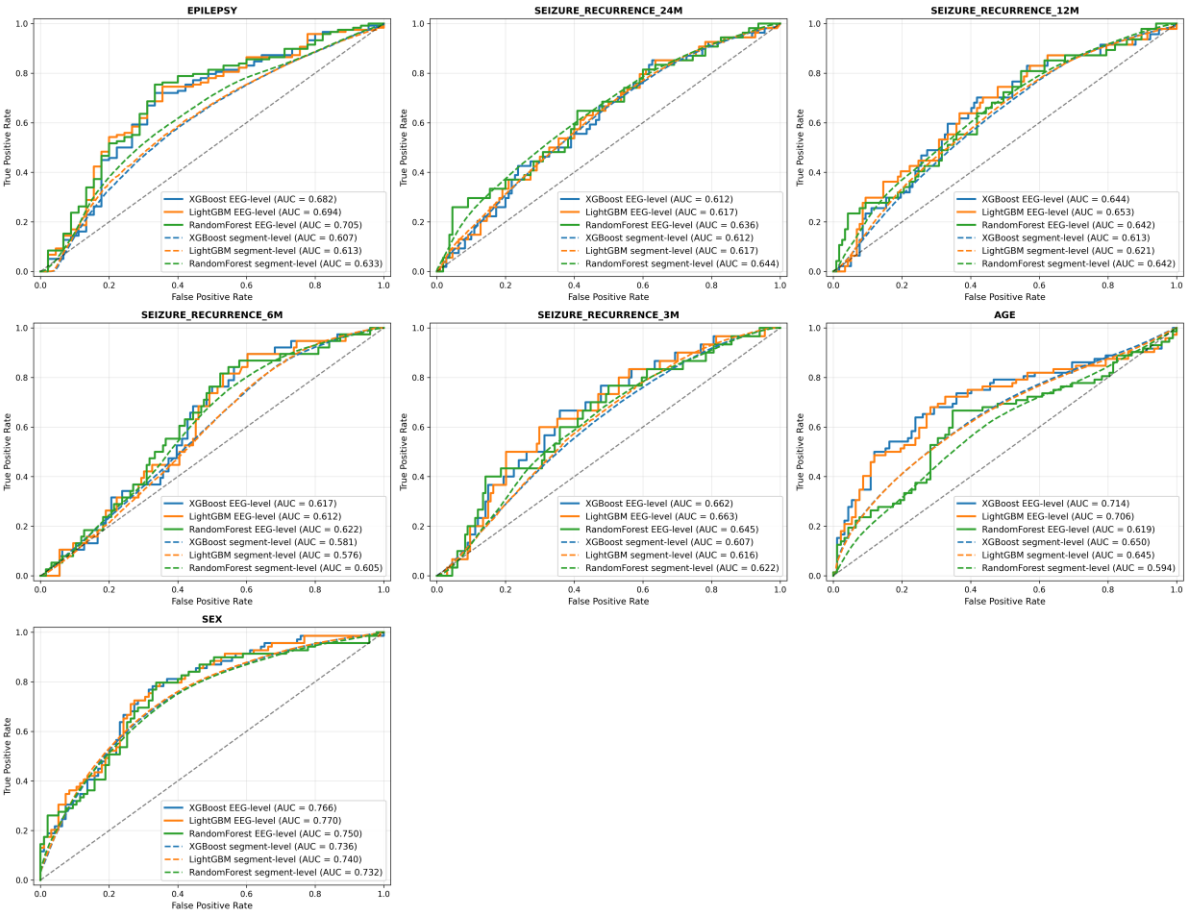


SimCLR, Simple Framework for Contrastive Learning of Representations; MLP, Multilayer Perceptron; Conv, Convolutional layer

https://secure.key4events.com/onsite/api.aspx?e=1960&img=Figure+1.png&ai=74095&op=getabstractimg&dirN=0

ROC curves of the SSL approach on downstream tasks

Figure 2. Receiver Operating Characteristic Curves for the SSL Approach at the Segment and EEG Levels for All Downstream Prediction Tasks



https://secure.key4events.com/onsite/api.aspx?e=1960&img=Figure+2.png&ai=74095&op=getabstractimg&dirN=1

P20
A Novel Biomarker of Anti-Seizure Medication Efficacy from Routine Scalp EEG

S. Pandiaraju[1], S.S. Sarma[1], J. Kang[2]

[1]Johns Hopkins University - Baltimore (United States), [2]Johns Hopkins Hospital - Baltimore (United States)

Introduction

Epilepsy remains a pervasive neurological burden, affecting approximately 70 million individuals globally. Despite the availability of numerous anti-seizure medications (ASMs), the standard of care remains a process of trial and error. Clinicians are forced to "titrate in the dark," relying on seizure diaries that are prone to under-reporting and therapeutic drug monitoring (TDM) that measures drug-serum levels rather than neural changes. Consequently, patients often endure months or years of ineffective ASM therapy, sustaining injuries, worsening psychosocial comorbidities, and facing an elevated risk of Sudden Unexpected Death in Epilepsy (SUDEP) before a successful regimen is identified or drug resistance is confirmed.

Current objective monitoring tools are insufficient for bridging this gap. Routine scalp electroencephalography (rsEEG) relies on qualitative, visual inspection of a brief, 20-minute snapshot of brain activity. Because interictal abnormalities are intermittent and seizures are rarely captured during such short recordings, standard visual analysis provides limited insight into a patient’s medication responsiveness. PharmacoEEG has attempted to quantify drug effects, but traditional analyses have primarily focused on static, channel-level features such as absolute or relative bandpower, coherence between a small number of electrode pairs, or summary spectral ratios. These features are considered “static” because they collapse the inherently time-varying and state-dependent behavior of the epileptic brain into single, time-averaged values. As a result, they fail to capture the dynamic network interactions and rapid fluctuations in connectivity that characterize a seizure-susceptible brain.

Methods

The innovation of this study lies in the application of control-theoretic network models to quantify the dynamical network organization of the epileptic brain. We model rsEEG as a linear time-varying system, where the state-transition matrix A(t) encodes how activity propagates across channels from one time step to the next. From A(t), we can derive the brain’s modes which represent the main directions or “dimensions” along which activity spreads across channels. We hypothesize that, in untreated and poorly treated epilepsy, this propagation is dominated by a small number of strongly constrained modes, resulting in signal redundancy that is consistent with pathological hypersynchrony. On the other hand, effective ASMs redistribute activity evenly across a richer set of modes. To test this, we compute Singular Value Spectral Entropy (SVSE) (Figure 1). Derived from the Shannon entropy of the normalized singular values of A(t), SVSE quantifies the diversity of these functional patterns. Low values indicate the rigid, redundant dynamics of the epileptic state, while high values reflect the distributed, resilient activity associated with effective treatment. Therefore SVSE serves as a quantitative biomarker that can distinguish between brain networks associated with effective versus ineffective antiseizure medication states.

We utilized a heterogeneous retrospective cohort of 15 patients with focal epilepsy on a variety of medications and dosages from Johns Hopkins Hospital. The experimental approach leveraged a rigorous, automated computational pipeline designed to extract high-fidelity dynamical features from noisy, clinical-grade rsEEG recordings.

Recordings were categorized based on clinical outcomes at the time of the EEG:
1. No med: Drug-naive or off-medication.
2. Ineffective: Patients experiencing seizures within 6 months of their previous seizure despite ASM therapy.
3. Effective: Patients with seizure freedom for at least 6 consecutive months
4. Weaned: Patients who were previously controlled but experienced seizure relapse upon weaning off medications.

Results

The analysis of SVSE revealed distinct, quantifiable differences in network dynamics correlated with therapeutic status. As visualized in our primary analysis (refer to Figure 2), there is a clear stratification of SVSE across treatment groups. The Ineffective medication group displayed significantly lower SVSE compared to the Effective medication group ($p=0.01$). This statistically distinct separation supports our hypothesis: ineffective treatment states are characterized by lower complexity and a more restricted state-space (lower SVSE), reflecting the underlying epileptogenic potential. Conversely, effective treatment shifts the network toward higher SVSE, implying a restoration of complex, diverse neural interactions and a move away from pathological synchronization.

The "Weaned" group exhibited SVSE values intermediate to the Ineffective and Effective groups. Although sample size is small ($N=2$), this observed trend is suggestive of re-emergence of epileptogenic state after medication wean.

Longitudinal analysis further validates biomarker's sensitivity. Subject 016 (represented by green triangles) demonstrated consistently high SVSE values while in the "Effective" category, outperforming the group mean. Similarly, Subject 017 (represented by squares) showed distinct SVSE states correlating with their clinical status. This intra-patient consistency suggests that DNM-derived fragility and entropy metrics are not merely population-level statistics but are viable for personalized tracking.

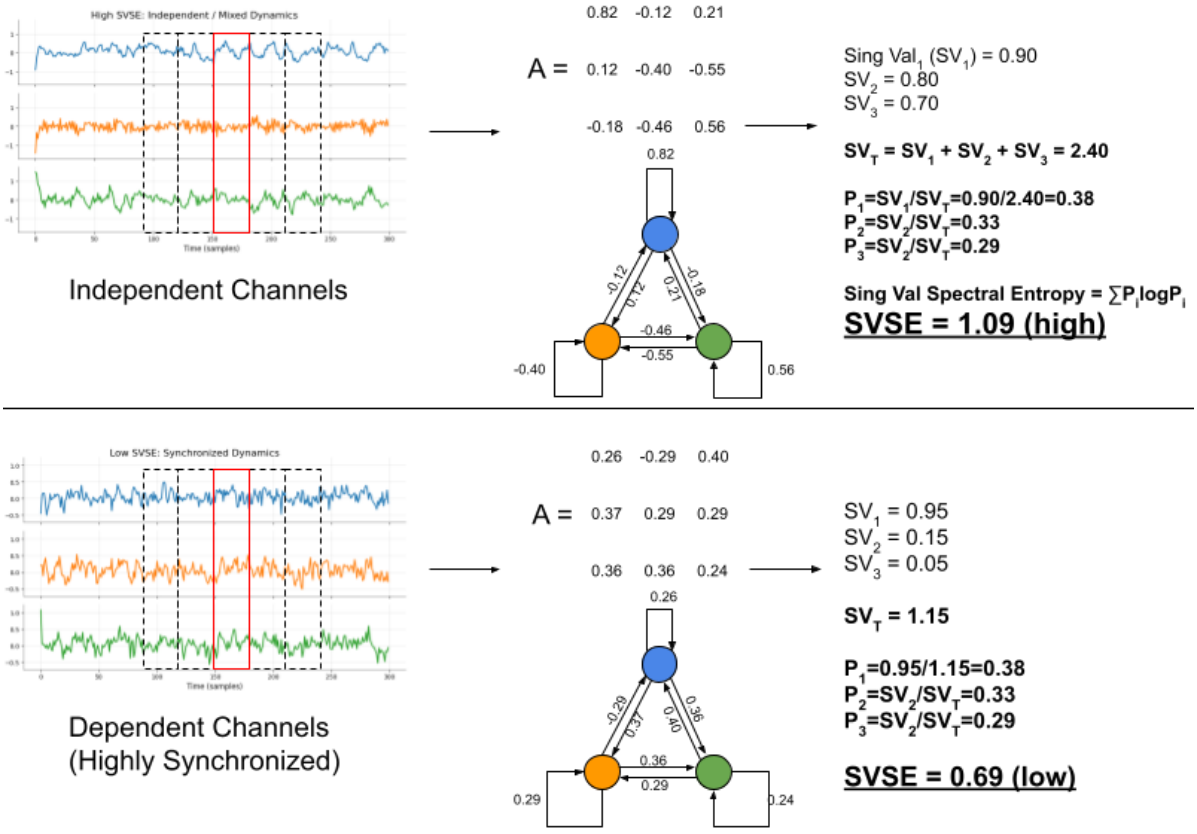
Conclusions

This study demonstrates that Singular Value Spectral Entropy (SVSE), derived from Dynamic Network Models of routine scalp EEG, is a promising and objective biomarker for assessing antiseizure medication (ASM) efficacy. By moving beyond static, channel-level features and instead quantifying the underlying dynamics of brain networks, SVSE captures meaningful differences between effective and ineffective therapeutic states. Our findings show that effective pharmacologic treatment is associated with broader engagement of multiple brain activity patterns, reflected by higher SVSE values.

Future work will focus on expanding this framework across several key areas. Increasing the number of patients and recordings will allow us to validate SVSE in larger and more diverse cohorts, helping to account for inter-individual variability. Stratifying analyses by focal epilepsy subtype, seizure onset zone, and relevant comorbidities will clarify whether distinct network signatures correspond to specific etiologies. Longitudinal studies will evaluate how SVSE evolves during medication titration, antiseizure medication withdrawal, or presurgical evaluation, providing insight into its sensitivity to temporal changes in treatment response. In parallel, predictive models

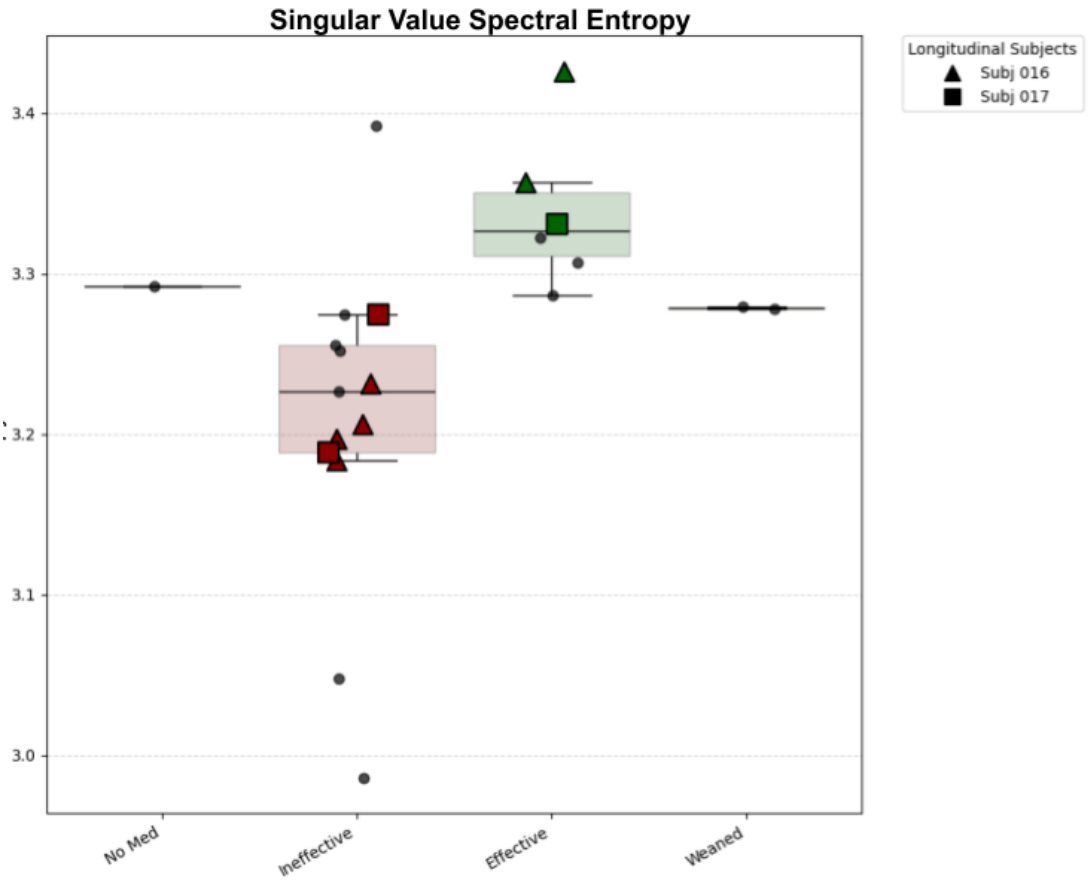
leveraging baseline or early-treatment SVSE values may help forecast future medication efficacy, enabling clinicians to make more informed decisions earlier in the care pathway. Finally, integrating SVSE with traditional pharmaco-EEG metrics and clinical variables will determine whether multimodal biomarker approaches offer additional gains in accuracy. Collectively, these efforts will further establish SVSE as a clinically actionable biomarker capable of advancing personalized epilepsy management.

Figure 1: Conceptualization of SVSE Computation



https://secure.key4events.com/onsite/api.aspx?e=1960&img=AI+in+Epilepsy+2026+Abstract.png&ai=74097&op=getabstractimg&dirN=0

Figure 2: Average SVSE Across Experimental Groups



https://secure.key4events.com/onsite/api.aspx?e=1960&img=SVSE-12-1.png&ai=74097&op=getabstractimg&dirN=1

P21

Higher-Level Eye Tracking Metrics Reveal Search Behaviour Differences in Persons with Epilepsy vs. Healthy Controls

A. Jansen[1], M. Steininger[1], S. Al-Haj Mustafa[1], N. Sido Bouzan[1], R. Surges[1], C. Helmstaedter[1], R. Von Wrede[1], B. Krüger[1, 2]

[1]Department Of Epileptology, University Hospital Bonn - Bonn (Germany), [2]Department Of Computer Science University Of Bonn - Bonn (Germany)

Introduction

Cognitive impairment is a frequently overlooked yet substantial burden for persons with epilepsy (PWE). This includes attention, working memory, and executive function, domains that are also related to coordinating eye movements and visual search [1].

Analyses of eye tracking data collected during the Trail Making Test (TMT), a subtest of the EpiTrack®, showed significant correlations between eye tracking measures (e.g., fixation count, pupil diameter etc.) and cognitive outcomes (e.g., EpiTrack score, completion times). This indicates that eye movement patterns provide insights into underlying cognitive processes and, conversely, that changes in cognitive functioning can be observed in eye behaviour [2].

However, clinical outcomes and raw eye tracking data provide limited insight into how PWE search, select, and process visual information. Higher-level metrics can characterise search efficiency and strategy beyond these traditional measures.

In this study, we developed and applied five higher-level visual search metrics to get information about cognitive dysfunction beyond the EpiTrack and to discover differences in search strategies between healthy controls and PWE. Specifically, we quantified the proportion of fixations and fixation time on relevant targets, failures to recognise search targets, and differentiated between global and local search behaviour.

Methods

2.1 Data Collection

The study included 96 persons with epilepsy (42.2 ± 17.6 years; 45 female, 51 male) and 36 healthy controls (30.9 ± 15.2 years; 25 female, 11 male) from the Department of Epileptology at the University Hospital Bonn. All participants completed the EpiTrack TMT subtasks, connecting nodes in ascending order (numbers in TMT-A; alternating numbers and letters in TMT-B). Each node selection is defined as one visual search task. During the task, participants wore Pupil Labs Neon mobile eye tracking glasses. The TMT sheets were instrumented with fiducial markers at the corners, enabling mapping of gaze positions onto the test sheet via computer vision algorithms. All data were annotated with timestamps indicating when the participant connected each node. For each task (TMT-A / TMT-B), we defined areas of interest (AOIs) on the sheet to categorise fixations into relevant, potentially relevant, irrelevant, and neutral. This forms the basis for all of our higher-level visual search metrics.

2.2 Visual Search Metrics

We quantified visual search efficiency and strategy using the following higher-level eye-tracking metrics:

Percentage of Relevant Fixation Count (PoRF): proportion of all fixations that fell on AOI labelled as relevant or potentially relevant; quantifying the proportion of fixations directed towards task-relevant targets.

Percentage of AOI Fixation Duration (PoD): proportion of total fixation duration spent on each AOI category (relevant, potentially relevant, irrelevant, neutral) divided by test duration

Unrecognized Targets (UT): rate of instances in which the currently relevant target node was fixated but not connected; capturing failures to recognise fixated targets.

Mean Fixation Distance (mFD): average Euclidean distance between successive fixation positions over the entire test duration; reflecting absolute spatial step size of the visual search.

Mean Fixation Distance relative to Target Distance (mFDTD): mean distance between successive fixations normalised by the distance between the current node and the next target node; capturing spatial step size relative to task demands.

2.3 Data Analysis

Data was annotated using the Pupil Labs cloud platform and further processed with Python for statistical analysis. All group comparisons between healthy controls and PWE were age-adjusted. Specifically, for each higher-level eye-tracking metric, we first regressed the metric on age across all participants, obtained the residuals, and then compared groups using two-sided Mann–Whitney U tests on these residuals, given non-normal distributions and unequal group sizes. For each variable, we quantify the age-adjusted group difference using a signed rank-biserial correlation as an effect size. To account for multiple comparisons across metrics, p-values were corrected using the Benjamini–Hochberg false discovery rate procedure, with statistical significance defined as $q < .05$.

Results

The analysis revealed several robust group differences in higher-level visual search metrics between PWE and controls. PoD on relevant AOIs was higher in controls than in PWE (PoD_R: $p = 0.009$, $q < 0.019$, $r_{rb} = 0.295$), indicating that controls directed their gaze towards relevant areas for longer periods, resulting in a more consistent allocation of visual attention to task-relevant information. In contrast, PoRF did not differ significantly between groups ($p = .998$, $q = .998$, $r_{rb} < 0.01$).

PWE showed a higher UT rate than controls ($p = .044$, $q = .058$, $r_{rb} = -0.23$). They more often fixated the correct target without recognising it as the next item to connect, leading to more missed targets relative to task demands and less efficient use of fixations.

Controls had higher mFD values ($p = .016$, $q = .026$, $r_{rb} = 0.27$), indicating a more global search strategy. This pattern remained for the target-normalised distance: mFDTD was higher in controls than in PWE ($p = .008$, $q = .016$, $r_{rb} = 0.31$), consistent with broader exploration in controls, whereas PWE relied on shorter, more local fixation shifts.

Conclusions

PWE show measurable differences in visual search behaviour compared to healthy controls, characterised by shorter fixation on relevant targets, more unrecognised targets, and a shift towards more local search patterns. These effects are small to moderate in size but consistent across multiple higher-level metrics. Beyond conventional outcome measures (EpiTrack), our metrics offer a process-level view of *how* PWE search for and select information by contrasting their behaviour with patterns observed in healthy controls. This may help distinguish specific cognitive profiles (e.g., attentional vs. planning deficits) and track changes over time. Together, our findings position higher-level eye tracking metrics as a promising basis for more fine-grained cognitive screening and for future AI-based decision support in epilepsy care.

Literature

[1] C. E. Elger, C. Helmstaedter, and M. Kurthen, “Chronic epilepsy and cognition,” The Lancet Neurology, vol. 3, no. 11, pp. 663–672, Nov. 2004.

[2] S. Al-Haj Mustafa, A. Jansen, M. Steininger, J. Müllers, R. Surges, R. von Wrede, B. Krüger, and C. Helmstadter, “Eyes on cognition: Exploring oculomotor correlates of cognitive function in patients with epilepsy,” Epilepsy & Behavior, vol. 173, p. 110562, 2025, doi: 10.1016/j.yebeh.2025.110562.

P22

Linking Higher-level Eye Tracking Metrics to High-Impact Antiseizure Medication in Epilepsy Patients

M. Steininger[1], A. Jansen[1], S. Al-Haj Mustafa[1], N. Sido Bouzan[1], R. Surges[1], C. Helmstaedter[1], R. Von Wrede[1], B. Krüger[1, 2]

[1]Department Of Epileptology, University Hospital Bonn - Bonn (Germany), [2]Department Of Computer Science, University Of Bonn - Bonn (Germany)

Introduction

People with epilepsy (PWE) face a dual burden: they experience recurrent seizures and often show cognitive impairment affecting attention, memory, and executive functions [1]. Seizures can be controlled with antiseizure medication (ASM), however, these drugs can further worsen cognitive performance, especially at higher doses or when several ASM are combined [2]. This creates a therapeutic trade-off between seizure control and cognitive function.
To characterize this trade-off in clinical practice, the EpiTrack® provides a standardized assessment of attention and executive function and is widely used to monitor cognitive performance and potential ASM-related cognitive side effects. Eye tracking data collected during the Trail Making Test (TMT), such as fixation duration, saccade velocity, and pupil diameter, systematically correlate with EpiTrack performance in PWE [3]. These findings support eye tracking as a sensitive, non-invasive marker of cognitive impairment in this population.
Building on this work, we focus on higher-level eye tracking metrics that describe visual search strategies rather than isolated oculomotor parameters. These metrics summarize how PWE explore a visual scene, for example, how systematically they scan relevant targets versus distractors, how often they revisit already inspected items and if they search in a local or global distance. If ASM alter search strategies, these changes are likely to translate into everyday difficulties, providing a clinically meaningful indicator of how strongly PWE are affected.
The present study examines how specific ASM profiles relate to eye tracking-based search behavior. We weighted medication according to their known cognitive impact and link these weighted medication profiles to higher-level eye tracking metrics. This approach aims to clarify how pharmacotherapy contributes to cognitive alterations at the level of visual search and to evaluate whether search strategy metrics can serve as sensitive markers of ASM-related cognitive side effects.

Methods

Eye tracking data of 96 PWE (42.21 ± 17.62, 51 male, 45 female) from the Department of Epileptology at the University Hospital Bonn was collected during the TMT test of the EpiTrack. Participants wore Pupil Labs Neon mobile eye tracking glasses. The TMT sheets were instrumented with fiducial markers at the corners and gaze positions were mapped onto the test sheet via Pupil Labs clouds' computer vision algorithms. We calculated correlations between all higher-level eye tracking metrics and two values expressing how much PWE are impacted by their ASMs. Due to non-normal distributions, we used Spearman rank correlations.

The following higher-level eye tracking metrics were calculated:

Percentage of Relevant Fixation Count (PoRF): proportion of all fixations that fell on areas of interest (AOI) labelled as relevant or potentially relevant; quantifying the proportion of fixations directed towards task-relevant targets.

Percentage of AoI Fixation Duration (PoD): proportion of total fixation duration spent on each AOI category (relevant, potentially relevant, irrelevant, neutral)

Unrecognized Targets (UT): rate of instances in which the currently relevant target node was fixated but not connected; capturing failures to recognise fixated targets.

Mean Fixation Distance (mFD): average Euclidean distance between successive fixation positions over the entire test duration; reflecting absolute spatial step size of the visual search.

Mean Fixation Distance relative to Target Distance (mFDTD): mean distance between successive fixations normalised by the distance between the current node and the next target node; capturing spatial step size relative to task demands.

The medication values were calculated:

High-impact ASM (hi-ASM) value: value ranging from 0 to 2, 0 if the patient is not taking any high-impact ASMs, 1 if the patient is taking at least one medium-impact ASM (valproic acid, carbamazepine, phenobarbital, clobazam) or 2 if the patient takes at least one high-impact medication (zonisamide or topiramate).

Medication Count: number of different ASM PWE take.

Results

The hi-ASM metric showed a significant positive correlation with PoRF ($\rho = 0.252$, $p = 0.048$). PWE rated as having high-impact ASM thus fixated relevant targets slightly more often. There was also a positive correlation between hi-ASM and UT ($\rho = 0.365$, $p = 0.003$) indicating that PWE with high-impact medication more often fixated, but did not recognize relevant targets. In contrast, there were no significant correlations between the hi-ASM and PoD across all AOI categories suggesting that problematic medication does not systematically change how long PWE fixate, but rather how effectively these fixations are used. For both mFD and target-normalised mFDTD, no significant associations with hi-ASM were found. None of the higher-level eye tracking metrics showed a significant correlation with medication count.

Conclusions

Our findings suggest that ASM-related cognitive side effects are reflected in specific aspects of visual search. PWE receiving high-impact ASM showed higher proportions of fixations on relevant or potentially relevant areas and, at the same time, more unrecognised targets, indicating that their fixations were used less efficiently rather than less often or for shorter periods. In contrast, fixation durations and spatial step sizes (mFD, mFDTD) were not systematically affected, this means the overall search pattern is not distinguishable from other PWE. Further, we found no association between medication count and search metrics, indicating that the specific ASM rather than the amount of different medications contributes to the altered search behaviour. Together, these findings indicate that eye tracking–based search metrics, particularly UT and PoRF, can capture subtle ASM-related cognitive side effects and may complement existing clinical tools such as the EpiTrack in characterising the cognitive side effects of pharmacotherapy.

Literature

[1] C. E. Elger, C. Helmstaedter, and M. Kurthen, “Chronic epilepsy and cognition,” The Lancet Neurology, vol. 3, no. 11, pp. 663–672, Nov. 2004.

[2] R. J. Quon, M. T. Mazanec, S. S. Schmidt, A. S. Andrew, R.M. Roth, T. A. MacKenzie, M. Sajatovic, T. Spruill, and B. C. Jobst, “Antiepileptic drug effects on subjective and objective cognition,” Epilepsy & Behavior, vol. 104, p. 106906, 2020, doi: 10.1016/j.yebeh.2020.106906.

[3] S. Al-Haj Mustafa, A. Jansen, M. Steininger, J. Müllers, R. Surges, R. von Wrede, B. Krüger, and C. Helmstadter, “Eyes on cognition: Exploring oculomotor correlates of cognitive function in patients with epilepsy,” Epilepsy & Behavior, vol. 173, p. 110562, 2025, doi: 10.1016/j.yebeh.2025.110562.

P23

MRI-negative epilepsy: Machine-learning distinction of the epileptogenic zone from FLAIR-T2 MRI

D. Stastna[1], J. Van Gompel[1]

[1]Neurosurgery, Mayo Clinic - Rochester (United States)

Introduction

MRI-negative epilepsy constitutes 20–30% of focal epilepsy and poses substantial diagnostic challenges. The designation “MRI-negative” is relative, as up to 39% of patients initially deemed lesion-negative on conventional protocols are later found to harbor subtle abnormalities on advanced post-processing. Thus, many cases reflect technical rather than true pathological absence. The most common substrates include focal cortical dysplasia (≈45%), gliosis (≈22%), hamartia with gliosis (≈13%), and less frequently hippocampal sclerosis (≈9%). Epileptogenic foci frequently involve mesial temporal, operculoinsular, and limbic/paralimbic regions, and patients often require invasive stereo-EEG. Advanced structural and functional imaging—including FDG-PET, SISCOM, MEG, and ASL—demonstrates high diagnostic accuracy, underscoring the limitations of routine MRI.

We hypothesized that epileptogenic cortex in MRI-negative epilepsy exhibits a reproducible quantitative signature on standard FLAIR-T2 MRI, and that a supervised machine-learning model trained on paired epileptogenic and contralateral regions could identify an interpretable radiomic “epi-fingerprint.”

Methods

We retrospectively included 20 patients with drug-resistant MRI-negative focal epilepsy who underwent SEEG. Preoperative FLAIR-T2 MRIs were pre-processed using N4 bias-field correction, rigid registration, standardized resampling, and z-score normalization. Seven filter families (e.g., Haar wavelet, Laplacian-of-Gaussian) enhanced textural contrast. SEEG CTs for each patient were co-registered to MRI, and epileptogenic (EPI) ROIs were defined as 3-mm spheres around ictal contacts, with mirrored contralateral NON-EPI ROIs to ensure pairing.(ITK SNAP; Python 3.14.0.).

From each ROI, 1,441 radiomic and deep-learning features were extracted, including first-order intensity measures, multi-scale texture features (GLCM, GLRLM, GLSZM,GLDM,NGTDM), shape metrics, and radial descriptors capturing core–rim gradients. Deep-learning features were derived using a compact 3D ResNet-18 (Python, 3.14.0).

To augment the small cohort, we applied conservative geometric augmentation to 5 original patients and an ADASYN-like procedure generating ~50% additional samples in training folds only.

An XGBoost classifier distinguished EPI from NON-EPI ROIs using nested grouped cross-validation. Outer 5-fold GroupKFold estimated generalizability; inner folds optimized hyperparameters. Evaluation used AUC, accuracy, precision, recall, F1-score, and bootstrapped confidence intervals. Per-patient “pairwise accuracy” assessed discrimination within matched

pairs. SHAP values ranked feature importance, supported by paired EPI–NON-EPI statistical comparisons.

Results

Our retrospective analysis included 20 patients (median age 26 years, range 17–67; 13 female) with drug-resistant MRI-negative focal epilepsy. Semiology was distributed across mesial temporal (n = 11) and frontal (n = 9) presentations. SEEG localized the epileptogenic zone (EZ) predominantly to the mesial temporal (n = 10), frontal (n = 7), medial cingulate (n = 2), and lateral temporal (n = 1) regions, with right-sided predominance (n = 14). All left-sided mesial temporal EZs were treated with LITT (n = 3), while right-sided management included LITT (n = 3), open resection (n = 3), RNS implantation for bilateral hippocampal involvement (n = 1), and ANT-DBS following failure of left LITT (n = 1). Among frontal lobe epilepsy cases, six patients underwent resection, four received VNS, two underwent RNS, and several underwent DBS cortical stimulation. All patients had gliosis as the dominant histopathological substrate when tissue was available, with no evidence of focal cortical dysplasia or sclerosis.

Overall, seven patients achieved seizure freedom, including three after anterior temporal lobectomy with amygdalohippocampectomy, three following frontal resections, and one after right mesial temporal LITT. An additional five patients achieved Engel Class II, three were Engel Class III, and five were awaiting definitive treatment at the time of analysis.

In the augmented paired dataset (25 EPI and 25 contralateral NON-EPI ROIs from 20 original and 5 synthetically generated patients), the grouped XGBoost classifier demonstrated robust discrimination despite visually normal MRI. Across outer folds, the mean ROC AUC was 0.94, with accuracy 0.82, F1-score 0.82, recall 0.81, and precision 0.83. Bootstrap analysis on pooled outer-test predictions yielded an AUC of 0.95 (95% CI 0.88–0.99), accuracy 0.82 (95% CI 0.73–0.90), and F1-score 0.82 (95% CI 0.70–0.91). On a patient level, the model correctly assigned a higher epileptogenic probability to the EPI ROI than to its contralateral counterpart in 27/30 evaluable patients (90%), indicating a consistent individual-level imaging signature.

SHAP-based feature ranking showed that the classifier relied predominantly on three feature families. First, high-frequency intensity asymmetry, dominated by wavelet-HHH first-order skewness and related skewness features, indicated that EPI ROIs exhibited heavier right-tailed intensity distributions, reflecting a subset of abnormally bright voxels within otherwise normal-appearing cortex. Second, fine-scale microtextural heterogeneity, captured by GLSZM and GLRLM measures such as wavelet-HLL_glszm_SmallAreaLowGrayLevelEmphasis, wavelet-HHH_glrlm_RunEntropy, and LBP-based run entropy, suggested many small, irregular low-intensity zones and highly disordered texture. Third, abnormal radial core–rim organization, including intensity–radius regression slope (roi_rad_reg_slope), core/outer shell ratios (roi_rad_ratio_core_outer), and a deep-learning–derived rim/core activation ratio (DL_RCAR_rim_core_ratio), indicated disrupted radial gradients with a focal abnormal core and heterogeneous rim. Shape parameters contributed modestly but were less influential. Paired statistical analysis confirmed that these intensity, texture, and radial descriptors were systematically shifted between EPI and NON-EPI ROIs across patients. Although limited by small sample size, single-center imaging, and conservative augmentation, the strong EPI–NON-EPI separation and alignment with gliotic histopathology support the robustness of these findings.

Conclusions

We hypothese that in MRI-negative focal epilepsy, the epileptogenic cortex is not structurally normal but exhibits a consistent microstructural phenotype that becomes detectable through quantitative FLAIR-T2 radiomics and machine-learning analysis. Using a paired, SEEG-guided

design, our XGBoost classifier distinguished epileptogenic from contralateral zones with high accuracy (AUC ≈ 0.95; 90% per-patient discrimination) and revealed a reproducible “epi-fingerprint” characterized by high-frequency intensity skewness, fine-scale textural disruption, and abnormal radial core–rim organization. These patterns align with subtle gliotic and microarchitectural alterations that remain invisible on conventional MRI. Importantly, feature-level explainability using SHAP demonstrated that the model’s decisions were biologically interpretable rather than "black-box" predictions. Together, these findings challenge the notion of true MRI-negativity and support the integration of radiomic post-processing and explainable machine learning into the presurgical evaluation of patients with drug-resistant focal epilepsy.

P24

Benchmarking EEG Functional Connectivity and Detecting Parkinson's Disease with Fully Hyperbolic Neural Networks

Y. Vakilna[1, 2], M. Carrasco-Gómez[1], N. Lotfi[1], S. Lhatoo[1], J. Mosher[1]

[1]Uthealth Houston - Houston (United States), [2]Rice University - Houston (United States)

Introduction

Parkinson's disease (PD) is usually diagnosed after substantial dopaminergic neuron loss, when available treatments primarily relieve symptoms rather than modify disease progression. There is a growing need for network level biomarkers that can capture early cortical reorganization and support screening and longitudinal monitoring. High density EEG (hdEEG) offers an accessible way to probe large scale brain networks, but functional connectivity estimates vary widely across methods, which limits reproducibility and translation. Hyperbolic neural networks provide a principled way to represent hierarchical network structure that is difficult to capture in Euclidean space. Here we benchmark two widely used connectivity measures, phase locking value (PLV) and amplitude envelope correlation (AEC), on a PD hdEEG dataset using a fully hyperbolic neural network (FHNN) link prediction task, and characterize group level connectivity differences between PD with dementia and healthy controls.

Methods

We analyzed resting state hdEEG from the publicly available University of Iowa PD cohort. Recordings were acquired with a 64 channel cap at 500 Hz, and source reconstructed on the ICBM152 template using a three shell boundary element model and sLORETA minimum norm estimation. Cortical activity was parcellated using the Schaefer 100 region atlas with Yeo 7 network labels. For each parcel, we extracted alpha band (8–12 Hz) source time series with a finite impulse response filter, motivated by prior evidence that alpha oscillations are altered in PD with cognitive impairment. Parcel wise connectivity matrices were computed for PLV and AEC on 4 s epochs with 50 percent overlap, then sparsified by proportional thresholding to retain the top 5 percent of edges before binarization. For each subject and connectivity estimator, we built a graph with one hot node features and trained an FHNN encoder operating in Lorentzian hyperbolic space. The network consisted of two fully hyperbolic graph convolution layers with Lorentz aggregation, optimized with Riemannian Adam on a link prediction task that discriminates observed edges from sampled non edges using a Fermi–Dirac decoder and a margin ranking loss. Training, validation, and test sets were formed with a fixed random seed and a stratified subject level split to preserve the balance of PD and control participants.

Results

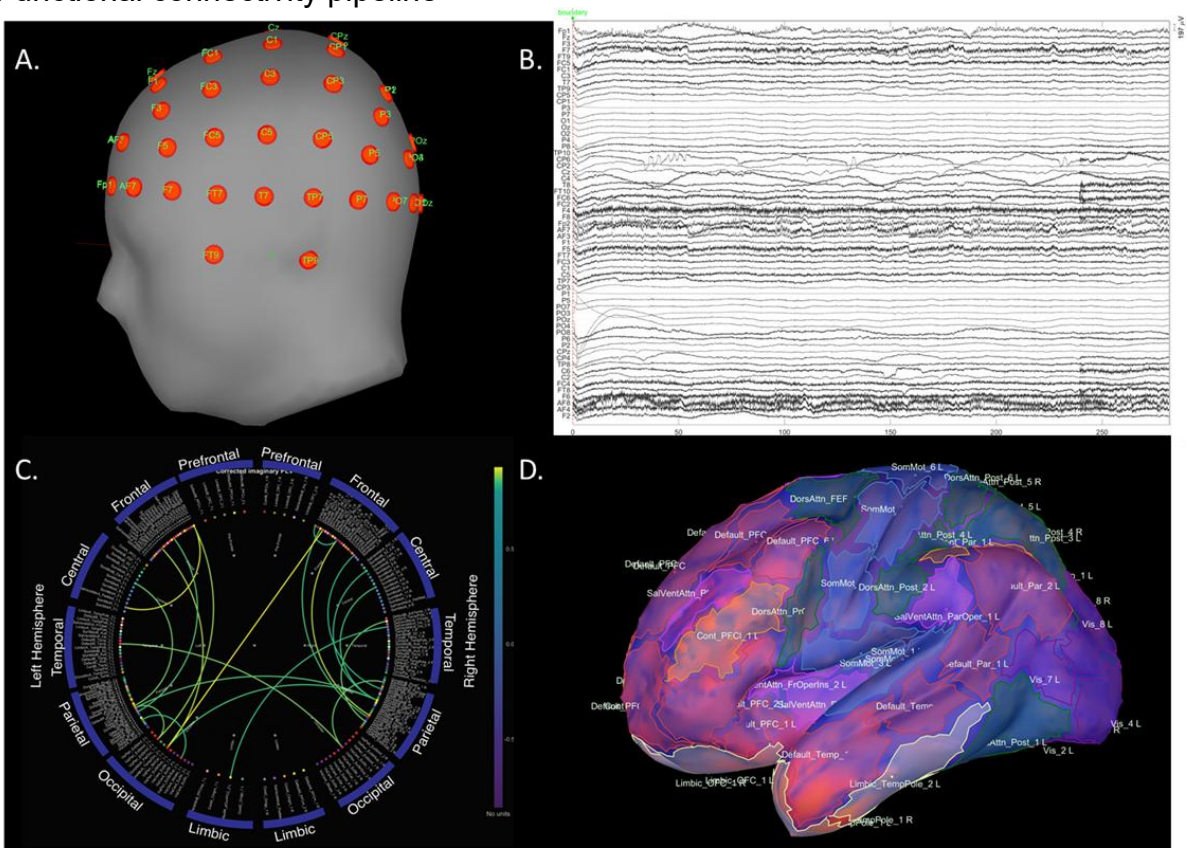
The present analysis focused on 19 PD patients with dementia (Montreal Cognitive Assessment score less than 22) and 14 demographically matched healthy controls with artifact free resting state recordings. Source reconstruction and inverse modeling produced physiologically plausible

cortical maps, with eye blink artifacts localizing to frontal regions and no obvious forward model artifacts. Group level comparison of connectivity revealed that PLV and AEC capture complementary aspects of network change. Thresholded PLV difference graphs showed widespread alterations spanning multiple large scale networks, whereas AEC emphasized more focal differences concentrated in occipital and parietal regions. During FHNN training, test accuracy, precision, and area under the receiver operating characteristic curve increased monotonically and the loss decreased without evidence of divergence, indicating stable optimization for both connectivity pipelines. Across the full training horizon, PLV based graphs consistently achieved higher accuracy, precision, and area under the curve, together with lower loss, than AEC based graphs. This pattern demonstrates that PLV edges align more closely with the hyperbolic geometry learned by the FHNN and preserve more information about the latent functional architecture under a fixed model configuration.

Conclusions

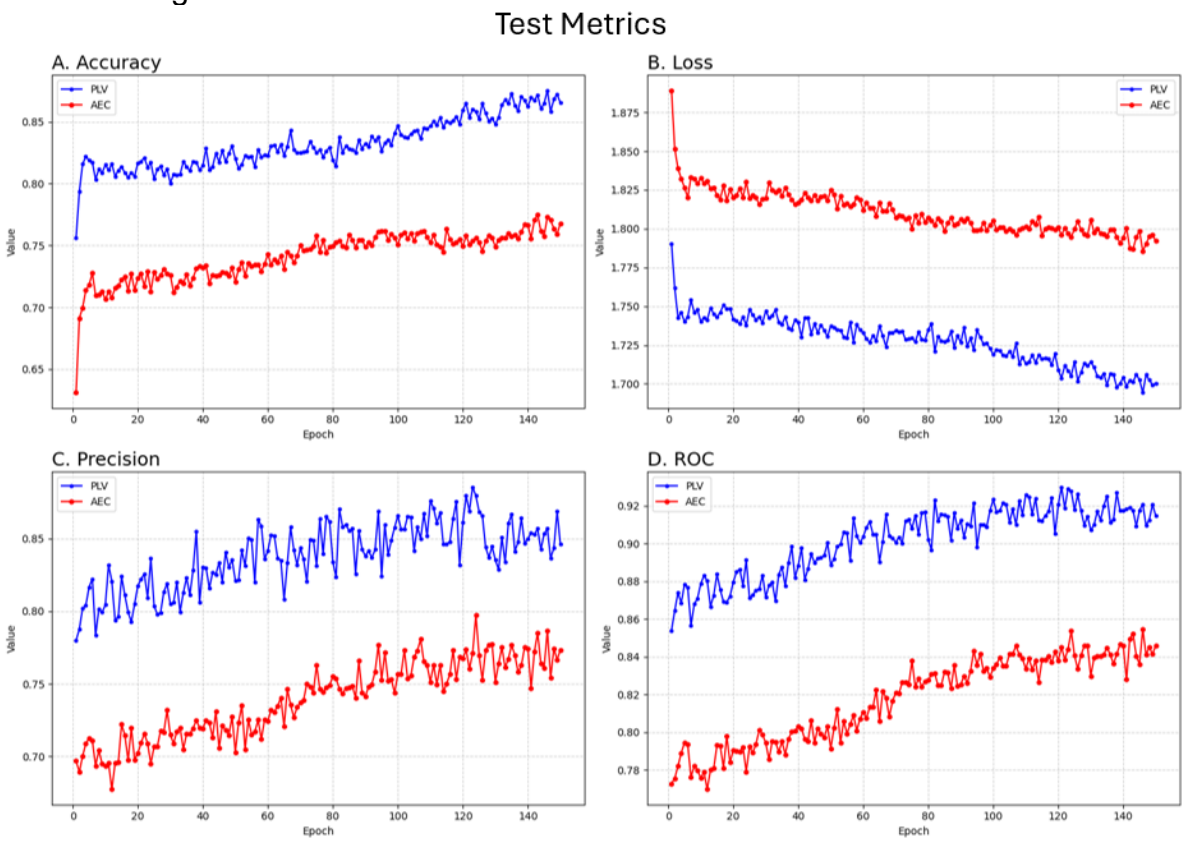
This work shows that fully hyperbolic neural networks can be used as a benchmarking tool to compare EEG functional connectivity pipelines in a clinically relevant PD cohort. PLV based connectivity is better preserved by FHNN link prediction than AEC under identical modeling and training conditions, suggesting that phase synchrony carries richer information about the underlying hierarchical network structure in alpha band resting state hdEEG for PD with dementia. At the same time, the distinct spatial patterns observed for PLV and AEC highlight that these estimators are not interchangeable and may capture complementary network abnormalities. Ongoing work will extend these hyperbolic embeddings to subject level classification of PD versus controls and to the prediction of cognitive status, with the goal of developing robust, interpretable network biomarkers that can support early detection and monitoring of neurodegenerative disease.

Functional connectivity pipeline



https://secure.key4events.com/onsite/api.aspx?e=1960&img=Picture1.png&ai=74106&op=getabstractimg&dirN=0

FHNN testing metrics



https://secure.key4events.com/onsite/api.aspx?e=1960&img=test_metrics.png&ai=74106&op=getabstractimg&dirN=1

P25

Machine learning analysis of motion signatures of febrile seizures and fatal seizures in home videos of young children

M. Harikar[1], D. Nhu[1], L. Kuhlmann[1], O. Devinsky[2], L. Gould[2], S. Sivathamboo[1, 3]

[1]Monash University - Melbourne (Australia), [2]Nyu Grossman School Of Medicine And Nyu Comprehensive Epilepsy Center - New York (United States), [3]Alfred Health - Melbourne (Australia)

Introduction

The causes underlying sudden unexplained childhood death (SUDC) are poorly understood. Terminal audio-videographic home recordings of seven children with SUDC demonstrated convulsive events preceding death in most cases. We do not know if febrile seizures differ from terminal seizures preceding SUDC. Here, we compared the motion patterns between visual recordings of febrile seizures with terminal seizures in children with SUDC.The causes underlying sudden unexplained childhood death (SUDC) are poorly understood. Terminal audio-videographic home recordings of seven children with SUDC demonstrated convulsive events preceding death in most cases. We do not know if febrile seizures differ from terminal seizures preceding SUDC. Here, we compared the motion patterns between visual recordings of febrile seizures with terminal seizures in children with SUDC.

Methods

This study was performed with ethical approval from the Institutional Review Boards of New York University (i14-01061) and Monash University (437/21).

Fatal seizure videos and febrile seizure videos were obtained from the SUDC Registry and Research Collaborative (SUDCRRC). The videos were annotated by two investigators (MH and SS) for video quality and the unequivocal clinical onset and termination for each seizure. Only recordings containing the entire seizure duration were included. We excluded incomplete videos, those without associated motor features, or those including subjects other than the child leaving or entering the frame during the recording of motor activity.

Video Pre-processing

The videos were edited to capture equivocal convulsive segments and cropped to ensure the child was upright and centred in each frame. These videos were loaded to Python 3.1 (Wilmington, Delaware, USA), converted to greyscale, rescaled to a unform spatial resolution of 64x64 pixels and resampled to 12 frames per second. Each video was then divided into non-overlapping 10-second clips.

Motion Energy Calculation

We calculated mean and standard deviation of motion energy, defined as the mean absolute pixel difference between consecutive normalized frames. We created equal temporal segments per seizure to calculate motion-wise features per segment, which corresponded to Quartile 1 (0–25th percentile, seizure onset), Quartile 2 (25–50th percentile, early seizure evolution), Quartile 3 (50–75th percentile, late seizure evolution), Quartile 4 (75–100th percentile, seizure cessation). We defined motion bursts characterised by motion energy exceeding the median threshold (50th

percentile) of motion in the seizure segment. Burst fraction was the fraction of video frames in burst state in a segment.

Statistical analysis

We calculated the mean and SD of motion energy, number of bursts, mean length of bursts and maximum length of consecutive bursts, and compared them among seizure segments and between fatal and febrile seizure groups.

Classification

We trained a histogram-based gradient-boosted classifier to differentiate between febrile and fatal seizure groups. Input features included seizure quartile; mean, maximum and SD of motion energy; number, maximum length and minimum length of bursts; and burst fraction. For each video, motion energy was normalised by subtracting the mean and dividing by the SD. Model performance was evaluated using leave-two-subject-out cross-validation, and reported as accuracy, area under the curve (AUC), sensitivity, specificity, and F1 scores. Feature importance was assessed using permutation importance scores.

Results

Twelve febrile seizures and 12 fatal seizures were initially reviewed. Seven febrile seizures and 8 fatal seizures that met the criteria were included (all from individual patients except for two videos in the febrile seizure dataset). All videos were taken by a wall-mounted or fixed crib camera, consisting of children in their crib/cot, and capturing seizures arising out of sleep.

Motion energy and burst fraction

Motion energy was largely consistent across febrile seizure segments but showed a clear decreasing trend as the seizure progressed in the fatal seizure group (Figure 1A). On segment-wise comparison of motion metrics between the two groups, there was a statistically significant difference in the burst fraction between fatal (0.68) and febrile seizures (0.49, p=0.02) in the first segment (Figure 1B). We found no statistically significant differences in mean motion energy, mean number of bursts, or mean and maximum lengths of consecutive burst fractions in remaining seizure segments.

GBM classifier

The feature matrix used for the GBM classifier was of shape (N, 8), where N was total number of 10-second seizure segments (N=135). The leave-two-subject out validation AUC was 0.75 ± 0.28, and accuracy 0.75 ± 0.17 (Table 1). The most informative feature was the maximum value of motion energy, with a permutation importance score of 0.31.

Table 1. Performance of a gradient boosted classifier algorithm in differentiating between fatal and febrile seizure groups based on motion energy features

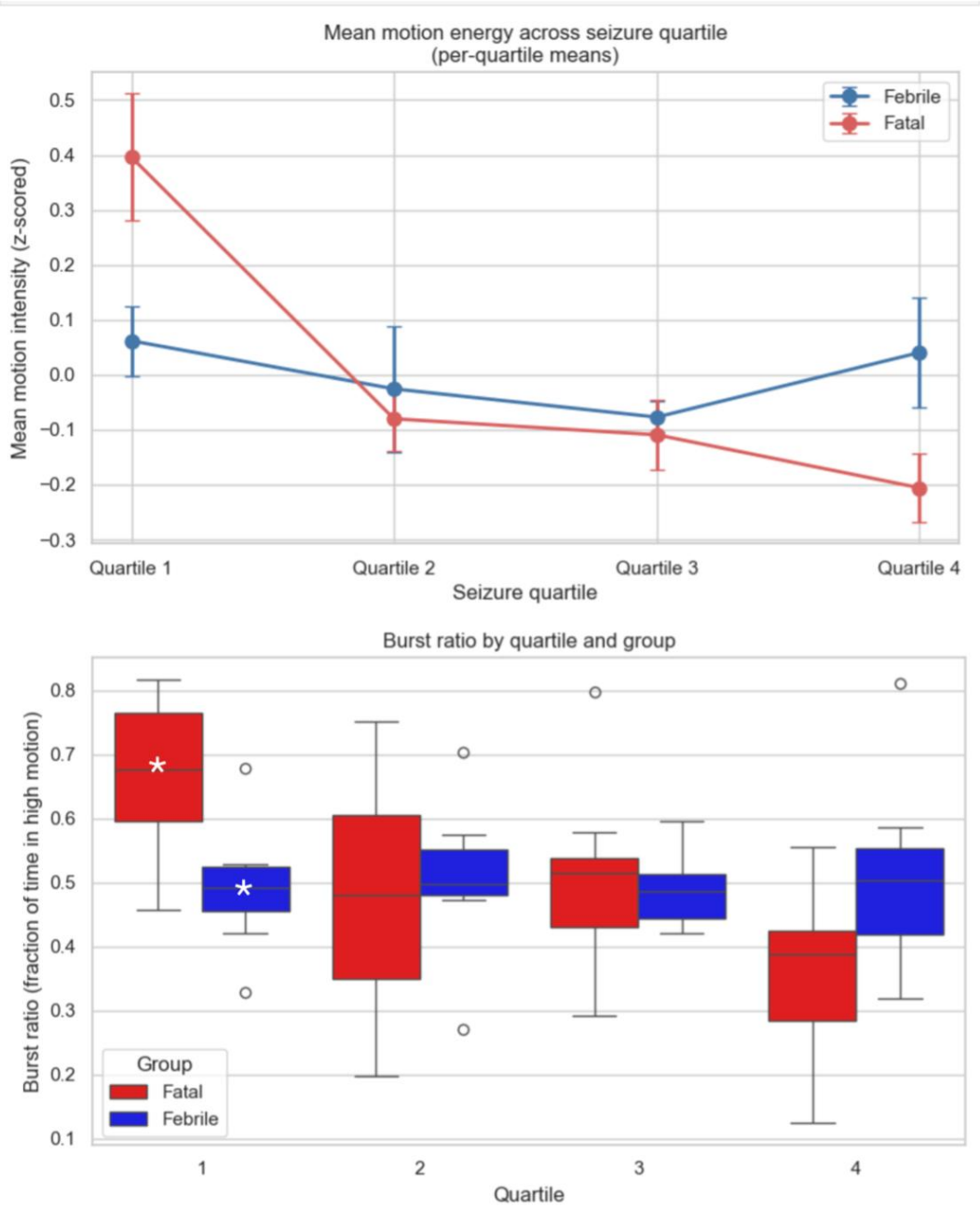
Performance metric (mean±SD)	Accuracy	AUC	Sensitivity	Specificity	F1 score
	0.75 ± 0.17	0.75 ± 0.28	0.65 ± 0.27	0.81 ± 0.23	0.64 ± 0.27

AUC, area under the curve; SD, standard deviation

Conclusions

We identified that fatal seizures had a higher proportion of high-intensity movements concentrated in the first segment of the seizure, which decreased to a level below that of febrile seizures in the last segment. We successfully implemented a gradient-boosted classifier to discriminate between fatal and febrile seizure groups despite the small size of the dataset. These findings, when validated on a larger dataset, may detect and alert at-risk seizures in the early-seizure period, and be integrated into existing technology to alert caregivers for real-time interventions.

Figure 1. (A) Change in mean motion energy and (B)



https://secure.key4events.com/onsite/api.aspx?e=1960&img=Figure+for+submission.png&ai=74111&op=getabstractimg&dirN=0

P26

XGBoost Model to Detect Cognitive Problem in Children With Epilepsy Using Network Parameter in Routine EEG

S.S. Kim[1], J.H. Moon[2], K.Y. Jung[3]

[1]Department of Pediatrics, Soonchunhyang University - Cheonan (Korea, Republic of), [2]Department Of Pediatrics, Hanyang University - Seoul (Korea, Republic of), [3]Department Of Neurology, Seoul National University College Of Medicine - Seoul (Korea, Republic of)

Introduction

Recently, there has been growing interest in attention problems and cognitive difficulties in children with epilepsy (CWE), along with an increasing demand for their screening and management. Cognitive problems in CWE are suspected to be related to the brain network, which is also related to the pathophysiology of epilepsy. However, brain network research is still mainly conducted on patients with intractable epilepsy using invasive electroencephalography (EEG) and functional magnetic resonance imaging (fMRI) for preoperative lesion resection, and there were few researches using routine EEG. Accordingly, the authors conducted this study to create an XGBoost model that detects intellectual disability using network parameters obtained from routine EEG.

Methods

The authors retrospectively reviewed the medical records of patients who were diagnosed epilepsy at one university hospital. Children diagnosed with epilepsy who completed the intelligence test (K-WISC-V) were included in the study. The average montage EEG recorded in the International 10-20 system was analyzed, and the subjects were divided into a group with cognitive decline or borderline (CD, FSIQ < 80) and a group with normal cognitive function (NC, FSIQ >= 80) according to the K-WISC-V. The Epochs on sleep 2 were used for brain connectivity analysis. The analysis was conducted according to the following steps. 1) Imaginary coherence (ImCoh) between each channel is calculated for each frequency band (delta [0.5-4.0 Hz], theta [4.0-8.0 Hz], alpha [8.0-12.0 Hz], sigma [12.0-16.0 Hz], beta [16.0-30.0 Hz], gamma [30.0-50.0 Hz]); 2) Using the ImCoh, graph theory analysis was performed and small worldness (omega) was calculated; 3) Using the omega value of each frequency band, we created an XGBoost model with FSIQ as the target feature (0: >= 80, 1: < 80). Subjects were randomly assigned to the training set and the test (hold-out) set at a ratio of 7:3. 4) The performance of the final model was evaluated using the test data set. Python 3.8.0 and its packages (MNE-python, networkx, xgboost) were used for the analysis.

Results

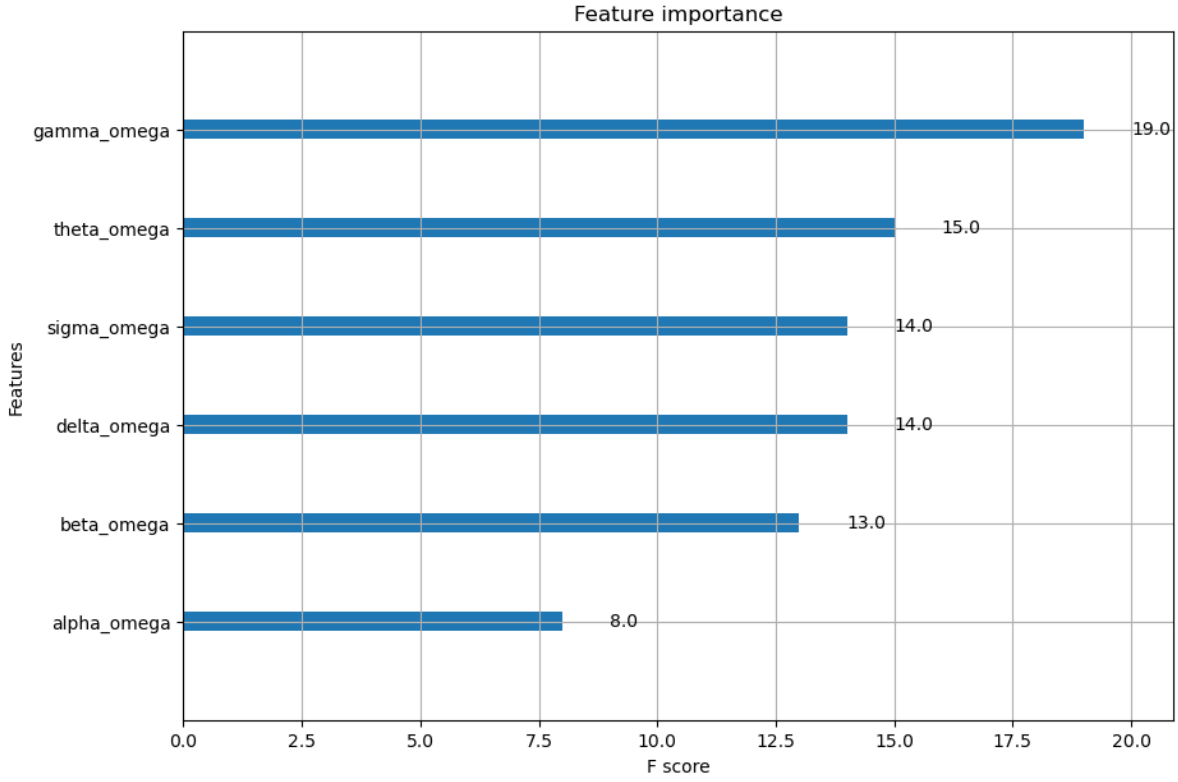
Finally 20 patients were included. There were 5 subjects (mean age 9.6 ± 2.9 year-old, 5 boys) in the CD group, and 15 subjects (mean age 7.2 ± 2.6 year-old, 7 boys) in the NC group. There was no statistically significant difference in small worldness in all frequency bands between the two

groups. The final model showed a good performance (accuracy 0.83, F1-score 0.66). The order of feature importance in model was gamma, theta, sigma, delta, beta, and alpha.

Conclusions

Despite the lack of subjects and severe target feature imbalance, this study demonstrated the feasibility of predicting cognitive problem in CWE using network parameter of routine EEG. Currently, the authors are conducting prospective data collection and developing a predictive model using multimodal data such as polysomnography (PSG), fMRI, and Magnetoencephalography (MEG).

Feature Importance Plot of XGBoost Model



https://secure.key4events.com/onsite/api.aspx?e=1960&img=fx_importance.png&ai=74113&op=getabstractimg&dirN=0

P27
Automatic detection of myogenic artifacts from iEEG using deep learning

V. Travnicek[1, 2], D. Mansilla[3, 4], K. Pijackova[5], F. Dubeau[3], P. Nejedly[5], P. Klimes[5], J. Cimbalnik[6], B. Frauscher[7]

[1]Nstitute Of Scientific Instruments, Czech Academy Of Sciences, Brno, Czech Republic - Brno (Czechia), [2]Nternational Clinical Research Center, St Anne's University Hospital, Brno, Czech Republic - Brno (Czechia), [3]Montreal Neurological Institute And Hospital, Mcgill University, Montreal, Quebec, Canada - Montreal (Canada), [4]Neurophysiology Unit, Institute Of Neurosurgery Dr. Asenjo, Santiago, Chile - Santiago (Chile), [5]Institute Of Scientific Instruments, Czech Academy Of Sciences, Brno, Czech Republic - Brno (Czechia), [6]International Clinical Research Center, St Anne's University Hospital, Brno, Czech Republic - Brno (Czechia), [7]Analytical Neurophysiology Lab, Department Of Neurology, Duke University Medical Center, Durham, Nc, Usa - Durham (United States)

Introduction

Myogenic artifacts are a major source of high-frequency contamination in intracranial EEG (iEEG). Because their spectral content overlaps with the frequency bands of high-frequency oscillations (HFOs), they can produce substantial numbers of false-positive HFO detections and compromise the reliability of HFO-based epileptogenic zone localization (Ren et al., 2019), (Gliske et al., 2020). These artifacts are highly variable in both time and frequency, with broad spectral distributions and strong inter-subject differences. Traditional deterministic rejection strategies - such as amplitude thresholding, wavelet filtering, and empirical mode decomposition - struggle to capture this variability, as they rely on fixed assumptions and quasi-stationarity that rarely hold in real clinical recordings (Muthukumaraswamy, 2013), (Chowdhury, 2021). As a result, they often generalize poorly across patients, motivating the development of a data-driven approach capable of learning the complex spectro-temporal signatures of myogenic bursts.

Methods

We developed a deep learning framework (Nejedly et al., 2020; Pijackova et al., 2023) for automated detection of myogenic artifacts in iEEG, leveraging a large, clinician-annotated dataset. The dataset included approximately 18,000 single-channel myogenic artifact annotations, manually labeled by a certified epileptologist across 19 patients from the Montreal Neurological Institute, McGill University, Canada. More than 150 hours of iEEG sampled at 2 kHz has been annotated (Travnicek et al., 2025). Each expert annotation was converted into a 2-second segment, yielding approximately 46,000 myogenic segments. To create a balanced training dataset, for every artifact segment we selected a corresponding artifact-free 2-second segment from verified clean intervals, resulting in a total of ~92,000 labeled segments.

We computed magnitude spectrograms using a short-time Fourier transform (STFT). A Hann window of 64 samples (32 ms at 2000 Hz) was used to maximize temporal precision for brief myogenic bursts. The window was advanced with 50% overlap to provide smooth temporal coverage. A 256-point FFT was applied to each frame, allowing adequate depiction of spectral

energy in the 50–300 Hz band. The resulting time–frequency matrices were truncated to the first 200 frequency bins (0–~390 Hz), covering the spectral range where myogenic contamination is most prominent. These spectrograms served as input to a convolutional–recurrent neural network (CRNN). The model first applies a 2D convolutional layer with 256 filters and a (kern × 3) kernel, followed by batch normalization and ReLU activation, enabling extraction of local spectro-temporal features. The output is reshaped and passed into a unidirectional GRU layer with 128 hidden units, designed to capture temporal dependencies across the 2-second window. After dropout ($p = 0.5$), a final fully connected layer produces a binary classification output (myogenic vs. normal).

Model training employed subject-balanced k-fold cross-validation, ensuring that training and test sets contained disjoint sets of patients. This approach mitigates patient-specific overfitting and provides an unbiased estimate of real-world generalization. Performance was evaluated using F1-score, Cohen's κ, AUROC, and AUPRC (true positives are defined as correctly identified myogenic artifacts).

Results

Across the three cross-validation folds, the model demonstrated consistently high performance on unseen subjects. Aggregate metrics across all folds showed an average F1-score of 0.9517 ± 0.0045, Cohen's κ of 0.9034 ± 0.0089, AUROC of 0.9851 ± 0.0034, and AUPRC of 0.9835 ± 0.0034, highlighting strong agreement with expert annotations and excellent discriminative ability.

Confusion-matrix analysis revealed uniformly low false-positive and false-negative rates, confirming robust performance for both classes. Importantly, the classifier maintained high performance even on subjects with artifact morphologies not represented in the training folds. This indicates that the model successfully learned generalizable features of myogenic activity, rather than overfitting to patient-specific patterns. These results substantially outperform deterministic artifact-rejection pipelines applied to the same dataset, which displayed markedly lower sensitivity to short transient bursts and produced substantially more false-positive classifications.

Conclusions

We present a high-performance deep learning method for detecting myogenic artifacts in intracranial EEG. Leveraging a large, expert-annotated dataset and a CRNN architecture tailored to spectro-temporal dynamics, the model achieved excellent performance with $F1 \approx 0.95$, $\kappa \approx 0.90$, and AUROC/AUPRC near 0.99 across subject-disjoint folds. These results substantially exceed the performance of traditional deterministic approaches, which are limited by strong assumptions about signal stationarity and morphology.

A key limitation of the present work is that all training data were sourced from a single clinical institution, which may restrict generalizability to recordings obtained under different hardware, clinical workflows, or patient populations. Addressing this limitation will be the focus of future work, in which we will incorporate multi-center datasets and broader patient diversity. This will enable the development of a universal, institution-agnostic solution for robust myogenic artifact detection in iEEG.

References:

Chowdhury, J. (2021) Analysis And Removal of Artifacts in Electroencephalographic Recordings Using Microstate Analysis and Randomization Statistics.

Gliske, S.V. et al. (2020) ‘Distinguishing false and true positive detections of high frequency oscillations’, Journal of neural engineering, 17(5), p. 056005.

Muthukumaraswamy, S.D. (2013) ‘High-frequency brain activity and muscle artifacts in MEG/EEG: a review and recommendations’, Frontiers in human neuroscience, 7, p. 138.

Nejedly, P. et al. (2020) ‘Multicenter intracranial EEG dataset for classification of graphoelements and artifactual signals’, Scientific data, 7(1), p. 179.

Pijackova, K. et al. (2023) ‘Genetic algorithm designed for optimization of neural network architectures for intracranial EEG recordings analysis’, Journal of neural engineering, 20(3). Available at: https://doi.org/10.1088/1741-2552/acdc54.

Ren, S. et al. (2019) ‘Redaction of false high frequency oscillations due to muscle artifact improves specificity to epileptic tissue’, Clinical neurophysiology : official journal of the International Federation of Clinical Neurophysiology, 130(6), pp. 976–985.

Travnicek, V. et al. (2025) ‘The state of vigilance and not antiseizure medication dosage drive variability in interictal epilepsy biomarkers’ Clinical neurophysiology : official journal of the International Federation of Clinical Neurophysiology, Available at: https://doi.org/10.1016/j.clinph.2025.2110825.

P28
Thalamocortical Connectivity During Epileptic Seizures: An SEEG Study

A. Geller[1], K. Tyner[1], A. Razmara[1], Z. Roccaforte[1], J. Thompson[1], D. Kramer[1], S. Ojemann[1], D. Satzer[2]

[1]University Of Colorado - Aurora (United States), [2]University Of Chicago - Chicago (United States)

Introduction

Focal epileptic seizures commonly spread into thalamocortical circuits. Disrupting this spread is an established modality for treating focal epilepsy, and is being explored as a treatment for generalized epilepsy. 3 thalamic structures targeted for epilepsy neuromodulation are the Anterior nucleus of the thalamus (ANT), Centromedian nucleus (CM), and Pulvinar. Because focal seizures do not in general start in the thalamus, it is rarely sampled with stereo-EEG (SEEG), which is primarily used to delineate the ictal onset zone (IOZ).

At our center, thalamic SEEG has been used for patients whose noninvasive evaluation suggests they may benefit from neuromodulation.

Methods

8 patients with medically-refractory focal epilepsy underwent stereo-EEG monitoring for IOZ localization, with one or more electrodes targeting the thalamus. We localized each patient's thalamic electrodes by coregistering their post- implantation CT to pre-imlantation MRI. We analyzed seizures to extract time-frequency characteristics for measures of functional connectivity between a single thalamic channel and all other recording channels. For each patient, we identified the dominant seizure onset pattern, and analyzed ex- amplar seizures to obtain time frequency spectrograms (TFSs) for imaginary coherence (IC) and Granger causality (GC). We aligned the connectivity TFS at seizure onset, and for each implanted contact, we obtained the z-scored mean TFS over each patient's dominant seizure pattern. We grouped contacts into IOZ and non-IOZ groups.

Results

For IC, IOZ-nonIOZ dissociation is seen for 1 of 4 ANT patients (VG), 2 of 2 CM patients, and 0 of 2 Pulvinar patients

For GC, IOZ-nonIOZ dissociation is seen for 1 of 4 ANT patients (DR), 2 of 2 CM patients, and 2 of 2 Pulvinar patients

For GC, forward-backward dissociation is seen for all 8 patients.

Conclusions

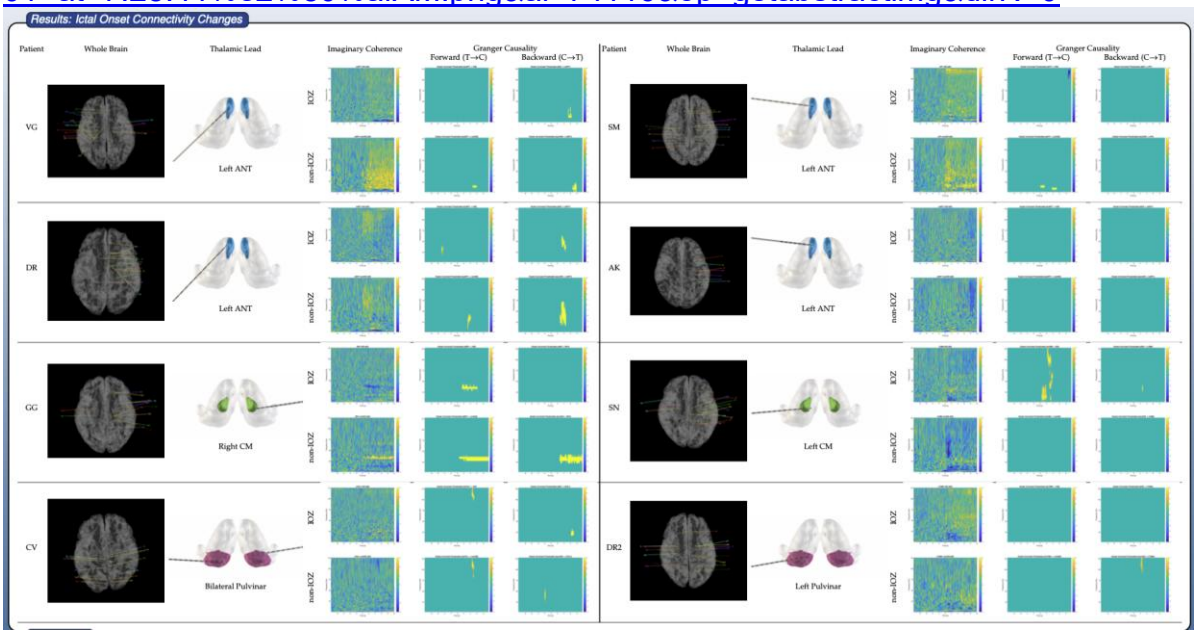
Prominent thalamocortical connectivity changes may be present at seizure onset even when thalamic raw EEG signal demonstrates no seizure spread.

Ictal onset connectivity changes may be narrow- or broadband.

In this limited sample, ictal onset connectivity changes vary across thalamic nuclei, raising the question of whether there might be consistent anatomic variation for these measures.

Patient	# seizures analyzed	Thalamic raw EEG spread	Ictal connectivity changes
VG	3	all	yes
SM	8	all	yes
DR	13	none	yes
AK	2	all	?
GG	16	none	yes
SN	12	none	yes
CV	10	3/10	yes
DR2	5	2/5	yes

https://secure.key4events.com/onsite/api.aspx?e=1960&img=Screenshot+2025-12-01+at+7.28.41%e2%80%afAM.png&ai=74116&op=getabstractimg&dirN=0



https://secure.key4events.com/onsite/api.aspx?e=1960&img=Screenshot+2025-12-01+at+7.28.32%e2%80%afAM.png&ai=74116&op=getabstractimg&dirN=1

P29
Interpretable Seizure Detection via a Hierarchical Vision-Based Pipeline

A. Kogan[1], A. Zamansky[1], I. Shimshoni[1], F. Benninger[2, 3], I. Goldberg[2, 3]

[1]Department Of Information Systems, University Of Haifa - Haifa (Israel), [2]Department Of Neurology, Beilinson Rabin Medical Center - Petah Tikva (Israel), [3]School Of Medicine, Tel Aviv University - Tel Aviv (Israel)

Introduction

Video-Electroencephalogram(VEEG) monitoring is essential for epilepsy diagnosis, yet clinicians must manually review long recordings to identify seizures and differentiate them from normal movements or psychogenic non-epileptic seizures(PNES). While EEG provides the primary diagnostic signal, the accompanying video often contains crucial information, especially in ambiguous cases. Although automated seizure-detection tools have improved, many challenges remain and most available algorithms are end-to-end deep learning systems that function as “black boxes”, limiting clinical trust and hindering integration into clinical settings. Clinicians need methods that not only detect seizure events but also explain why a segment was flagged and how its movement patterns correspond to recognizable seizure signatures.

To address this need, we present an interpretable hierarchical computer-vision framework for segment-level seizure detection in long VEEG recordings. Deep learning is used solely for feature extraction and initial frame-level screening, while final decisions are made by a transparent Random Forest classifier(RF) operating on temporally aggregated physical motion features. This separation yields measurable, clinically aligned descriptors of movement and supports trustworthy, explainable event detection.

Methods

Our dataset, collected from the VEEG unit at Rabin Medical Center Beilinson Hospital includes 11 adult patients with 16 clinically diagnosed seizure events (15 epileptic, 1 PNES) and matched non-seizure segments. For this preliminary analysis, we evaluated the framework on 5 seizure videos and 5 patient-matched normal videos.

Our pipeline consists of four major stages: feature extraction, deep-learning classification, segment construction, and interpretability-driven segment-level refinement. Each stage is designed to produce clinically comprehensible information.

1. Feature Extraction Across Three Visual Modalities

To capture the richness of motor behavior during seizures, the system analyzes three complementary visual fields: (a)The full-body bounding box centered on the patient, capturing gross motor activity(bbox), (b)A crop of the facial region, capturing facial micro-expressions and head movement(face), (c)A joint feature-set combining both modalities(bbox_face).

For each modality, frame-level feature vectors combine several complementary signals:

Visual motion descriptors: embeddings extracted via the CLIP and ResNet50 networks, reduced with t-distributed Stochastic Neighbor Embedding(t-SNE) capture appearance changes over time, yielding low-dimensional, meaningful trajectories for visualizing patient movement.

Geometric body features derived from the bounding boxes, including width, height, center position, and their temporal derivatives.

Movement-complexity measures: applying Singular Value Decomposition(SVD) to short (48-frame) windows of geometric signals produces a mathematical indicator of motion organization. The SVD component ratio is especially informative, showing stark differences between voluntary movements and rhythmic or chaotic seizure activity.

Accumulated motion distance: Euclidean distances calculated between consecutive embeddings over a 216-frame sliding window, to quantify the intensity of movement.

2. Frame-Level Screening Using Deep Learning

A lightweight Multi-Layer Perceptron(MLP) trained on all extracted features classifies each frame as seizure or non-seizure. To prevent patient leakage and promote generalization, the model is trained using Leave-One-Subject-Out (LOSO) cross-validation across seven patients, ensuring each test subject is fully unseen during training. The MLP outputs a continuous probability stream, which is informative but inherently noisy and sensitive to fluctuations, especially in recordings with frequent voluntary movements or subtle events, necessitating additional temporal and segment-level processing.

3. Segment Construction via Hysteresis Thresholding

To stabilize the frame-level probabilities, we apply hysteresis thresholding with two decision levels: a high threshold that initiates a seizure segment once confidence rises sufficiently, and a lower threshold that keeps the segment active until the evidence clearly drops off. The intent is to mirror how neurologists interpret seizures, allowing brief fluctuations without prematurely considering an event has ended. The output is a set of temporally coherent candidate segments that move forward to the final classification stage.

4. Interpretable Segment-Level Random Forest Filtering

Finally, each hysteresis-derived segment is then converted into an interpretable feature summary and evaluated by a Random Forest(RF) classifier. For every candidate event, we aggregate all geometric and motion features, including those used by the neural network, using their minimum, mean, maximum, and standard deviation values. This allows the RF to assess motion patterns over a longer temporal window and distinguish true seizures from normal behavior on a per-segment basis. A final post-processing step merges nearby events and discards segments that are too short to be clinically meaningful, producing realistic and clinically interpretable seizure detections.

Results

Across the five seizure events, the system showed strong, stable performance over a wide range of thresholds. For mid-range RF decision thresholds (0.3-0.6) and IOU criteria(0.1-0.5), both bbox and bbox_face models achieved perfect detection(TP=5, FN=0, FP=0), yielding Precision = Recall = F1 = 1.0. This broad plateau indicates high robustness without reliance on precise tuning. Outside this region, performance declined predictably: lower thresholds introduced 1-2 false positives while maintaining F1 scores(0.83-0.91), whereas strict RF or IOU thresholds(≥0.7) reduced sensitivity to TP=4 (F1 ≈ 0.89). The most extreme settings suppressed detections entirely, serving as upper-bound controls rather than clinical operating points.

Overall, the hierarchical pipeline achieved high sensitivity with minimal false alarms across its threshold range. Segment-level feature aggregation effectively filtered out normal movements that superficially resembled seizures, maintaining stability across parameters. In addition to strong quantitative performance, the system produces fully interpretable 3-RF modality plots(bbox, face, bbox_face) and a weighted ensemble graph, aligning confidence trajectories, RF decisions, and ground truth on a common timeline. These multimodal visualizations provide transparent explanations to justify the classification of each segment, supporting trustworthy and clinically actionable review.

Conclusions

This research shows that a hierarchical, interpretable computer-vision framework can achieve clinically reliable seizure detection even with a small patient cohort. By separating feature extraction, frame-level screening, and segment-level decision-making, the system produces transparent outputs that clinicians can easily review, an essential requirement for real-world VEEG workflows. The Random Forest preserved high sensitivity while effectively suppressing false positives, and this robustness strengthens clinical viability while reducing reliance on hyperparameters. Future work will expand the patient cohort, incorporate additional tracked body parts, and retrain larger networks to improve RF predictions across operating conditions. This will enable the system to link decisions to specific body-part dynamics, further enhancing interpretability. Overall, these results suggest that combining deep learning with interpretable temporal features offers a practical, clinically trustworthy path toward deployable AI-assisted seizure detection.

P30

Automated Noninvasive Mapping of Epileptogenic Regions in Children with Drug-Resistant Epilepsy Using Magnetoencephalography and Dictionary Learning

H. Partamian[1, 2], S. Jahromi[1, 2], M.S. Perry[1], E. Tamilia[3], J.R.M. Madsen[3], J. Bolton[3], S.S.D. Stone[3], P.L. Pearl[3], C. Papadelis[1, 2]

[1]Cook Children's Healthcare System - Fort Worth (United States), [2]The University Of Texas At Arlington - Arlington (United States), [3]Boston Children's Hospital - Boston (United States)

Introduction

Magnetoencephalography (MEG) plays a key role in the presurgical evaluation of patients with drug-resistant epilepsy (DRE). When appropriately applied and accurately interpreted, MEG can expand the number of candidates eligible for further surgical evaluation, enhance surgical planning, and potentially improve the likelihood of achieving postoperative seizure freedom. Yet, the clinical utility of MEG as a presurgical evaluation tool is limited by the time-consuming, labor-intensive, and prone to bias and error, manual data review and process. Here, we propose a fully automated framework that characterizes epileptogenic brain regions noninvasively by applying dictionary learning methods to MEG data recorded from patients with DRE. We hypothesize that dominant epileptogenic brain sources extracted through dictionary learning will localize epileptogenic sources accurately and more efficiently than traditional spike-based source localization methods.

Methods

We retrospectively analyzed MEG data recorded from 20 good-outcome (Engel I) children and young adults with DRE who underwent neurosurgery (follow-up 2 years). Using a dictionary learning approach, we transformed the MEG data into temporal maps and brain networks. We first computed the continuous wavelet transform (CWT) to define the spectro-temporal characteristics of each channel (**Fig. 1a**). Next, we averaged the spectrogram power in seven physiologically relevant bands: *δ* (1–4 Hz) , *θ (*4–8 Hz), *α* (8-12 Hz), *β (*12–30 Hz), *γ (*30–80 Hz), *sb (*1–80 Hz), and *rb* (80-250 Hz). We then applied non-negative matrix factorization across the CWT of all the channels in all bands to identify sparse brain networks and the corresponding activation times in the temporal map (**Fig. 1b**). We then identified the sensor-level epileptogenic network (Sen-EN) as the one having frequent interictal epileptiform discharges (**Fig. 1b**). We estimated the concordance of active data segments of Sen-EN in the temporal map with the clinically annotated interictal MEG events. Using dynamic Statistical Parametric Mapping (dSPM), we localized the sources at each active Sen-EN time point and then averaged the source maps across the time points to define a source-level epileptogenic network (Sor-EN) (**Fig. 1b**). Using the resection as the gold standard for defining the EZ, we assessed the ability of Sor-EN to delineate the EZ by estimating three properties: (i) focality, measuring the proximity of the identified network electrodes to each other; (ii) overlap with resection; and (iii) distance from resection (**Fig. 1b**).

Results

At the sensor level, the proposed framework identified the Sen-EN, whose activity in the temporal map was consistent with the clinically annotated spikes in the MEG recordings. Specifically, the concordance between the active time points of the Sen-EN and the annotations was 74%, 76%, and 73% in the δ, θ, and α bands, respectively. At the source level, the Sor-EN showed a focality of 9 ± 2.7 mm, an average distance of 5.6 ± 0.7 mm from the surgically resected region, and an overlap of 72% with the resection.

Conclusions

Our proposed framework identifies a Sen-EN that is concordant with the clinically annotated interictal spikes, automatically. Moreover, the framework identifies a Sor-EN that delineates the EZ. The dictionary learning framework offers a powerful and fully data-driven approach for noninvasive delineation of epileptogenic brain regions from interictal MEG data. Such an automated process could enhance presurgical planning and improve surgical outcomes in patients with DRE.

P31

Video-based Detection of Generalized and Focal to Bilateral Tonic-Clonic Seizures via Automated Patient-Focused Motion Identification

D.A. Galindo Lazo[1, 2], I. Sarzo Wabi[1, 2], A. Jahani[3, 2], R. Saber[3, 4], M. Robert[2], J.P. Millan Sandoval[5], A. Robin[5], G. D'onofrio[5], D.K. Nguyen[5, 3], E. Bou Assi[2, 6]

[1]Institute Of Biomedical Engineering And The Department Of Pharmacology And Physiology, University Of Montreal - Montreal (Canada), [2]Research Center Of The University Of Montreal Hospital Center - Montreal (Canada), [3]Department Of Neuroscience, Université De Montréal, Montreal - Montreal (Canada), [4]Research Center Of The University Of Montreal Hospital Center - Montreal (Canada) - Montreal (Canada), [5]Neurology Service, University Of Montreal Hospital Center - Montreal (Canada), [6]Department Of Neuroscience, Université De Montréal, Montreal - Montrea (Canada)

Introduction

One-third of people with epilepsy remain drug-resistant, experiencing unpredictable seizures that can lead to injury and, in some rare cases, sudden unexpected death. Automated video monitoring offers a non-invasive approach to seizure detection, particularly in settings where EEG is not feasible or cannot be continuously deployed. Existing video-based methods are often trained on short, curated clips recorded under controlled conditions (e.g., fixed camera angles that keep patients centered and clearly visible in bed). Models developed in these settings may overestimate real-world performance. In contrast, continuous patient monitoring frequently involves variable viewpoints, partial occlusions, and inconsistent lighting, and methods that depend on joint or pose estimation often degrade substantially under these conditions. To better reflect these constraints, we trained a deep learning model on randomly sampled one-hour video recordings from our epilepsy monitoring unit (**EMU**), where camera viewpoints vary and patients can move freely within the room. The model detects generalized and focal to bilateral tonic-clonic seizures (**GTCS/FBTCS**) by first localizing the patient and then extracting motion cues within the identified patient region.

Methods

Database: We analyzed recordings of patients staying at the University of Montreal Hospital Center (**CHUM**) EMU between 2019 and 2024. Data were collected across four EMU rooms equipped with ceiling-mounted cameras capturing video (1080x720-pixel) at 30 frames per second (**fps**). The cameras provide a 340° pan range, a 105° tilt range, and infrared capability to support night-time monitoring.

For the patient identification module, we used 1,590 labeled frames from 131 patients. For the seizure detection algorithm, the dataset contained 907 randomly selected recordings (mean duration 56.3 minutes; range: 5.3–61 minutes), totaling 872 hours of footage from 44 patients with 74 GTCS/FBTCS. All seizure events were annotated by board-certified epilepsy fellows using synchronized video-EEG, with seizure periods defined from clinical onset to electrical offset.

Patient identification algorithm: We fine-tuned a You Only Look Once (**YOLOv8**) object-detection model to distinguish patients from visitors or staff and to automatically localize the patient within bounding boxes. A temporal tracking post-processing was developed utilizing Kalman filters and a moving average of the patient's last known coordinates (C) to manage two scenarios: 1) If the detection failed, the predicted position from the Kalman filter was temporarily retained during video sequences; and 2) to prevent misidentification, in cases of an abrupt positional jump (where the current detection deviated significantly from C), the system retained the predicted location.

Seizure detection algorithm: Motion was quantified from patient subframes at 15 fps using a Recurrent All-Pairs Field Transforms (**RAFT**) model, which estimates a dense 2D vector field representing pixel-wise displacement between consecutive frames. Using the calculated patient-only motion streams, we trained a lightweight residual R(2+1)D convolutional neural network with factorized 3D convolutions. The continuous motion stream was fed to the model using a 3-second sliding-window with 50% overlap. A firing-power postprocessing method was used, defined as a moving average of the raw model output over a 9-second window. A detection was triggered when this smoothed output exceeded a predefined threshold (θ=0.6).

Performance evaluation: Seizure detection performance was assessed using a patient-wise 3-fold cross-validation scheme where the data was divided for training/validation (~57/28%), and held-out testing (15%), ensuring the testing set did not contain any patient used for the initial YOLOv8 detection training. The patient-localization module was assessed using precision and recall, while seizure detection performance was quantified with event-based metrics, including sensitivity, precision, false-alarm rate (**FAR**) per 24 hours, and detection latency.

Results

Patient identification algorithm: Performance was evaluated on a held-out test subset of 244 frames (20 patients), yielding 97.6% recall and 96.8% precision. The detector remained robust under challenging conditions, reliably distinguishing the patient even when covered by bedsheets, bedtables, or occluded during staff intervention, both during daytime and nighttime recordings.

Seizure detection algorithm: Performance metrics were computed on a held-out test set comprising 123 hours of video from 6 patients. The footage spanned a realistic distribution of motion contexts, including scenes with external motion artifacts within the patient subframe (e.g., staff checkups, patient care, and visitor interaction) as well as periods with minimal disturbance such as sleep or routine daily activities.

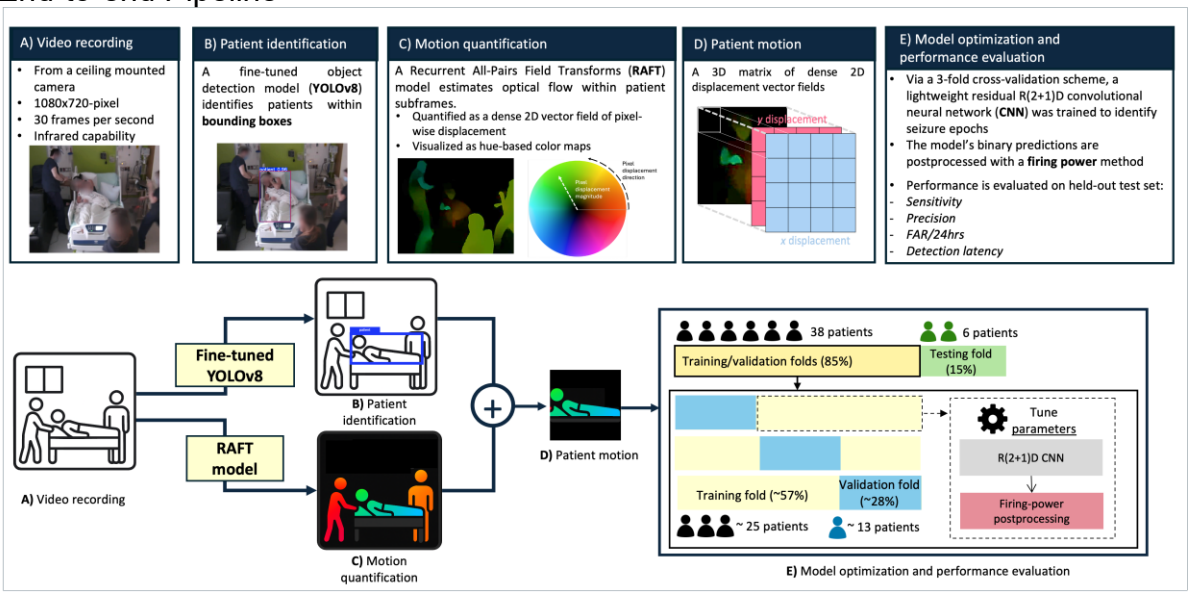
The model achieved 100% sensitivity, correctly detecting 9/9 seizures. One seizure where the patient was completely alone in the room and medical staff intervened late (92 seconds after seizure offset) was correctly detected, suggesting that patient motion alone, free from intervention motion artifacts, is sufficient for detection. Mean detection latency relative to clinical onset was 67.8 seconds [range: 10s–195s]. When referenced to the moment of seizure generalization, latency decreased to 20.0 seconds [range: 7s–52s].

Precision reached 60%, corresponding to a FAR/24hrs of 1.17. Importantly, no false alarms were raised when external personnel entered the patient's subframe. The total six false alarms were highly concentrated, with 5/6 occurring in a single patient. They originated primarily from some sudden, sharp patient movements, such as kicking during sleep, making the bed and throwing the bedsheet, or abrupt hand gestures while talking, indicating a minor weakness in differentiating certain high-velocity patient movements from seizure activity.

Conclusions

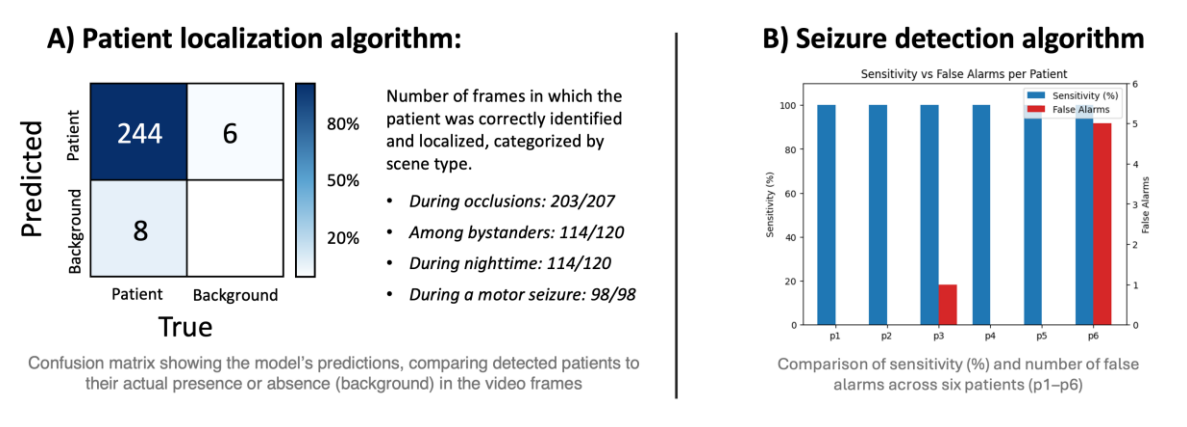
We present a preliminary video-based motion analysis pipeline that achieves promising performance for the detection of GTCS/FBTCS from uninterrupted one-hour video recordings. By effectively isolating patient movements from the rest of the frame, the system filters external motion artifacts produced by staff and visitors. This approach yielded 100% sensitivity with a FAR/24hrs of 1.17 and a latency of 20 seconds. Our results show that the system can detect seizures based on the patient's intrinsic movement pattern, while demonstrating preliminary robustness against false alarms caused by external activity (e.g., staff doing routine checkups or late interventions). Although the current analyzed recordings are not multi-day continuous streams, they retain the operational complexity of long bedroom monitoring. Future work will focus on integrating the full, continuous patient's stay at the EMU recordings to include richer non-seizure motion patterns necessary to further eliminate the observed causes of false alarms. We will also implement nested cross-validation strategies to ensure robust generalizability.

End-to-end Pipeline



https://secure.key4events.com/onsite/api.aspx?e=1960&img=Methodology.png&ai=74124&op=getabstractimg&dirN=0

Performance of algorithms



https://secure.key4events.com/onsite/api.aspx?e=1960&img=Results.png&ai=74124&op=getabstractimg&dirN=1

P32

An Artificial Intelligence Framework for Surgical Outcome Prediction in Drug Resistant Epilepsy through the Combination of Interictal Biomarkers

S. Jahromi[1], H. Partamian[1], M.S. Perry[1], E. Tamilia[2], J. Bolton[2], S.S.D. Stone[2], P.L. Pearl[2], C. Papadelis[1]

[1]Cook Children's Healthcare System - Fort Worth (United States), [2]Harvard Medical School - Boston (United States)

Introduction

Precise delineation of the epileptogenic zone (EZ) remains a critical challenge for achieving seizure freedom in patients with drug-resistant epilepsy (DRE). While the seizure onset zone (SOZ) derived from ictal intracranial EEG (iEEG) provides a conventional approximation of the EZ, interictal abnormalities, such as spikes and ripples, offer abundant, non-seizure-dependent markers of the epileptogenic tissue. Although spatiotemporal and spectral features of spikes and ripples have been linked to the EZ, they may lack sufficient specificity when considered individually; in contrast, their combination may provide complementary information regarding the underlying epileptic network. We hypothesize that integrating these multimodal interictal features within a machine-learning framework will enable a more precise delineation of the EZ, and that the resection of the zone predicted by the combined features will be associated with favorable surgical outcomes.

Methods

We retrospectively analyzed iEEG recordings from 62 children with drug-resistant epilepsy undergoing neurosurgery, classified as having good (n = 34, Engel I) or poor (n = 28, Engel ≥ II) surgical outcomes. We coregistered pre-operative MRIs of the patients with the electrode implantation CT scans to map the locations of iEEG electrodes and to define the resected brain regions. The resection volume of good outcome patients was used as the gold standard to define the EZ. For each patient, an artifact free five-minute interictal iEEG segment was selected. The iEEG segments were notch-filtered to remove power-line noise and its harmonics. The iEEG segments were then bandpass-filtered into spike (1–80 Hz) and ripple (80–250 Hz) frequency bands. From these filtered signals, we detected spikes and ripples automatically and computed the features (Figure 1a). For each channel, we computed the features by extracting spike and ripple rates (SR and RR, respectively), number of co-occurrences of spikes and ripples (SRC), spike and ripple propagation indices (SPI and RPI, respectively; Figure 1b-d), and average spike and ripple powers (SP and RP, respectively; Figure 1e). In addition, we computed spatial features separately for spikes and for ripples, based on the Euclidean distance of each channel from (i) the channels with the highest spike rate (ED-HSR) and the highest ripple rate (ED-HRR); (ii) the channels with the highest average spike power (ED-HSP) and the highest average ripple power (ED-HRP); and (iii) the spike and ripple propagation onset channels (ED-SO and ED-RO, respectively; Figure 1f).

We evaluated the ability of individual features to predict the epileptogenic zone by using Youden's method to determine optimal thresholds that distinguished epileptogenic from non-epileptogenic

electrodes. Additionally, using three feature sets based on spikes (SF), ripples (RF), and the combined features (CF) as predictors, and resection as target, we trained three Random Forest classifiers to define the predicted epileptogenic zone (PEZ) using the leave-one-patient-out cross-validation (LOPO-CV) approach. We compared the performance of the EZ prediction classifiers with that of individual features by computing their area under the receiver operating characteristic curves (AUC-ROC) for predicting the resection of good outcome patients. We also compared the performance of the Random Forest classifiers in good vs. poor outcome patients by comparing the overlap of their PEZ with resection (O_{RES}), the distance of PEZ from the resection (D_{RES}), and the focality of PEZ (F_{PEZ}) as the average distance between electrodes of the PEZ. We applied two-sided non-parametric *Wilcoxon signed-rank* test for all paired comparisons and *Wilcoxon rank-sum* test for non-paired comparisons between good- and poor-outcome patients. We also assessed whether resecting the PEZ provided by the individual features and the Random Forest classifiers could predict surgical outcome, using *Fisher's exact test*, and compared their outcome predictability using the *McNemar's* test.

Results

Individual interictal features, including spike and ripple rate, spectral power, and propagation metrics, were elevated inside vs. outside resection in good-outcome patients, suggesting their potential to characterize the EZ ($P<0.05$, Wilcoxon signed-rank test). The PEZ defined by the spike rate and spatial features had an overlap >60% with resection in good outcome patients. Resecting the PEZ delineated by most of the features predicted surgical outcome. We found that SF and CF classifiers achieved higher area under the receiver operating characteristic curve (AUC-ROC: 0.9 and 0.89, respectively) compared to most individual features, except for ED-SO (0.9, $P< 0.05$ in all cases) (Figure 2a). We found that SF, RF, and CF had a higher O_{RES} and lower D_{RES} in good- compared to poor-outcome patients ($P<0.01$). No differences in F_{PEZ} were observed between poor- and good-outcome patients for any of the feature sets. Notably, SF had a median $O_{RES} \approx 74$ (56-95) %, RF had a median $O_{RES} \approx 75$ (50-85) %, and CF had a median $O_{RES} \approx 74$ (57-94) % (Figure 2b). Finally, PEZ predictions were strongly associated with postsurgical outcome, with the CF model achieving the highest sensitivity (85%), specificity (71%), and overall accuracy (79%) (Figure 2c). The comparison of outcome predictability across individual features and classifiers showed that resection of the PEZ predicted by the CF classifier performed as well as or better than individual features and the other classifiers in predicting surgical outcome (Figure 2d).

Conclusions

Our findings highlight the potential clinical utility of combining spatiotemporal and spectral interictal features for training machine learning classifiers for the analysis of patients with DRE. The resulting models provide accurate predictions of both the true EZ and surgical outcomes in most patients with DRE.

P33

Cross-device generalization of EEG foundation models: A critical evaluation for ambulatory seizure monitoring

G. Lo Bianco[1], C. Ambroise[1], F. Baudouin[1], O. Gloaguen[1], L. Ribière[1], S. Chambon[1], M. Pendelio-Boyer[1]

[1]Reliev Technologies - Nantes (France)

Introduction

Long-term seizure monitoring requires practical alternatives to the cumbersome full montage EEG used in hospital settings. Minimal EEG systems, such as the Byteflies Sensor Dot with 2 EEG derivations [1], offer a promising ambulatory solution, but the severe reduction in spatial dimensionality fundamentally complicates automatic seizure detection. Foundation models (FMs) have garnered significant interest in the EEG community by offering large pre-trained representations capable of accelerating development and improving generalization across diverse downstream tasks [2].

Yet, a critical gap remains: the generalization capabilities of existing FMs are rarely assessed against data from novel, highly reduced spatial montages and completely different hardware platforms. FMs are typically validated only on data similar to their pre-training source and their pre-training hardware.

This work addresses this gap by benchmarking a fine-tuned downstream predictor built on EEGPT [3], pre-trained on EEG signal recorded with traditional 10-20 montage (18 bipolar derivations). We evaluate its performance on seizure detection using data collected from the minimal, two-channel Byteflies SensorDot, assessing the model's ability to transfer knowledge across this extreme hardware and spatial shift.

Methods

1. Training Paradigms and Model Initialization

The core of this study is a spatial & hardware generalization challenge. We benchmark the ability of a downstream seizure predictor, the "seizure detection Head", derived from a large FM.

The SZ Head is derived from a 4.7M parameter EEGPT-base1 encoder [3]. This encoder was obtained following a self-supervised pre-training on a proprietary 100,000-hour dataset, collected using a 10-20 EEG montage involving 18 bipolar derivations.

We define three distinct learning setups:

- From-Scratch (FS): The model is initialized with random weights, establishing the performance baseline without pre-trained knowledge.

- Partial Fine-Tuning (PFT): The model encoder is initialized with pre-trained weights but is frozen during training, allowing only the downstream predictor (Head) weights to be updated. This assesses the quality of the raw pre-trained representation, essentially using the encoder as a fixed feature extractor.

- Full Fine-Tuning (FFT): The model encoder is initialized with pre-trained weights and is trainable, allowing both the encoder and the downstream predictor (Head) weights to be updated.

2. Downstream Predictor Architectures

We compare two types of seizure detection heads attached to the EEGPT encoder: the MLP Head and the Transformer Head. The MLP Head uses mean temporal pooling across the tokens returned by the encoder before feeding the aggregated vector into a Multi-Layer Perceptron. The Transformer Head uses a dedicated Classification Token ([CLS] token) appended to the sequence of tokens returned by the encoder; the prediction is derived solely from this token's final representation.

All models were trained and evaluated on the Byteflies SeizIT2 dataset [4]. This clinical dataset includes 125 patients (~ 12,000 hours of recording, 886 annotated seizures, focal and general), but importantly, utilizes only two bipolar EEG channels among T7-P7, T7-T8, T8-P8, a major spatial reduction from the pre-training set. To ensure patient-independent generalization and a fair comparison, we performed a 5-fold cross-validation (CV). The 125 patients were partitioned in five folds of 25 patients each, then regrouped into training (3 folds, 75 patients), validation (1 fold, 25 patients) and test (1 fold, 25 patients) sets. At each iteration, the model was trained on the 3 folds and assessed on the left-out test fold, ensuring robust cross-subject generalization over the entire dataset.

Performances were assessed using the following event based metrics: Area under the Precision Recall curve (AUC-PR) and Recall as a function of the number of False Alarm Rate per hour (#FP / h) [5]. AUC-PR is given with 95% confidence intervals (CI) estimated with bootstrap with replacement. For direct comparison against the current state-of-the-art, we include the performance of a specialized deep learning model, the "Specialized baseline." This baseline is a high-performing CNN-Transformer architecture trained entirely from scratch and highly tuned to work on the target data.

Results

The experiments demonstrate that equipping Foundation Model with an MLP head using Full Fine-Tuning (FFT) is the most effective strategy and exhibits comparable performances as the specialized baseline. In contrast, using a Transformer head leads to poorer performance across all learning setups. Overall, Full-Fine-Tuning yields the best AUC-PR, followed by Partial Fine-Tuning finally followed by From-Scratch training. The performance gap is tight for false alarms per hour, but the specialized baseline slightly outperforms the foundation model.

Conclusions

This work validated the Foundation Model's ability to overcome significant hardware and data distribution shifts. We demonstrated that the model's pre-trained representations could be fine-tuned to achieve high-performing seizure detection on EEG data recorded using entirely different hardware (Byteflies vs. standard) and a distinct electrode montage.

Consistent with recent literature [6], the foundation model does not yet significantly outperform the specialized approach. This is primarily attributed to the inherent difficulty of cross-device generalization and the nascent stage of the EEG foundation model literature. New approaches propose to encode the spatial electrode positions into a learned compressed latent space, which could be a great lead for improving this challenging cross-device generalization [7,8].

Bibliography

[1] Harvey W et al. (2020). The Byteflies Sensor Dot: A Multimodal... [Abstract 238]. Proc. Am. Epilepsy Soc. Annual Meeting.

[2] Lee N et al. (2025). A Comprehensive Review of Biosignal Foundation Models. TechRxiv.

[3] Wang G et al. (2024). EEGPT: Pretrained transformer... Proc. Adv. Neural Inf. Process. Syst. (NeurIPS 37), pp. 39249–39280.

[4] Bhagubai M et al. (2025). SeizeIT2: Wearable Dataset Of Patients With Focal Epilepsy. Sci. Data, 12, 1228.

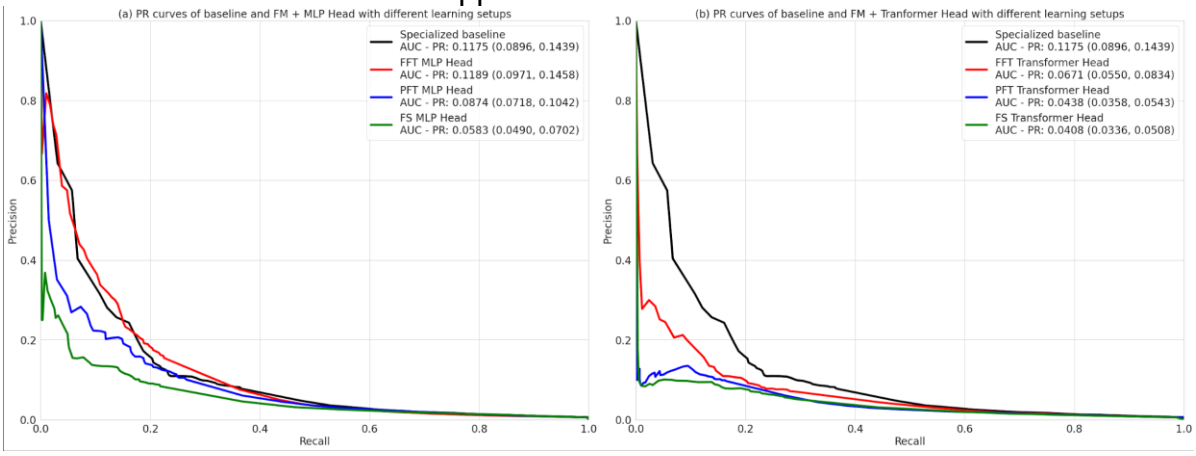
[5] Dan J et al. (2025). SzCORE as a benchmark... arXiv (arXiv:2505.18191).

[6] Lee N et al. (2025). Are large brainwave foundation models capable yet?... ICML 2025

[7] Xiao, Q., et al. (2025). BrainOmni: A Brain Foundation Model for Unified EEG and MEG Signals. NeurIPS 2025

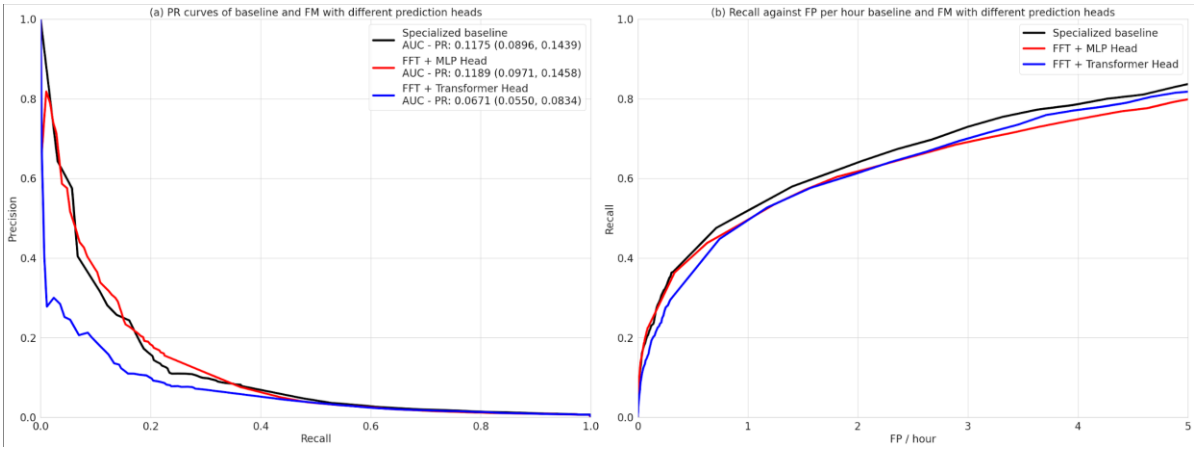
[8] Döner, B., et al. (2025). LUNA: Efficient and Topology-Agnostic Foundation Model for EEG Signal Analysis. NeurIPS 2025

PR curves of baseline and FM approaches



https://secure.key4events.com/onsite/api.aspx?e=1960&img=figure_1.png&ai=74126&op=getabstractimg&dirN=0

PR curves and Recall vs FP per hour



https://secure.key4events.com/onsite/api.aspx?e=1960&img=figure_2.png&ai=74126&op=getabstractimg&dirN=1

P34

Between Simplicity and Oversimplification: Unlocking the Clinical Value of Latent-EEG AI Analysis

L. Krumm[1], C. Meisel[2]

[1]Phd Student - Berlin (Germany), [2]Professor - Berlin (Germany)

Introduction

Clinical decisions in neurology face a fundamental challenge: traditional clinical scores are often too simple to capture the complexity of brain states, while bio-signal data such as EEG may be too complex to extract all information over extended periods of time in routine neuro- intensive care settings. Machine learning algorithms can extract valuable information from EEG, yet they typically reduce this complexity to binary predictions. For ethically consequential decisions, reliance on such black-box outputs remains limited and often defaults to simple clinical scores. We propose that the solution lies not in better classifiers, but in structured latent space representations that make EEG patterns directly interpretable.

The present work builds on our previous publication (Krumm et al., "Toward a Universal Map of EEG," Annals of Neurology, 2025), where we developed a latent space representation of established brain states. This map is a semantic, continuous space encompassing healthy states (wake and sleep stages) and pathological states (burst suppression and the ictal- interictal injury continuum (IIIC)).

Methods

We projected multi-day EEG recordings from cardiac arrest (CA, n=578) and subarachnoid hemorrhage (SAH, n=100) patients into our map of EEG. From each patient's trajectory, we extracted spatial and temporal measures quantifying movement through latent space (e.g., angle to healthy reference clusters, fraction of time near pathological patterns, state transition entropy; see Krumm et al. for details). Using Linear Discriminant Analysis (LDA), we identified the three most discriminative features for distinguishing good from poor outcomes in each cohort. We then performed binary outcome classification using a hyperplane decision boundary in this three-dimensional feature space. Beyond binary classification, we tracked each patient's trajectory over time, measuring how discriminative features evolve throughout the recording. To enable comparison across diseases, we defined a relative similarity score that normalizes the pathological-to-healthy relationship. For each EEG segment, we compute cosine similarity to both a pathological anchor (e.g. burst suppression for CA; GRDA for SAH) and a healthy anchor (e.g. wake). The relative similarity score is then: (cos_healthy + 1) / ((cos_healthy + 1) + (cos_pathological + 1)), yielding values from 0 (pathological) to 1 (healthy). This accounts for different spatial arrangements, as burst suppression lies closer in the map to wake than IIIC patterns do, enabling direct comparison between cohorts. To present our trajectories in an accessible framework, we developed an interactive visualization tool. The tool displays each patient's movement through latent space over time and quantifies the percentage of time spent close to different vigilance states. Users can select specific timepoints to inspect the

corresponding EEG segment alongside the three most similar pre- labeled reference patterns from the UM-EEG, and retrieve patients with similar trajectories for case comparison.

Results

For CA patients, the most discriminative features were angle to healthy, state transition entropy, and fraction of time in burst suppression (AUC 0.83 [95% CI: 0.74–0.90]). For SAH patients, the top features were angle to healthy, angle to REM, and fraction of time in GRDA

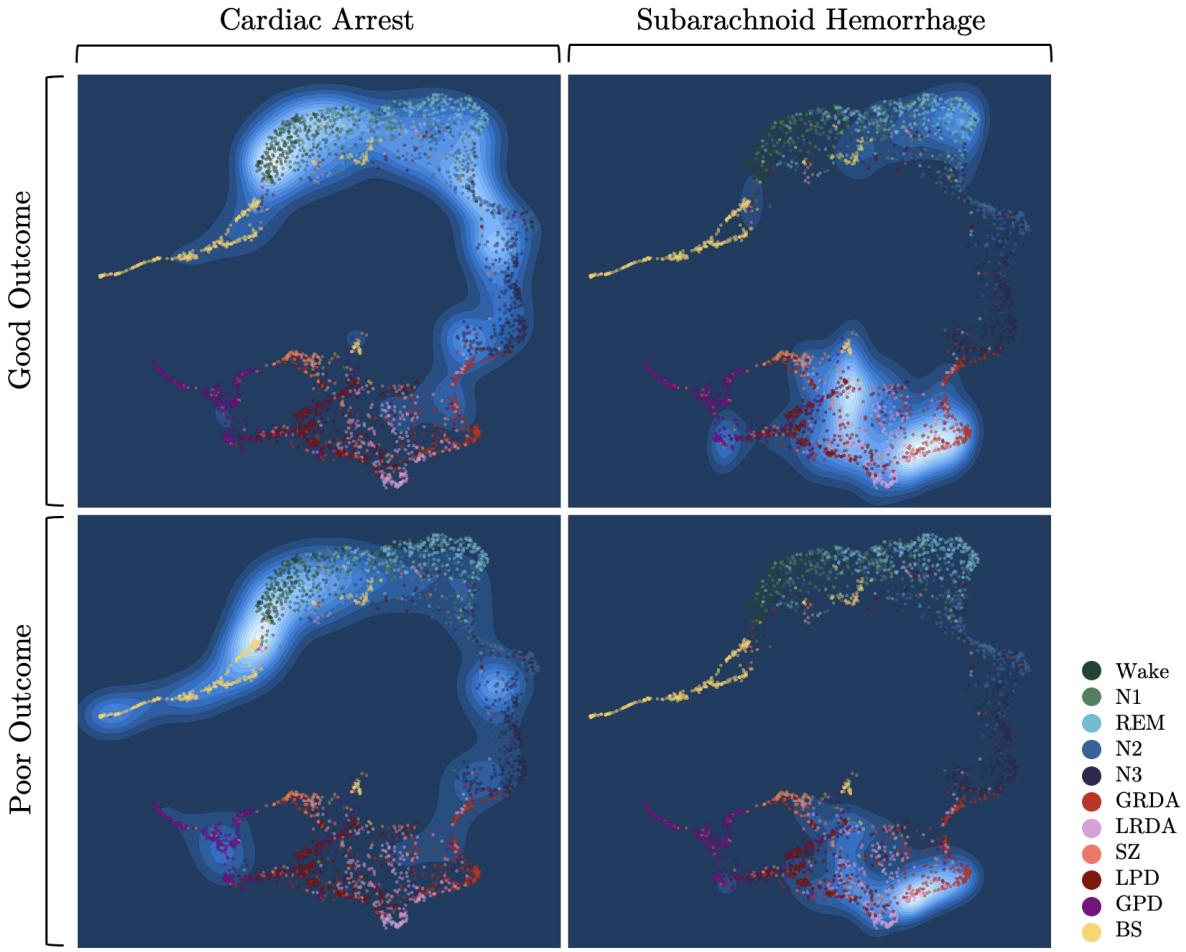
(AUC 0.78 [0.69–0.87]). While binary prediction delivered moderate results, it also offered limited clinical utility as a benchmark.
Temporal analysis revealed more nuanced separation. When tracking discriminative features over time, good and poor outcome groups showed non-overlapping 95% confidence intervals early in the recording period. Good outcome patients positioned closer to healthy reference states throughout their trajectories, while poor outcome patients remain near pathological patterns.

Both cohorts shared a common discriminative feature (angle to healthy) but exhibited distinct disease-specific trajectories reflecting their underlying pathophysiology: CA patients clustered near burst suppression, while SAH patients moved along the IIIC continuum. Despite these differences, analyses revealed a simple and shared principle: poor outcome patients remained closer to pathological regimes, while good outcome patients trended toward healthy reference clusters.

Conclusions

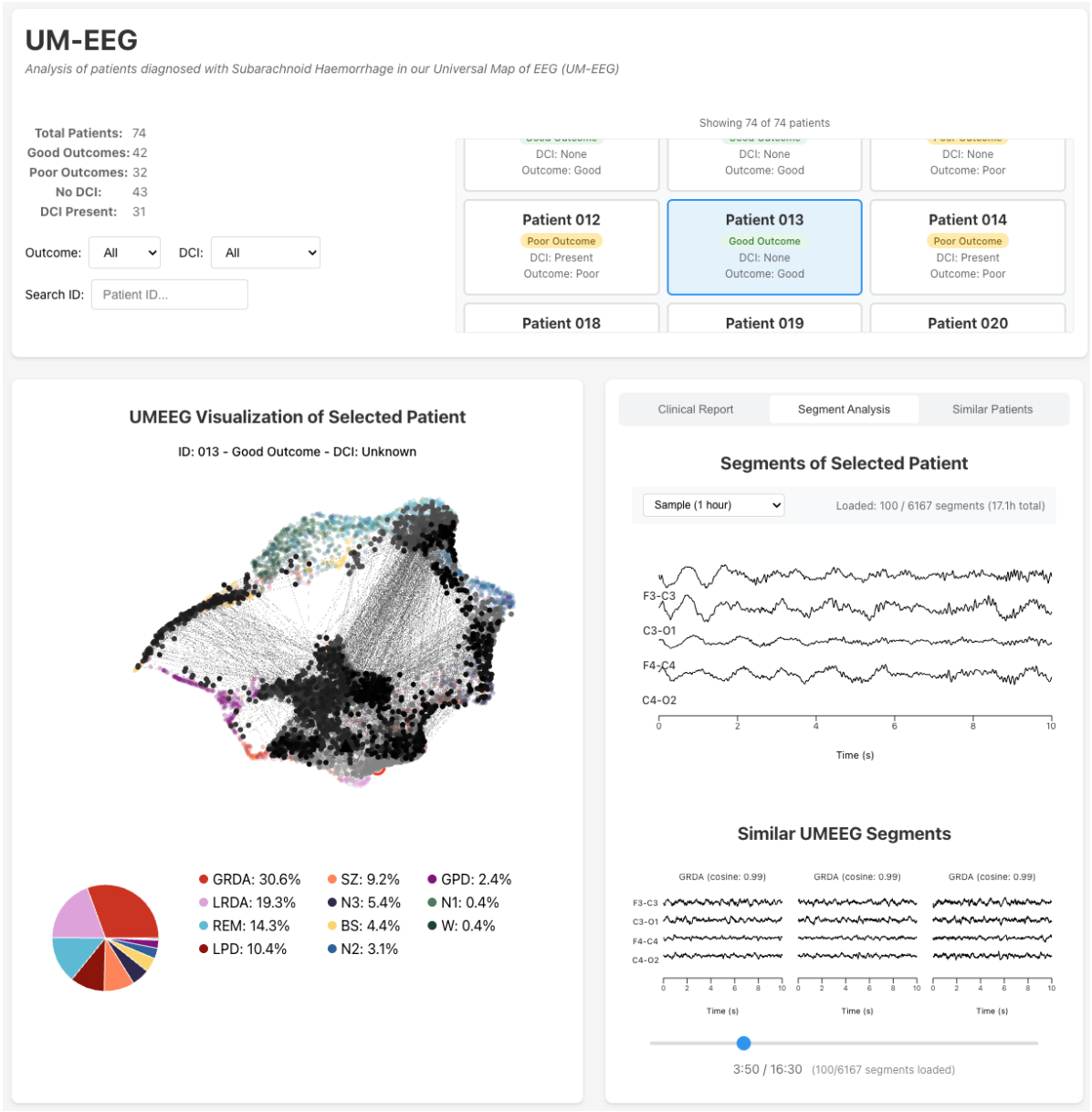
Tracking patient trajectories through structured latent space may provide richer clinical information than binary outcome prediction. By visualizing how brain states evolve relative to healthy and pathological references, clinicians can assess not just predicted outcome but how a patient is progressing and which EEG patterns characterize their recording. This bridges the gap between oversimplified clinical scores and complex bio-signal data while remaining interpretable. Our interactive visualization tool makes this information directly accessible, allowing clinicians to explore patient trajectories, inspect underlying EEG patterns, and compare them with similar cases. Consistent results across CA and SAH populations suggest this framework may generalize to other neuro-intensive care settings where continuous EEG monitoring has yet to reach its full potential.

CA and SAH movements in UMEEG



https://secure.key4events.com/onsite/api.aspx?e=1960&img=ca_sah.png&ai=74127&op=getabstractimg&dirN=0

Preliminary Interactive Tool



https://secure.key4events.com/onsite/api.aspx?e=1960&img=umeeg.png&ai=74127&op=getabstractimg&dirN=1

P35
Video-Based Seizure Detection Challenge: Automated Model Development and Classification with Codex Agents

M. Kramer[1, 2]

[1]Boston University - Boston (United States), [2]Johns Hopkins University - Baltimore (United States)

Introduction

We implemented an agent-based workflow for the video-based seizure detection challenge.

Methods

Two Codex agents were defined: (i) a Worker agent that implemented and evaluated performance; and (ii) an Advisor agent with high-level reasoning that interpreted results and recommended next experiments. Agents were provided the general instruction to use the information on the competition website to maximize F1 (tie-breaker sensitivity). Following these initial instructions, Agents operated directly on the 5-second training clips, with limited human oversight.

Results

The Agents performed over 400 training/evaluation iterations, modifying classification strategies. Performance improved from ~0.5 F1 to a final out of fold F1 of 0.782 with sensitivity 0.898.

Conclusions

We conclude that agent-based model development with limited human oversight can achieve strong performance on this challenging time-series classification task, providing exciting opportunities to address existing classification problems and highlighting the limits of manual feature-engineering alone.

P36

Evaluating Performance of a Neonatal and Infant Seizure Detection Algorithm in the Presence of Expert Annotation Variability

T. Puranik[1], S. Anand[1], A. Gupta[1], L. Tozzi[1], C. Juliano[2], M. La Vega-Talbott[3], R.R. Said[4], A. Singh[5], J.E. Sullivan[6], N.S. Abend[7], C. Hahn[8], B. Kamousi[1]

[1]Ceribell Inc. - Sunnyvale (United States), [2]Division Of Newborn Medicine, Icahn School Of Medicine At Mount Sinai - New York City (United States), [3]Department Of Neurology And Department Of Pediatrics, Icahn School Of Medicine At Mount Sinai - New York City (United States), [4]The University Of Texas Southwestern Medical Center - Dallas (United States), [5]Medical College Of Wisconsin - Milwaukee (United States), [6]University Of California San Francisco School Of Medicine - San Francisco (United States), [7]Division Of Neurology, Children's Hospital Of Philadelphia And Departments Of Neurology And Pediatrics, University Of Pennsylvania - Philadelphia (United States), [8]Division Of Neurology, The Hospital For Sick Children, And Department Of Pediatrics University Of Toronto - Toronto (United States)

Introduction

Seizures occur frequently during the first year of life (Annegers et al., 1995), and high seizure burden during this period is associated with adverse outcomes (Alharbi et al., 2023). Continuous EEG monitoring remains the gold standard for identifying neonatal seizures (Wusthoff et al., 2025), but timely review by specialists is not always feasible (Boylan et al., 2010). Additionally, the predominance of focal and short seizures in this age group (Sandoval et al., 2021) make interpretation challenging and contribute to high inter-rater variability between expert reviewers (Massey et al., 2019). Clarity for Neonates and Infants (Ceribell, Inc., Sunnyvale, CA) is a new FDA-cleared algorithm designed to detect seizures from EEG data recorded in neonates and infants. In this retrospective study, we evaluated Clarity's seizure detection performance using the annotations from pediatric electroencephalographers as a reference standard. We also compared performance for recordings that were unanimously labeled as containing seizures or no seizures by experts.

Methods

We analyzed conventional EEG studies from 715 patients from three hospitals (Kravis Children's Hospital, Children's Medical Center Dallas, and Children's Wisconsin). Patients ranged from birth to one year of age, with post-menstrual ages between 22 and 94 weeks. Overall, the dataset contained 101 preterm, 384 term, and 230 infant EEG studies. Each EEG was reviewed by two or more board-certified pediatric epileptologists who annotated seizures using established criteria (Tsuchida et al., 2013). Reviewers were unaware of the Clarity algorithm output or the annotations of other reviewers. The majority opinion of reviewers was used as the final label. Reviewer agreement was evaluated pair-wise at the recording level using Cohen's kappa and percentage agreement when a pair of reviewers had an overlap of at least 40 cases (Cohen, 1960).

We then ran the Clarity for Neonates and Infants algorithm, post-hoc, on the EEGs. For every 10 seconds of non-overlapping EEG, the algorithm detected seizures and calculated a seizure

burden, defined as the percentage of the previous 5 minutes that the algorithm identified as containing seizure activity. We evaluated algorithm performance for the full dataset, as well as for the subset of EEGs for which reviewers were unanimous. Specifically, we tested whether the algorithm correctly identified recordings that exceeded three predefined maximum seizure burden threshold: ≥10%, ≥50%, and ≥90%. For each threshold, we calculated sensitivity, specificity, and negative predictive value.

Results

The overall seizure prevalence in the dataset was 10%. Reviewers had unanimous agreement in 64% of recordings containing seizures and in 90% of recordings not containing seizures. Cohen's kappa between reviewer pairs varied from 0.103 (poor agreement) to 0.503 (moderate agreement). Percent agreement between reviewers ranged from 71% to 86%.

Among recordings with unanimous labels, Clarity showed high performance across all seizure burden thresholds with sensitivity 85–90%, specificity 81-97%, and negative predictive value 99–100%. Specifically, the sensitivity, specificity, and negative predictive value were 88%, 81%, 99% for the ≥10% threshold, 85%, 93%, 99% for the ≥50% threshold, and 90%, 97%, 100% for the ≥90% threshold.

Over the entire cohort, including EEGs for which reviewers disagreed, the algorithm maintained high performance, with sensitivity 88–94%, specificity 78–96%, and negative predictive value 99–100%. The sensitivity, specificity, and negative predictive value were 88%, 78%, 98% at the ≥10% threshold, 88%, 92%, 99% at the ≥50% threshold, and 94%, 96%, 100% at the ≥90% threshold.

Conclusions

We observed poor to moderate agreement between expert pediatric epileptologists in labeling neonatal and infant EEG recordings for seizures. The Clarity algorithm demonstrated strong performance in detecting and ruling out seizures. Additionally, the high performance was consistent across cases with varying levels of interrater agreement. These results indicate the potential of Clarity for automated continuous EEG monitoring in the first year of life.

P37

Performance of Ceribell Clarity AI Algorithm in Identifying Adult Status Epilepticus in a Large Real-World Dataset

J. Parvizi[1], A. Gupta[2], M. Armenta Salas[2], L. Tozzi[2], S. Karunakaran[2], T. Puranik[2], B. Kamousi[2]

[1]Department Of Neurology And Neurological Sciences, Stanford University School Of Medicine - Palo Alto (United States), [2]Ceribell, Inc. - Sunnyvale (United States)

Introduction

Timely identification of nonconvulsive seizures and status epilepticus (SE) and continuous monitoring of these events in acute and emergency settings remain a challenge. Point-of-care electroencephalography (POC EEG) systems coupled with artificial intelligence (AI) algorithms to monitor seizure activity offer a possible solution. The clinical utility of such systems depends on the algorithms' performance in representative EEG samples acquired in their intended settings. Here, we evaluate the eighth version of the Clarity AI algorithm (Ceribell, Inc.) for the detection of suspected SE in a large adult POC EEG dataset, when compared against expert consensus.

Methods

We ran a retrospective cohort study, randomly selecting 1340 anonymized adult EEGs, recorded with the Ceribell system. The EEGs were independently reviewed by two or more EEG experts, who annotated seizure and other abnormal patterns while blinded to other annotations and algorithm output. In accordance with clinical guidelines, cases with seizure duration ≥4.5min or with cumulative 20% seizure burden per hour were labeled as status epilepticus. For each EEG, the majority consensus (≥70%) was used as reference standard. For a subset of eight reviewers who each annotated at least 300 files, individual performance for status epilepticus detection was estimated against majority consensus. Separately, Clarity AI_{v8} was run post-hoc for all EEGs, maximum 5-min seizure burden (SzB) and hourly seizure burden were obtained per file. We used a threshold of ≥90% for the 5-min SzB and ≥20% seizure burden per hour to determine presence of status epilepticus. We estimated the algorithm's performance relative to expert consensus, estimating sensitivity, specificity, and negative predictive value to detect status epilepticus. Finally, we examined reviewer annotations for EEGs that were categorized as status epilepticus by Clarity AI_{v8}, but not by majority consensus.

Results

The majority consensus determined that 25 of 1340 EEGs met criteria for SE, 51 had seizures, 210 had highly epileptiform patterns (HEPs), 188 had other pathological abnormalities, and 866 were diffusely slow or normal. Across all files, Clarity AI_{v8} captured 24 of 25 status epilepticus cases, for a sensitivity of 96.0%, a specificity of 94.8%, and a 99.9% negative predictive value. The single missed status epilepticus case had a maximum SzB of 10%.

Reviewers had significant interrater variability in sensitivity to detect SE when compared against majority consensus, ranging from 28.6% to 100%. Specificities ranged from 83.3-99.5%. When assessed across the same subset of cases as each reviewer, Clarity AI_{v8} had sensitivities ranging from 95.2-100%, and specificities ranging from 93.0-95.0%. Relative to individual human reviewers, Clarity's sensitivity was close to the 75th percentile, while its specificity was similar to the 50th percentile.

Among the cases categorized as status epilepticus by Clarity AI_{v8}, but not determined as such by the majority consensus, seven were labeled as having seizure, 51 as HEPs, 5 as other abnormalities, and 6 as normal or slowing. In 4 of the latter 6 cases, at least one reviewer noted seizures that met criteria for status epilepticus, and all had abnormalities marked by at least one reviewer.

Conclusions

In this large real-world dataset, Clarity AI_{v8} demonstrated high performance for detecting status epilepticus at levels similar to or higher than individual expert reviewers. The high negative predictive value and low rate of false positives in cases without any pathology supports the algorithm's clinical utility in ruling out status epilepticus for neuro-emergencies and optimizing management decisions.

P38

Pediatric Clarity - Evaluating the performance of a novel AI algorithm in detecting seizures in a real-world Pediatric dataset

A. Gupta[1], T.Y. Little[1], T. Puranik[1], J. Parvizi[2], B. Kamousi[1]

[1]Ceribell, Inc. - Sunnyvale (United States), [2]Department Of Neurology And Neurological Sciences, Stanford University School Of Medicine - Palo Alto (United States)

Introduction

Critically ill pediatric patients, especially those presenting with acute encephalopathy, have a high occurrence of seizure that often lack clinical signs (Abend, et al. 2015, Herman, et al., 2015; Wilson, et al., 2015). Timely monitoring and review of electroencephalography (EEG) is essential for reliable and prompt identification of seizures. Ceribell Inc. (Sunnyvale, CA) received FDA clearance for an artificial intelligence (AI) seizure detection algorithm for patients aged 1 year and older. This study aims to quantify the performance of this algorithm in identifying seizure episodes in a real-world pediatric cohort.

Methods

We retrospectively collected 719 real-world pediatric EEGs recorded with Ceribell hardware in intensive care units, general floor, and emergency departments. The EEGs were independently annotated by 2 or more experts. EEGs were parsed into 10-second seizure or non-seizure epochs following experts' consensus, to create episode-level labels. At the EEG level, cases were labeled as status epilepticus (SE) if they had at least one seizure episode with cumulative time in seizure ≥4.5 in a 5-min window, according to expert consensus. The Pediatric Clarity algorithm was run post-hoc for all EEGs, which outputs binary seizure detections every ten seconds of recording. The seizure Burden (SzB) was calculated as a 5-minute moving average of these binary outputs. We quantified the performance of the algorithm in two different ways: 1) at the episode level, for identifying seizure episodes with true SzB ≥1% and ≥90% over any 5-minute window; and 2) and the EEG level, for identifying suspected SE cases, using a ≥90% SzB threshold.

Results

There were 194 episodes with at least 1% true SzB, and 27 long seizure episodes with 90% or greater true SzB. For detecting long seizure episodes, the AI algorithm had a sensitivity of 92.5% (95% CI 77%-100%). Algorithm-generated seizure burden exceeding 90% threshold had a false alert rate of 0.05/Hr (0.04 - 0.08). When considering all seizure episodes, including shorter seizures, the sensitivity was 84.0% (71%-92%), and the false alert rate was 0.19/Hr (0.16-0.23).

The majority consensus determined that 20 EEGs had at least one episode that met the SE criteria, and 51 had seizures. The algorithm correctly identified 19 out of 20 SE cases, for a sensitivity of 95.0% (75%-100%) with a specificity of 93.8% (92%-96%). At the EEG level, the algorithm had a negative predictive value of 99.8% (99%-100%), for ruling out status epilepticus.

Conclusions

In this real-world dataset, the Pediatric Clarity algorithm demonstrated strong performance for continuously monitoring and detecting seizure episodes across a range of seizure durations. Further, the algorithm's strong performance translated to high sensitivity and specificity at detecting and ruling out status epilepticus present at any point during the recording. The low false alert rate for long seizure episodes (0.05/Hr) and high negative predictive value to rule out status epilepticus at the file-level (99.8%), highlights the value of this algorithm as a triaging tool in pediatric neuroemergencies.

P39

Evaluating Human Pose Estimation Models in Hospitalized Patients Undergoing Video-EEG

E. Juarez-Martinez[1], J. Min[1], J. Eagles[2], S. Schuele[1], E. Kimchi[1]

[1]Northwestern University. Department Of Neurology - Chicago (United States), [2]Northwestern Medicine. Department Of Neurology - Chicago (United States)

Introduction

Monitoring activity in patients hospitalized on neurological services can be crucial for timely detection of conditions such as falls, delirium, and epilepsy, but remains labor-intensive. Continuous video-based monitoring using pose estimation models offers a scalable approach to optimize clinical assessments. Model performance in clinical populations, however, where variable postures and complex environments pose unique challenges, remains unexplored. Here, we aim to evaluate the performance of human pose estimation models in hospitalized patients undergoing video EEG monitoring and explore the foundations for integrating automated motion analysis into inpatient neurological workflows.

Methods

Three state-of-the-art pre-trained human pose estimation models, YOLOv11, OpenPose, and MMPose, were evaluated in hospitalized patients undergoing video-EEG. Each model generated coordinates and confidence scores for facial and body keypoints. From each video, ten frames were randomly selected and manually annotated (17 common keypoints) as ground truth and used to define model-specific confidence thresholds. Sensitivity, specificity, balanced accuracy, and ROC curves were computed. Keypoint location accuracy was calculated using the Euclidean distance between ground truth and model detections.

Results

We analyzed 46 videos from 24 hospitalized patients with diverse racial backgrounds. OpenPose showed the highest AUC in 52% of the cases (24/46 videos) and the highest overall balanced accuracy (BA) of 72% (53% sensitivity and 89% specificity). For specific body regions, OpenPose performed better for facial detection (BA 77%, 64% sensitivity, 90% specificity) while MMPose performed better for upper body (BA 74%, 62% sensitivity and 85% specificity) and lower body detection (BA 66%, 48% sensitivity and 82% specificity). Distance errors were smaller for facial keypoints than for torso and limbs.

Conclusions

State-of-the-art computer vision models face challenges when applied to clinical videos. Models were more accurate for facial keypoints, and least accurate for lower body keypoints. Enhancing

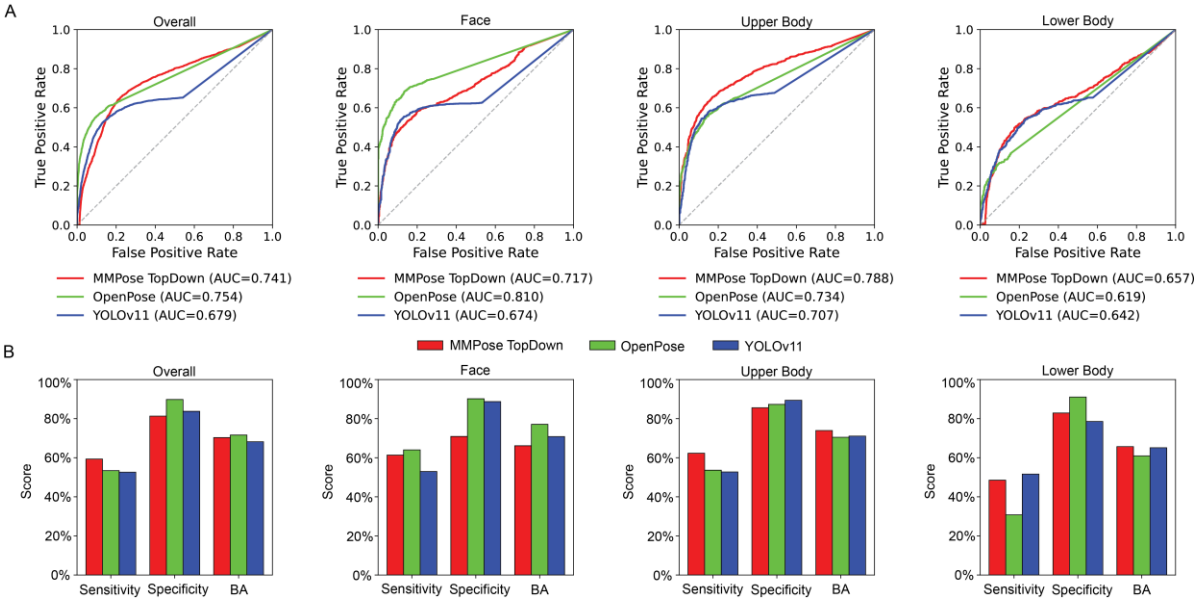
automated computer vision performance for inpatients in neurological services may require custom models based on clinical datasets or the development of novel, potentially combined, model approaches. Together, these preliminary results introduce an approach that could enhance real-time clinical evaluations in inpatient neurology settings, offering a path toward real-time detection of falls, delirium, and seizures as the model performance continues to improve.

Human Pose Estimation keypoints



https://secure.key4events.com/onsite/api.aspx?e=1960&img=Figure+1.Body+keypoints+overlay.png&ai=74134&op=getabstractimg&dirN=0

Performance Pose Estimation models



https://secure.key4events.com/onsite/api.aspx?e=1960&img=Figure2.ROC.png&ai=74134&op=getabstractimg&dirN=1

P40

A Calibrated Ensemble Framework for Video-Based Differential Diagnosis of Sleep-Related Hypermotor Epilepsy and Parasomnias

F. Sassi[1], R. Cordani[2, 3], A. Castelnovo[4, 5], L. Tassi[6], F. Provini[7, 8], P. Proserpio[9], F. Odone[1, 10], M. Casadio[1], M. Moro[1, 10], L. Nobili[2, 3]

[1]Department Of Informatics, Bioengineering, Robotics And Systems Engineering (dibris), University Of Genova, 16146, Genova, Italy - Genoa (Italy), [2]Child Neuropsychiatry Unit, Irccs Istituto Giannina Gaslini, Member Of The Ern Epicare, 16147, Genova, Italy - Genoa (Italy), [3]Dipartimento Di Neuroscienze, Riabilitazione, Oftalmologia, Genetica E Scienze Materno-Infantili (dinogmi), University Of Genova, 16146, Genova, Italy - Genoa (Italy), [4]Sleep Medicine Unit, Neurocenter Of Southern Switzerland, Ente Ospedaliero Cantonale, Lugano, Switzerland - Lugano (Switzerland), [5]Faculty Of Biomedical Sciences, Università Della Svizzera Italiana, Lugano, Switzerland - Lugano (Switzerland), [6]“c. Munari” Epilepsy Surgery Centre, Niguarda Hospital, Milan, Italy - Milan (Italy), [7]Department Of Biomedical And Neuromotor Sciences, Bellaria Hospital, University Of Bologna, Bologna, Italy - Bologna (Italy), [8]Irccs Institute Of Neurological Sciences Of Bologna, Bellaria Hospital, Bologna, Italy - Bologna (Italy), [9]Irccs San Raffaele Hospital, Milan, Italy - Milan (Italy), [10]Machine Learning Genoa (malga) Center, University Of Genova, 16146, Genova, Italy - Genoa (Italy)

Introduction

Leveraging modern video-based deep learning approaches in the clinical domain presents unique challenges, including limited data availability, ethical constraints, and the difficulty of obtaining expert annotations In this context, differentiating between epilepsy subtypes and parasomnias using video data alone presents distinct challenges, since the visual differences between classes are subtle and frequently limited to short, temporally circumscribed events. In this work, we introduce a clinically oriented pipeline designed for video data to support the differential diagnosis of Sleep-Related Hypermotor Epilepsy (SHE), Disorders of Arousal (DOA), and REM Sleep Behavior Disorder (RBD). Although the dataset remains limited in size, our findings demonstrate robust classification performance, underscoring the feasibility of integrating video-based, AI-assisted differentiation of seizures and parasomnias into real-world medical workflows.

Methods

Dataset

The dataset was acquired by different centres: (i) the Sleep Medicine Center, “C. Munari” Center of Epilepsy Surgery, Niguarda Hospital, Milan, Italy; (ii) the Sleep Medicine Unit, Neurocenter of Southern Switzerland, Regional Hospital of Lugano, Lugano, Switzerland; (iii) the Child Neuropsychiatry Unit, IRCCS G. Gaslini, Genoa, Italy and (iv) the Bellaria Hospital in Bologna. The retrospective study was approved by the Niguarda Hospital ethics committee (ID 939–12.12.2013), and all patients gave written consent for the use of video recordings for research and publication purposes.

The dataset contains video-polysomnographic (v-PSG) recordings used for sleep-related paroxysmal events and includes 68 videos from 60 different participants collected in real-world clinical settings leading to heterogeneity in video quality, duration, lighting, and background. We included only videos meeting two criteria: (1) at least one fully recorded habitual episode, and (2) a definitive diagnosis of Sleep-Related Hypermotor Epilepsy (SHE) or Disorders of Arousal (DOA) or REM Behavior Disorder (RBD). SHE episodes were grouped into four Semiology Patterns (SP): (i) SP1: elementary motor signs, (ii) SP2: unnatural hypermotor movements, (iii) SP3: integrated hypermotor movements and (iv) SP4: gestural behaviours with high emotional content. The diagnosis was established by two expert physicians based on both video inspection and electroencephalography (EEG) analysis and it was considered as our ground truth.

Implemented pipeline

We approached the problem of differentiating between epileptic and parasomnic episodes as an action recognition task. Initial experiments focused on training a single multiclass classifier using pretrained spatiotemporal models. As architecture, we selected a SlowFast Network for its capacity to capture both rapid, fine-grained movements and broader contextual motion patterns through a dual-pathway architecture. However, even with transfer learning from large-scale datasets (i.e., Kinetics-400), models failed to generalize.

To address this, as showed in Figure 1, we restructured the task into 15 binary classification problems, one for each pair of classes, and then combined the outputs to produce a single prediction for each sample. We initially applied three static aggregation rules to obtain the final predicted class: (i) summation of probabilities, (ii) majority voting, and (iii) weighted voting based on validation accuracy; however, these approaches did not yield the desired performance. We therefore concatenated the probability outputs of all binary classifiers and shifted towards dynamic aggregation methods such as stacking (ensemble learning). In this way, each video is represented by a 30-dimensional feature vector (15 classifiers × 2 probabilities), which is then used to train a meta-classifier to predict the final class. We conducted a grid search over four families: Logistic Regression (LR), K-Nearest Neighbours (KNN), Random Forests (RF), and fully connected Neural Networks (NN). While this decomposition may introduce redundancy, it simplifies the learning task and reduces confounding effects between clinically overlapping classes.

The dataset consists of 68 videos across six classes, four SHE semiology patterns (SP1–SP4), DOA, and RBD. Both the validation and test sets include 2 videos per class, each from different participants. The training set is structured as follows: SP2, SP3, DOA, and RBD each contain 8 videos; SP1 contains 7 videos; and SP4 contains 5 videos. The test set is used exclusively to evaluate the performance of the whole pipeline, and no videos from test participants were seen during training and validation.

Results

All 15 binary classifiers achieved at least 75% accuracy (3/4 videos), indicating that decomposing the task into binary problems allowed the models to better capture the discriminative patterns between classes. Classifiers involving SP1 consistently reached 100% accuracy across all pairings, indicating that elementary motor signs are clinically distinct and identifiable. In contrast, classifiers involving SP2 and SP3 plateaued at 75%, reflecting the clinical overlap between these two semiology patterns. Similarly, classifiers involving SP4 performed well against other SHE subtypes but dropped to 75% when compared with RBD, which is consistent with the resemblance between RBD dream-enactment behaviors and the emotional–gestural features typical of SP4. Finally, parasomnias (DOA vs. RBD) were fully separable (100% accuracy).

We evaluated how the calibrated probabilistic outputs could be combined to obtain multiclass predictions. Among the three static aggregation methods, weighted voting achieved the best performance at 75% (9/12), while majority voting and probability summation reached 66%. Despite

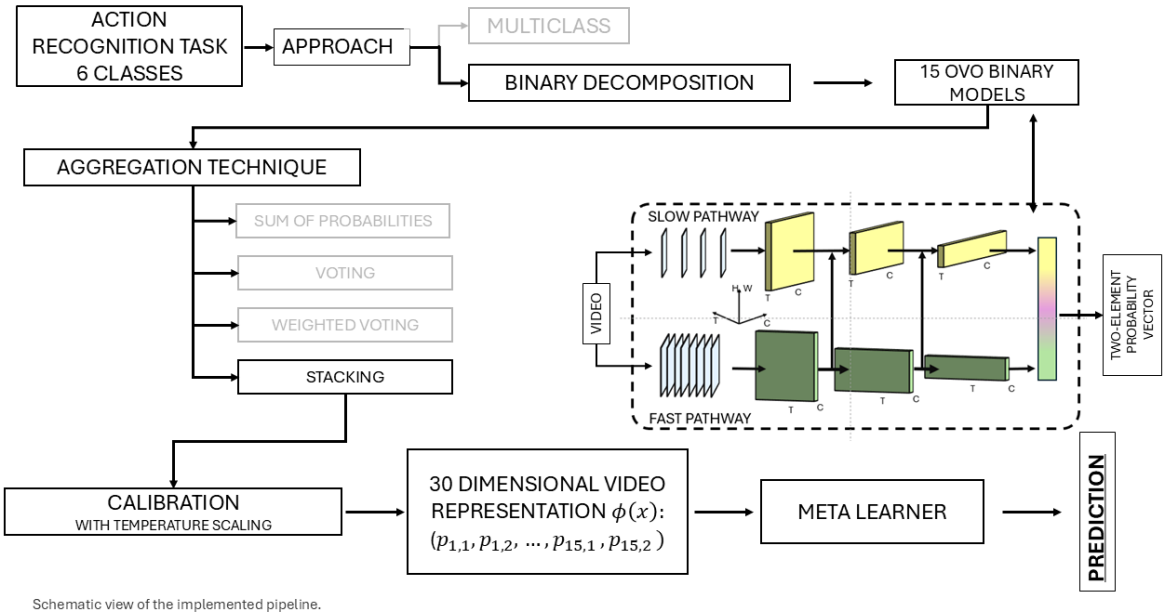
performing reasonably well, these methods plateaued and did not offer further improvements, motivating the shift to more flexible approaches such as stacking. Among the meta-classifiers, logistic regression achieved the highest test accuracy at 83.3% (10/12), followed by KNN at 75%, while Random Forests and Neural Networks reached 66.7%.

Conclusions

Our results, consistent with prior works, highlight the inherent challenges of multiclass classification in video-based clinical contexts. Specifically, the combination of high intra-class variability and substantial visual overlap between epilepsy subtypes and parasomnias leads to ambiguous decision boundaries that are difficult to learn from limited annotated data. However, stacking on calibrated binary outputs can effectively capture nonlinear interactions and provide robust multiclass predictions for real-world clinical video analysis, even in small cohorts.

By aligning experimental outcomes with established clinical observations, our results indicate that this approach enhances robustness and reliability in video-based medical classification, even under conditions of data scarcity. In conclusion, this work provides a concrete foundation for developing clinically deployable AI tools to support the differential diagnosis of nocturnal events.

Schematic view of the implemented pipeline.



Schematic view of the implemented pipeline.

https://secure.key4events.com/onsite/api.aspx?e=1960&img=implemented_pipeline.png&ai=74135&op=getabstractimg&dirN=0

P41
Adaptive rest–active state detection of generalized convulsive seizures

S.C. Meuwly[1], A. Spahr[1], A. Bernini[1], P. Ryvlin[1]

[1]Neurodigital@neurotech, Department Of Clinical Neurosciences, Lausanne University Hospital (chuv), University Of Lausanne, Lausanne, Switzerland - Lausanne (Switzerland)

Introduction

Several automated algorithms for generalized convulsive seizure (GCS) detection already exist. High nocturnal sensitivity is required to avoid missing critical events associated with a risk of sudden unexpected death in epilepsy (SUDEP). At the same time, reducing false alarms during daytime or active periods remains a major challenge. An adaptive approach that distinguishes rest and active states could allow seizure detection systems to adjust their sensitivity according to the patient's physiological context. The aim of this study was to develop and evaluate a wearable-based algorithm capable of reliably detecting rest versus active periods, as a first step toward adaptive convulsive seizure detection for improved SUDEP prevention.

Methods

We developed a convolutional neural network combined with a bidirectional long short-term memory network (CNN-BiLSTM) to classify rest and active states using wrist-worn accelerometer data. This model allowed us to capture both short-term motion patterns and longer temporal context. The rest label was derived from sleep annotations and used as a proxy for low-activity nocturnal periods. The model was trained on the CAPTURE-24 dataset, which contains approximately 24 hours of wrist accelerometer recordings from 151 participants, with ground truth annotations obtained from body-worn cameras during the day and sleep diaries at night. Data were split subject-wise, with 80% used for training and 20% for validation. To assess generalizability, the trained model was further evaluated on two independent datasets: (1) a Wear OS smartwatch dataset, consisting of 30 consecutive days of wrist accelerometer data from one participant, recorded on an Android-based smartwatch with a custom app and manually annotated sleep periods, and (2) the Severity dataset, collected during a multicenter clinical study in epilepsy monitoring units across Switzerland and Europe, where patients with epilepsy wore a wrist-worn band collecting 3D accelerometer data during their stay and wakefulness state before seizures was clinically annotated, with 97 convulsive seizures. For the Severity dataset, model predictions were evaluated on the 15-minute pre-seizure window for GCS occurring during sleep.

Results

On CAPTURE-24, the model achieved 96% overall accuracy over 3882 hours of recordings with no evidence of overfitting. On the Wear OS smartwatch dataset, overall accuracy reached 95% over ~34 days of recordings, with most errors concentrated around sleep–wake transitions, partly reflecting lack of precision in the labeling of these transitions. On the Severity dataset, the model achieved 88% accuracy over 48 pre-GCS sleep segments.

Conclusions

This study demonstrates that rest and active states can be robustly detected from wrist-worn accelerometer data using a CNN-BiLSTM architecture across multiple datasets. Integrating this rest–active classifier with existing GCS detection algorithms could enable adaptive sensitivity, reducing false alarms during active periods while maintaining high sensitivity during rest, when SUDEP risk is highest.

P42

Unsupervised Machine Learning for Automated Detection of Interictal Epileptiform Discharges in SEEG

A. Cohen[1], I. Goldberg[2], F. Benninger[2], M. Ramot[3]

[1]Weizmann Institute Of Science - Karkoor (Israel), [2]Rabin Medical Center - Petah Tikva (Israel), [3]Weizmann Institute Of Science - Rehovot (Israel)

Introduction

Interictal Epileptiform Discharges (IEDs) are critical biomarkers for diagnosing epilepsy and guiding surgical planning, yet their identification in stereo-EEG (SEEG) remains labor-intensive and highly dependent on expert annotations. Current AI automated approaches rely on supervised learning, requiring large, high-quality labeled datasets that are often difficult to obtain in clinical environments. To address this limitation, we developed an unsupervised machine-learning framework designed to detect IEDs directly from raw SEEG signals without expert-provided labels. This study evaluates the model's ability to isolate epileptiform events, characterize their structure, and match expert epileptologist assessments.

Methods

We analyzed several hours of sEEG recordings collected through our collaboration with Belinson Hospital - Rabin Medical Center from patients with drug-resistant epilepsy.

To validate model performance, we used two fully labeled patients from the same institution, covering both day and night conditions, with 15 minutes of annotated data per condition. In addition, we incorporated an external labeled dataset from a publicly available source, consisting of 5 patients, each providing 5 minutes of expert-annotated IED events, to support benchmarking and cross-dataset validation.

The pipeline consists of: (1) preprocessing steps to remove artifacts, improve signal to noise ratio, filter out bad channels and outliers (2) feature extraction using IED's morphology features (3) unsupervised clustering to isolate high-probability IED candidates. Cluster archetypes were interpreted using characteristic waveform morphology, temporal and spatial distribution. Performance was evaluated by comparing detected IED windows to independent epileptologist annotations, using metrics including Cohen's kappa, precision and recall.

Hyperparameters and thresholds were tuned automatically without using labeled data.

Results

The unsupervised model successfully identified distinct clusters corresponding to IED morphology, detecting both isolated spikes and multi-channel epileptiform discharges. When benchmarked against expert annotations, high-probability clusters were confirmed by epileptologist to represent clinically valid IED events.

Across the clinically annotated patients from our collaborating medical institution, the model achieved high and consistent performance, with an average precision of 94.5%, recall of 83%, and Cohen's κ = 0.81, demonstrating agreement levels comparable to supervised approaches despite requiring no labels.

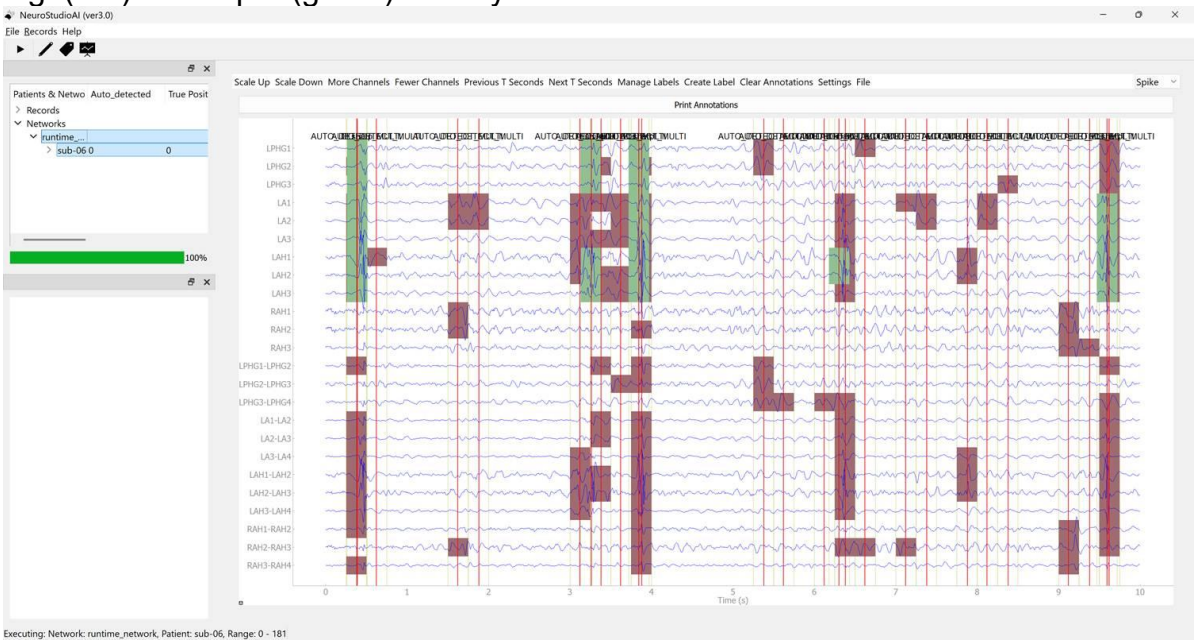
Evaluation on an external publicly available dataset showed greater variability, with recall ranging from 81% to 88% and reaching 100% in two cases; precision ranged from 0.23 to 0.81, and Cohen's κ ranged from 0.34 to 0.80. We attribute this variability partly to differences in annotation methodology: our clinical dataset was labeled using a streamlined one-click user interface dedicated application enabling dense and consistent marking, whereas the public dataset likely relied on conventional manual annotation that may emphasize only the most prominent IEDs.

Conclusions

These findings demonstrate the feasibility of unsupervised machine learning for reliable IED detection in SEEG and highlight its potential to reduce reliance on expert labeling. Analysis of additional patients is currently underway, and expanded results will be presented at the conference. While preliminary, the current outcomes show promising performance across diverse recording conditions.

A key direction for future work is to investigate the events detected by the model that were not marked by neurologists, in light of the well-established inter-rater variability in IED annotation. This raises an important ground-truth question: whether such model-identified events represent true epileptiform activity that may be overlooked in manual review, borderline patterns of uncertain significance, or non-IED signal fluctuations. Understanding this distinction will be essential for defining more objective benchmarks and improving both clinical and algorithmic interpretation of IEDs

algo(red) and expert(green) overlayed annotations



https://secure.key4events.com/onsite/api.aspx?e=1960&img=algo_vs_expert_annotations.png&ai=74138&op=getabstractimg&dirN=0

P43
Optimising machine-learning detection of generalized convulsive seizures

L. Goffinet[1], A. Spahr[1], A. Bernini[1], S. Lahtoo[2], Y.S. Vakilna[2], C. Baumgartner[3], J.P. Koren[3], L. Imbach[4], S. Beniczky[5], S. Rheims[6], M.E. Fabricius[7], I. Beuchat[1], D. Atienza Alonso[8], C.E. Bardyn[1], P. Ryvlin[1]

[1]Neurodigital@neurotech, Department Of Clinical Neurosciences, Lausanne University Hospital (chuv), University Of Lausanne, Lausanne, Switzerland - Lausanne (Switzerland), [2]Department Of Neurology, Mcgovern Medical School, The University Of Texas Health Science Center At Houston, Houston, Tx, Usa - Houston (United States), [3]Karl Landsteiner Institute For Clinical Epilepsy Research And Cognitive Neurology, Medical Faculty Sigmund Freud University, Vienna, Austria - Vienna (Austria), [4]Swiss Epilepsy Center, Klinik Lengg, Zurich, Switzerland - Zurich (Switzerland), [5]Department Of Clinical Neurophysiology, Aarhus University Hospital, Aarhus, Denmark - Aarhus (Denmark), [6]Department Of Functional Neurology And Epileptology, Hospices Civils De Lyon, Lyon 1 University, Lyon, France - Lyon (France), [7]Department Of Clinical Neurophysiology, Copenhagen University Hospital Rigshospitalet, Copenhagen, Denmark - Copenhagen (Denmark), [8]Embedded Systems Laboratory, Epfl, Lausanne, Switzerland - Lausanne (Switzerland)

Introduction

Machine learning algorithms using wearable sensors data to detect generalized convulsive seizures (GCS) are trained on data acquired in epilepsy monitoring units (EMU) where patients have very limited motor activity. Accordingly, they typically achieve high sensitivity and low false alarm rate (FAR) over independent datasets acquired in the same controlled environment but display significantly higher FAR in real-world ambulatory use, particularly during daytime. We aimed at tackling this issue by incorporating real-world 3D-accelerometry data acquired in non-epileptic subjects in the training and validation of a GCS detection algorithm.

Methods

Three 3D-accelerometry datasets were used for this study: 1) three EMUs datasets that pooled 435 persons with epilepsy (PWE), including 85 who had a total of 126 GCS (40.739 hours of recordings), 2) the public dataset CAPTURE-24 which includes 151 normal individuals with precisely curated daily activities (3883 hours of recordings), 3) a private dataset of 49 migraine patients with no history of seizures (45.063 hours of recordings).

We developed an ambulatory-aware ensemble model using the same method as that previously published for EpiSave (Spahr et al. Epilepsia 2025). The model was trained on data from 37 PWE with 53 GCS and the 151 normal controls from CAPTURE-24 and validated on data from 398 PWE with 73 GCS (sensitivity and FAR) and 49 migraine patients (FAR only). Overall, FAR was assessed over 81,683 hours of recording. We further tested the impact of lightweight event-level post-detection false-alarm classifier, using a fixed temporal window of accelerometer data preceding the detection to reject false alarms without creating new detections.

We compared the performance of the EpiSave model exclusively trained on EMU data with the ambulatory-aware model, with and without the adding of the post-detection classifier to both models. Performance was assessed on continuous recordings and stratified by setting (EMU vs ambulatory) and time of day (16 hours of day vs 8 hours of night). Whenever possible, we chose as an optimal model quantile, that providing 100% sensitivity with the lower FAR.

Results

For day-time recording, both the EpiSave and ambulatory-aware models reached 100% sensitivity. In contrast, the ambulatory-aware model reduced the daytime FAR observed on ambulatory real-world data by 24-fold, down to 0.005/hour, translating into 1 daytime false alarm every 12.5 days. The post-detection classifier only marginally improved this metric down to 1 daytime false alarm every 13.6 days.

In contrast, for nighttime recordings, the ambulatory-aware models did not improve the FAR. The post-detection classifier reduced the FAR observed with the EpiSave algorithm from 0.003/hour to 0.002/hour in the EMU, and from 0.009/hour to 0.005/hour on ambulatory recordings while keeping 100% sensitivity. These latter figures likely overestimate the true sleep FAR in ambulatory setting since the migraine recordings used for this evaluation included patients with nocturnal migraine attacks.

Conclusions

Incorporating ambulatory data from non-epileptic subjects during training and adding a post-detection classifier provide complementary strategies to lower FAR in ambulatory PWE while preserving optimal sensitivity.

P44
Circadian Rhythms in High-Frequency Oscillations from Long-Term Intracranial EEG

S. Das[1], A. Alsammani[1, 2], W. Stacey[3], S. Gliske[1]

[1]University Of Nebraska Medical Center - Omaha (United States), [2]Delaware State University - Dover (United States), [3]University Of Michigan - Ann Arbor (United States)

Introduction

High-frequency oscillations (HFOs) have emerged as promising biomarkers of epileptic tissue. Their potential clinical relevance has expanded rapidly through automated detectors and machine-learning models that quantify temporal dynamics of HFOs. While circadian and ultradian patterns in epilepsy have been reported, the temporal dynamics of HFOs across the 24-hour cycle remains less understood. Understanding whether HFO rates exhibit consistent circadian fluctuations could enhance our mechanistic understanding of epilepsy and support future chronobiological strategies in seizure forecasting. This study aimed to quantify circadian rhythmicity in HFO rates across a large cohort of epilepsy patients undergoing long-term intracranial EEG (iEEG) monitoring as presurgical evaluation.

Methods

We retrospectively analyzed continuous, multi-day, sleep-scored iEEG recordings from 68 patients with drug-resistant epilepsy at the University of Michigan between 2016 and 2022. Patients with daylight savings time transitions were excluded to avoid any potential changes in circadian pattern, resulting in 66 subjects. HFOs were detected using a validated automated algorithm (qHFO detector), which is based on the Staba HFO detector. The qHFO detector which includes rigorous artifact rejection steps, was applied to non-overlapping 10-minute windows, resulting in a highly specific set of HFOs. To assess 24-hour cyclicity in the HFO rates, we computed the first Fourier moment of the HFO rate for each patient within multiple states of vigilance, which is proportional to the first circular moment. Given the uneven distribution of sleep stages, statistics were too limited to analyze the 24-hour cyclicity of every sleep stage individually. We combined all the NREM stages together and compared with the awake state. To assess rhythmicity, we computed the first circular statistic, corrected for non-uniform sampling and incomplete data using the unfolding method. The unfolding method is a patient-specific AI model that used simulated data along with their recording times. We further examined if the zeniths of HFO rates clustered within relevant time windows.

Results

We observed significant 24-hour cyclicity in HFO rates during NREM sleep in all 66 patients ($p < 0.05$) and in awake state in 65 subjects ($p < 0.05$). These findings demonstrate that circadian modulation of HFO dynamics is a robust feature of intracranial electrophysiology rather than an artifact of uneven sampling. Across subjects, 36 of 66 individuals had their NREM HFO zenith

between 10 PM and 4 AM ($p = 7.51 \times 10^{-8,}$ binomial test). 28 out of 65 patients had the zenith of the awake stage HFO rates between 2 PM and 6 PM ($p = 1.25 \times 10^{-7}$, binomial test).

Conclusions

Circadian patterns in HFO rates are statistically significant and differ across vigilance states. Although the magnitude of rate fluctuations is modest, their consistency across patients suggests that a possible biological rhythm influences HFO dynamics. The use of the AI-based unfolding model strengthens the validity of these findings by correcting for irregular sampling in long-term clinical recordings. These results support the integration of time-of-day information into AI-driven seizure prediction, biomarker modeling to aid in making clinical decisions. Circadian information may improve the interpretability of HFO-based assessments of epileptogenic tissue and could enhance chronobiology-informed therapeutic strategies.

Patient Demographics

Table 1: Patient Demographics	
Subjects	N = 68
Age category	Children: 16 (age 5-17) Adults: 50 (age 20-66)
Gender	Male: 29 Female: 37
Engel Class I	N=15

https://secure.key4events.com/onsite/api.aspx?e=1960&img=Table+1.png&ai=74141&op=getabstractimg&dirN=0

P45

Detecting Seizure Onset Patterns as Temporal Objects: A Generalizable YOLO-Based Framework for iEEG Time Series Event Detection

Z.C. Xu[1], C. Aguila[1], H. Wu[1], N. Sinha[1], B. Litt[1], E. Conrad[1]

[1]University Of Pennsylvania - Philadelphia (United States)

Introduction

Seizure onset patterns (SOPs) are a great biomarker that strongly correlates with a patient's underlying epilepsy network, allowing us to localize the seizure onset zone and predict outcomes after clinical intervention. Currently, SOP classification is a manual, visual review by clinicians, a process that is time-consuming, plagued by subjectivity and poor reliability. In turn, this makes large, multi-cohort studies not feasible, limiting our ability to translate findings into definitive clinical practice. We propose to leverage a large, diverse cohort to build a novel, robust and generalizable deep learning framework to detect and classify SOPs. Ultimately, this tool will pave the way for a new generation of personalized, data-driven surgical planning.

Methods

We developed a YOLO-based temporal object-detection framework to identify seizure onset patterns (SOPs) from intracranial EEG (iEEG). The model adapts the YOLOv8 architecture for 1-D biosignals by using a1-D CNN encoder and modifying the detection head to predict temporal bounding boxes (start time, end time) and class labels. Each 60 second iEEG segment is treated as a "temporal image," enabling the model to simutanously localize multiple events at different scales within the same clip. We benchmark the detector against standard classification approaches and human annotator in performance.

Results

We analyzed over 2,000 seizures collected from 179 patients at the Hospital of the University of Pennsylvania (HUP) and 29 patients at the Medical University of South Carolina (MUSC). SOPs were categorized into five clinically validated patterns: low-voltage fast activity (LVFA), low-frequency repetitive spiking (LFRS), sharp activity, rhythmic slow spikes, and bursts of polyspikes. Our YOLO-based detection model achieved a preliminary mean Average Precision at IoU = 0.5 (mAP@0.5) of 0.783, demonstrating robust performance across diverse implant types and recording environments. Using these SOP classifications and patient outcome data, we will evaluate whether particular patterns can identify strong surgical candidates and predict treatment success two years after intervention. We will also use SOPs classification to predict seizure onset zone/epileptogenic zone.

Conclusions

This framework delivers a fast, scalable alternative to human classification and traditional univariate sliding-window SOP classifiers by treating SOPs as temporal objects, enabling precise, multi-scale event localization with very high speed. Its flexible YOLO-based design easily supports extension to additional epileptiform biomarkers, such as spikes and HFOs, making it well-suited for large, heterogeneous datasets. Ultimately, this framework will help uncover how specific onset patterns relate to underlying epilepsy networks, improving our ability to target interventions with greater precision.

P46

Preictal High-Connectivity States in Epilepsy: Multicenter Evidence and Network-Level Mechanisms Underlying Seizure Emergence

A. Tauste Campo[1], N. Medina[1], M. Vila-Vidal[2], M. Centeno[3], E. Conde[3], M. Carreño[3], A. Donaire[3]

[1]Universitat Politècnica De Catalunya - Barcelona (Spain), [2]Brainfocus Labs Sl - Barcelona (Spain), [3]Hospital Clínic Barcelona - Barcelona (Spain)

Introduction

Identifying reliable preictal biomarkers is essential for advancing neuromodulation strategies in drug-resistant epilepsy. Prior single-center work suggested that seizures are preceded by transient reconfigurations of large-scale functional connectivity, specifically through short-lasting (≈0.6 s) high-connectivity states (HCSs) that increase in probability in the hours before seizure onset. These findings imply that seizure susceptibility may arise from global network processes rather than from strictly local changes in neuronal activity.
Here we extend this line of work by (i) assessing the reproducibility of HCS dynamics in a multicenter cohort of patients with drug-resistant epilepsy and (ii) characterizing the network mechanisms underlying their emergence using detailed SOZ network analyses and a low-dimensional stochastic model.

Methods

Continuous intracranial EEG recordings were analyzed from 12 patients recruited across multiple clinical centers, including both stereo-EEG (sEEG) and electrocorticography (ECoG) implants. For each patient, we computed dynamic functional connectivity during long interictal and preictal periods and performed unsupervised clustering to identify recurring network states. The cluster with the highest average global connectivity was designated as the HCS.
To evaluate preictal modulation, we quantified the probability of HCS occurrence in sliding windows across hours preceding each seizure. Statistical comparisons between interictal and preictal periods assessed whether HCSs systematically increase before seizure onset.
In two patients with high-resolution clinical characterization of the seizure-onset zone (SOZ), we investigated whether preictal HCS increases were associated with changes in SOZ network centrality, measured using graph-theoretic indices that capture influence and integration within the global network (e.g., eigenvector centrality).
Finally, to examine the underlying mechanism of HCS emergence, we constructed a low-dimensional stochastic network model linking network topology to dynamic state transitions. By manipulating the connectivity structure between SOZ and non-SOZ nodes, we tested whether gradual shifts in network topology could generate the observed increase in preictal HCS probability without requiring changes in mean connectivity.

Results

Across the multicenter cohort, we found consistent preictal increases in HCS probability in most patients, demonstrating that these phenomena generalize across diverse clinical settings, electrode configurations, and epilepsy etiologies, highlighting their potential applicability as broad preictal biomarkers.
In the two patients with detailed SOZ localization, the rise in HCS probability was closely accompanied by increasing SOZ network centrality. As the preictal period unfolded, SOZ contacts became progressively more integrated within the large-scale functional network, showing stronger influence on global connectivity patterns.
The stochastic network model reproduced the empirical preictal increase in HCS probability through topological reorganization, specifically by modulating the relative weighting of SOZ–non-SOZ connections. Importantly, the model captured the observed rise in HCS occurrence without requiring an overall increase in mean connectivity. This indicates that preictal HCSs arise from network-level structural shifts rather than from simple global amplification of connectivity

Conclusions

Our results provide strong multicenter evidence that high-connectivity states constitute a reproducible preictal biomarker across diverse clinical environments. The convergence of empirical findings and computational modeling supports the interpretation that HCS emergence reflects a network-topological mechanism, in which the seizure-onset zone becomes increasingly integrated into the broader functional network during the approach to a seizure. These findings highlight the importance of large-scale network reconfiguration in seizure generation and underscore the value of connectivity-based biomarkers for seizure forecasting and neuromodulation strategies.

Influence of SOZ dynamics on HCS proba

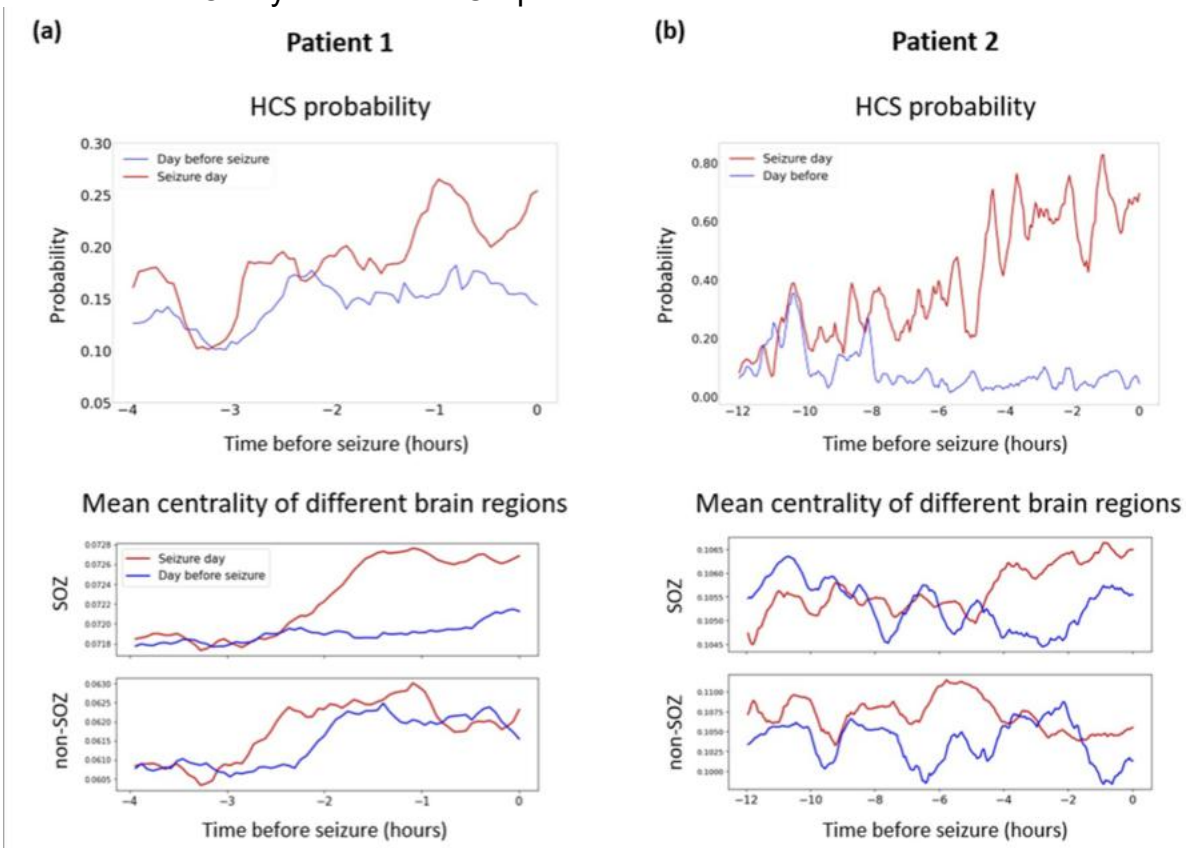


Figure 6. Influence of the SOZ dynamics on pre-seizure HCS probability (Comparison between P1 and P2). 6(a) The HCS probability and the mean centrality of each brain region are shown for P1 (similarly to figure 5 (a,b)). 6(b) Top: In P2 it is observed that the HCS probability increases as the seizure approaches, specifically, this happens from 8 h before the onset of the seizure. 6(b) Bottom: it shows an increase in the mean centrality located in the SOZ channels that is evident from 5 h before the onset of the seizure.

https://secure.key4events.com/onsite/api.aspx?e=1960&img=Fig_1.png&ai=74143&op=getabstractimg&dirN=0

Network model reproducing HCS probability

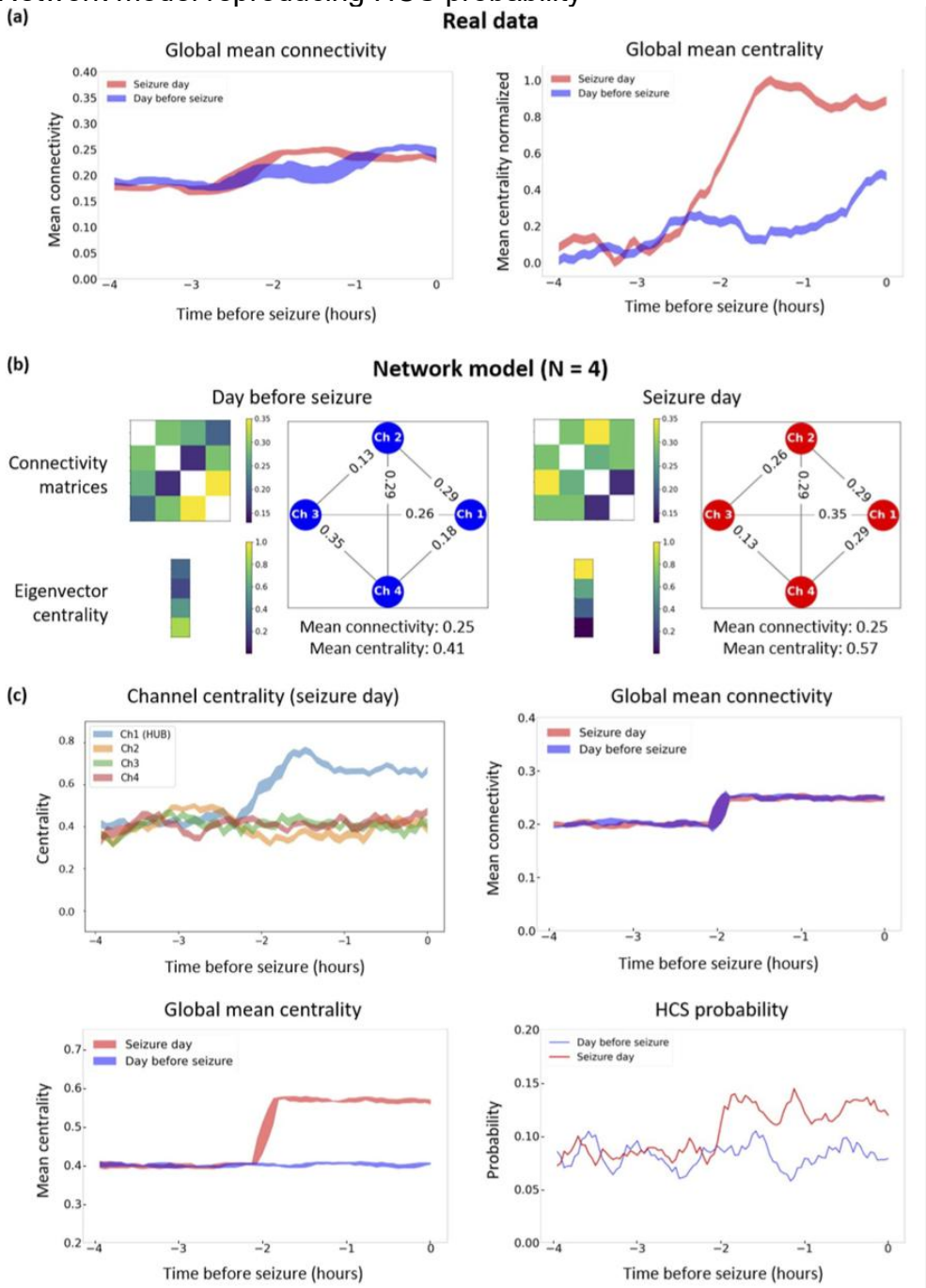


Figure 7. Low-dimensional stochastic network model reproducing the preictal emergence of HCS probability. (a) The global mean connectivity and centrality of patient P1 are presented, illustrating how dynamic shifts in centrality—rather than connectivity—guided the development of our stochastic network model. (b) The model generates random correlation matrices ($N = 4$) with fixed mean connectivity and varying mean centrality by rearranging correlation coefficients. This rearrangement results in two distinct network states: one with a connectivity matrix whose coefficients produce nodes without dominant centrality, (Left panel: lower mean centrality, as observed the day before the seizure) and another exhibiting a hub-like structure (Right panel: higher mean centrality, resembling the seizure day). Network states are classified based on eigenvector centrality, prioritizing centrality over connectivity. (c) Results averaged over 40 simulation runs. The four subfigures are interrelated, as they arise from the same set of simulations. The figure shows that HCS probability (bottom right) increases significantly only when a specific channel (e.g. Ch1) exhibits dominant centrality (top left), forming a hub (analogous to the clinical SOZ). This centrality dominance explains the observed rise in HCS probability during the 2 h preceding a seizure. The corresponding global mean connectivity and mean centrality are presented in the top right and bottom left figures, respectively. Together, these results demonstrate that the increase in HCS probability emerges as a topological property of the model. Time ranges are selected for each patient to highlight the pre-seizure effect.

https://secure.key4events.com/onsite/api.aspx?e=1960&img=Fig_2.png&ai=74143&op=getabstractimg&dirN=1

P47

EpiScalp™: A Novel AI-Based Tool for Improved Epilepsy Diagnosis in the EMU From Interictal Scalp EEG

M. Hays[1], G. Kamali[1], M.R. Streett[2], R. Miranda[3], G. Ren[3], A. Walia[4], S. Johnson[5], D. Wyeth[5], A. Gotshalk[1], J. Hopp[4], M. Sperling[5], J. Gonzalez-Martinez[6], S. Sarma[7]

[1]Neurologic Solutions - Baltimore (United States), [2]Johns Hopkins Hospital - Baltimore (United States), [3]University Of Maryland Medical Center - Baltimore (United States), [4]University Of Maryland School Of Medicine - Baltimore (United States), [5]Thomas Jefferson University Hospital - Philadelphia (United States), [6]University Of Pittsburgh School Of Medicine - Pittsburgh (United States), [7]Johns Hopkins University - Baltimore (United States)

Introduction

50 million people worldwide are diagnosed with epilepsy, but around 30% may be misdiagnosed. High misdiagnosis rates stem from lack of a simple, reliable diagnostic process, instead relying on a comprehensive clinical evaluation. Visual inspection of a routine scalp EEG can confirm an epilepsy diagnosis if interictal epileptiform discharges (IEDs) are present, but the rarity of IEDs often leads to inconclusive diagnoses. Patients may undergo further evaluation in the Epilepsy Monitoring Unit (EMU) for days to weeks to capture a seizure or IEDs on EEG while off anti-seizure medications and sleep-deprived to provoke seizures. Despite this costly process, some patients have insufficient or no seizures, leading to an inconclusive or possibly incorrect diagnosis. By characterizing imbalances within brain networks that can't be visually detected from EEG, dynamical network models (DNMs) offer a potential solution to this clinical need for accurate diagnoses from inconclusive EEG. While EpiScalp™, our patent-protected algorithm that applies these DNMs to predict epilepsy diagnoses, was originally validated in routine EEG, we sought to investigate its performance in the noisier EMU setting over a continuous 24-hour period with a more heterogeneous patient population.

Methods

Through an NINDS SBIR Phase 2 grant, EpiScalp™ is being currently used in a research study at three level 4 Epilepsy Centers (Johns Hopkins Hospital, Thomas Jefferson University Hospital, and University of Maryland Medical Center), where we are gathering the first 24 hours of continuous scalp EEG data for patients undergoing evaluation in the EMU.

Conceptually, EpiScalp™ finds EEG patterns which are below the threshold of traditional visual interpretation. Dynamical network models (DNMs) derived from EEG data characterize the internal properties of the brain and are necessary to reveal the epileptic network connections and the underlying dynamics of seizure generation. By approximating short windows of EEG as linear time invariant models, DNMs were built to characterize the network's N to N channel interactions across time. We hypothesized that neural fragility and source-sink index, two DNM-derived metrics previously used to localize the epileptogenic zone (EZ) in intracranial EEG recordings, can be applied to interictal scalp EEG to distinguish between epileptic and non-epileptic brains. The EZ nodes in epileptic brains would have high fragility and behave like strong sinks, while these dynamics would be absent in non-epileptic brains.

In our application, a "sink” node is a region in the brain network that is being highly influenced by other nodes but is not influential itself. In contrast, a “source” node is a region that is highly influential to other nodes in the network but is not being influenced by other nodes. Our source-sink claim is that the seizure onset zone (SOZ) in epilepsy is a group of nodes that are being highly inhibited between seizures and thus cannot influence the network, i.e., the SOZ are sink nodes in between seizures. Prior to seizure initiation, however, the SOZ nodes transition into sources, due to a lapse in inhibition or an excitatory trigger, that then generate a seizure. We define “fragility” as a network level measure of how sensitive a brain region is to small perturbations in the network. A region is considered fragile if a small change in that node’s outgoing or incoming network weights causes a disproportionate disruption in the model’s ability to maintain stable dynamics, which indicates that the region plays a critical role in seizure initiation or propagation pathways. In other words, fragile nodes are regions where the DNM is least tolerant to perturbation — consistent with seizure-generating or seizure-related circuits.

We trained a logistic regression model on 442 routine scalp EEG patients (221 epilepsy, 169 non-epilepsy randomly upsampled to 221) to predict epilepsy and non-epilepsy diagnoses using fragility, source-sink, and bandpower features. The model was tested on 164 EMU patients, where features and the resulting prediction probability were computed for each 20-minute segment of the first 24 hours of EMU recordings separately. The predicted probabilities were converted to epilepsy risk scores ranging from 0-100, and the median score over all 20-minute segments was used to compute an overall risk score for each patient.

Results

EpiScalp™ was tested on 164 patients who were assigned the following clinical labels after EMU monitoring: definitely epilepsy (n=93), definitely non-epilepsy (n=35), or uncertain (n=36). EpiScalp™ made a predictive definitive diagnosis for 116 of those patients with 83% accuracy, 86% sensitivity and 73% specificity (0.85 AUC).

Conclusions

While EpiScalp™ was trained on routine scalp EEG with a maximum duration of 1 hour, these results validate its use over an extended period of continuous 24-hour monitoring. These findings are significant because they show that EpiScalp™ can provide an accurate, quantitative metric to assist in diagnosing inconclusive EEG in the EMU within the first 24 hours, possibly removing the need for lengthy multi-day/week stays in the EMU and decreasing the financial burden on the healthcare system.

P48

Quantifying Evolving Brain Network Dynamics and Impairments in Infantile Epileptic Spasm Syndrome Using EEG Microstates and Source Localization

F. Gao[1], T. Jiang[1], W. Gao[2]

[1]Children's Hospital, Zhejiang University School Of Medicine - Hangzhou (China) - Hangzhou (China), [2]Cornell University - New York (United States)

Introduction

Infantile epileptic spasms syndrome (IESS) is a severe early-onset epilepsy frequently associated with profound neurodevelopmental impairment, yet the mechanisms by which seizures disrupt the developing brain remain poorly understood. Occurring during a critical period of synaptogenesis, epileptic spasms may impose a secondary insult on vulnerable neural networks, leading to lasting cognitive deficits. However, direct, time-resolved evidence of how brain networks dynamically transition from interictal to ictal states in IESS is lacking, and existing EEG microstate studies are largely limited to interictal, sensor-level analyses without cognitive linkage. To address these gaps, this study presents an integrated framework combining EEG microstate analysis with source localization to characterize ictal network dynamics, quantify seizure-related network injury, identify disruptions underlying visual and attentional deficits, and evaluate microstate parameters as potential biomarkers of brain injury in IESS.

Methods

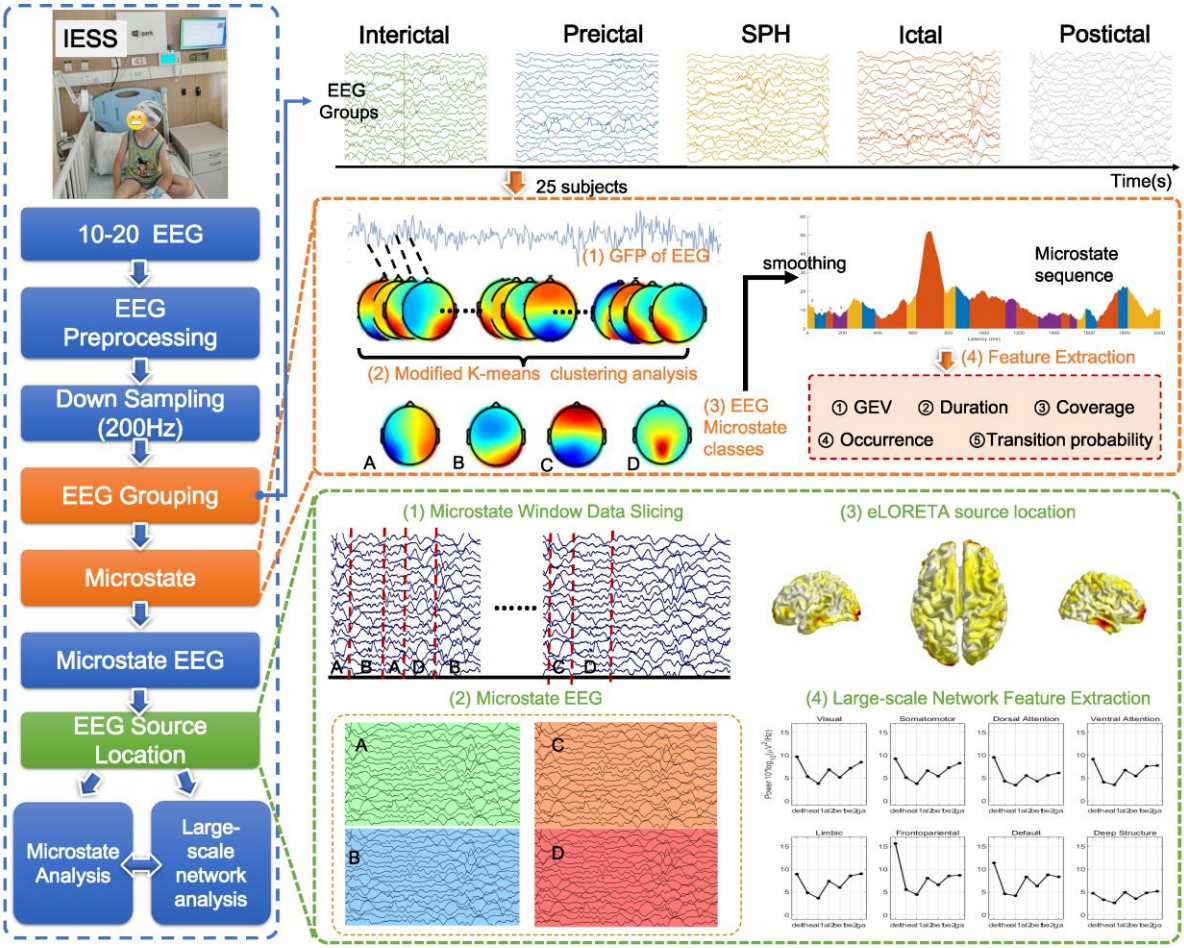
We put forward a dynamic network analysis framework that integrates EEG microstate analysis and source localization to quantify the changes in BFNs throughout infantile spasms and appraise their functional impact. Specifically: (1) Using K-means clustering, we identified four dominant microstate patterns (A–D) across five seizure phases: interictal, preictal, the Seizure Prediction Horizon (SPH), ictal, and postictal. Paired t-tests were employed to compare the dynamic microstate parameter changes between these phases, aiming to evaluate the effect of seizures on the global topological stability. (2) For each frequency band, we constructed microstate - specific signals per phase, mapped these signals to the cortex via source localization, and then utilized multifactorial ANOVA (Analysis of Variance) to assess the changes in the distribution of cortical current density. (3) After parcellating the cortex into eight functional networks, we calculated the mean current density per band per network and applied multifactorial ANOVA to identify the phase-specific impairments.

Results

Analysis of EEG data from 25 infants demonstrated that epileptic spasms significantly perturbed BFN homeostasis, particularly within core cognitive/motor networks (visual, default mode, attention, and sensorimotor networks), thereby impeding the typical progression of cerebral lateralization. Notably, alterations in power spectral density (PSD) within the alpha, beta, and gamma bands were identified as robust biomarkers of seizure-induced functional impairment.

Conclusions

These findings yield novel mechanistic insights into how infantile spasms disrupt brain function. The identified biomarkers provide critical guidance for optimizing clinical interventions aimed at mitigating cognitive and motor deficits in affected infants.



https://secure.key4events.com/onsite/api.aspx?e=1960&img=Fig1.png&ai=78698&op=getabstractimg&dirN=0

P49

Brain State Dynamics Underpinning Native and Task-Based Seizure Suppression Networks in Drug-Resistant Epilepsy

M. Khoshkhou[1], R.P. Lesser[2], S.V. Sarma[3], J.Y. Kang[4]

[1]Johns Hopkins University - Baltimore (United States), [2]Johns Hopkins University School Of Medicine - Baltimore (United States), [3]Johns Hopkins University School Of Medicine-Baltimore (united States) - Baltimore (United States), [4]Johns Hopkins University School Of Medicine-Baltimore (united States) - Baltimore (United States) - Baltimore (United States)

Introduction

Intracranial EEG (iEEG) monitoring during presurgical evaluation of patients with drug-resistant epilepsy (DRE) provides a rare opportunity to study seizure dynamics and termination at high spatiotemporal resolution. In this setting, clinicians observe both native, spontaneous seizures and induced seizure activity elicited during electrical mapping. Notably, cognitive tasks—such as mental arithmetic or spelling—can sometimes interrupt induced seizures in real time, successfully terminating nearly half of attempts. This phenomenon suggests the presence of latent network-level mechanisms capable of halting pathological activity, mechanisms not fully utilized by current electrical or pharmacological interventions. Although seizure initiation and propagation are relatively well characterized, the mechanisms that terminate seizures remain poorly understood, and it is unclear whether cognitive effort recruits the same network pathways used in native, spontaneous termination or why some suppression attempts succeed while others fail. The network reorganization that enables the brain to transition out of the ictal state has rarely been quantified. A deeper understanding of how the brain naturally terminates seizures—and how cognitive engagement can modulate these pathways—could reveal more effective network targets for stimulation therapy in DRE.

Methods

We studied 5 patients. For each patient, we estimated the dynamical network model $X(t+1) = A(t)X(t)$, where $X(t)$ represents the multichannel iEEG activity at time t and $A(t)$ is a square matrix describing how signals at each electrode influence one another. We computed $A(t)$ throughout the recording and used these matrices to derive the sink index (SI), a measure that reflects how strongly each node is driven by the rest of the network. At every time point, the collection of SI values across channels forms an SI vector that captures the network's instantaneous configuration. To identify recurring patterns in these configurations, we clustered all SI vectors using k-means, yielding a set of brain-state centroids that represent the dominant network configurations. Assigning each SI vector to its nearest centroid allowed us to track the brain's movement through these states with high temporal precision. We further examined the structure of these transitions using principal component analysis (PCA), which provided a low-dimensional trajectory illustrating how the network activity evolved across time during each recording. We then characterized features of the seizure suppression network by quantifying the amount and direction of incoming influence to each node. For every brain state, we extracted metrics such as polarity in sink–source space and the covariance of a fitted two-dimensional Gaussian distribution describing the spatial arrangement of channels. These features summarize how the network organizes itself

as it enters different brain states and offer interpretable markers for distinguishing suppression-related patterns.

Results

Across five patients, we analyzed multiple native seizures as well as induced seizures collected during mapping procedures. Induced seizures fell into two categories: those in which cognitive engagement (arithmetic or spelling tasks) appeared to halt the event, and those in which cognitive effort did not interrupt seizure progression. Across these patients, we examined nine native seizures and eleven induced seizures, seven of which were successfully interrupted by the cognitive task and four of which were not. Native seizures tended to progress through a small set of relatively stable ictal brain states and followed longer PCA trajectories, reflecting greater destabilization and broader propagation through established epileptic networks. In contrast, successful task-induced suppression produced the shortest trajectories, with the network moving toward a stable interictal state shortly after task initiation. These sessions often displayed brief fluctuations between ictal and interictal states before settling, suggesting that cognitive engagement recruits network pathways that support stabilization. Polarity analysis showed a consistent trend: polarity increased at seizure onset, indicating a shift from more balanced interactions to more asymmetric incoming and outgoing influences. Polarity remained elevated during the ictal period and decreased as the seizure approached termination. In failed suppression sessions, trajectories were longer, and instead of transitioning directly into an interictal configuration, the network shifted toward a postictal pattern before gradually returning to non-seizure levels. Overall, these observations suggest that brain-state transitions, trajectory length, and changes in polarity help differentiate patterns associated with native termination, successful suppression, and unsuccessful attempts.

Conclusions

These findings indicate that seizure termination is accompanied by recognizable changes in brain-state organization, network trajectories, and polarity patterns. Successfully suppressed sessions showed rapid stabilization into interictal, non-seizure states, whereas unsuccessful attempts evolved toward postictal activity before recovery. These distinctions highlight the network pathways that support seizure termination and may guide more effective, network-informed strategies for treating DRE. Describing these distinctions offers a clearer view of how stabilizing pathways emerge and supports ongoing efforts to develop network-informed approaches for seizure control in DRE.

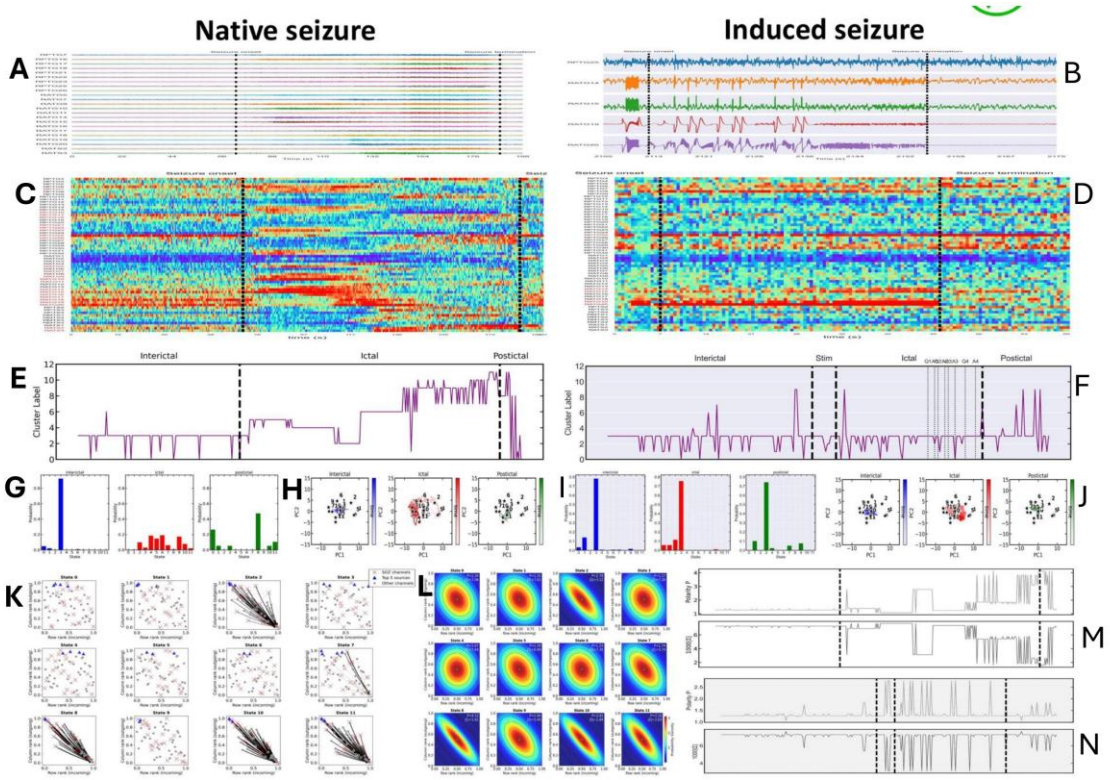


Figure 1. Native seizure dynamics contrasted with an induced seizure that was successfully terminated through cognitive engagement. (A–D) iEEG traces and sink index heat map aligned to ictal onset and termination. (E–F) Temporal evolution of the brain states. (G-J) State occupancy probability and principal component analysis (PCA) trajectories.(K) Seizure-suppression network for each brain state. (L) Two-dimensional Gaussian distribution of sink index rank-space. (M–N) Polarity time courses for native (M) and induced (N) seizures.

https://secure.key4events.com/onsite/api.aspx?e=1960&img=Figure_1.jpeg&ai=74149&op=getabstractimg&dirN=0

P50

Preoperative network integration predicts late postoperative seizure relapse in temporal lobe epilepsy: graph theory and machine learning analysis

V. Karpychev[1], R. Roth[1], K. Davis[2], D. Drane[1], A. Bagić[3], P. Dugan[4], H. Pardoe[4, 5], A. Parashos[6], N. Laxpati[7], R. Kuzniecky[8], L. Bonilha[9]

[1]Department Of Neurology, School Of Medicine, Emory University - Atlanta (United States), [2]Department Of Neurology, University Of Pennsylvania - Philadelphia (United States), [3]University Of Pittsburgh Comprehensive Epilepsy Center (upcec), Department Of Neurology, University Of Pittsburgh Medical Center (upmc) - Pittsburgh (United States), [4]Department Of Neurology, New York University - New York (United States), [5]Florey Institute Of Neuroscience And Mental Health - Melbourne (Australia), [6]Department Of Neurology, Medical University Of South Carolina - Charleston (United States), [7]Department Of Neurosurgery, School Of Medicine, Emory University And Children's Healthcare Of Atlanta - Atlanta (United States), [8]Department Of Neurology, Zucker School Of Medicine, Hofstra/northwell, Hempstead - New York (United States), [9]Department Of Neurology, School Of Medicine, University Of South Carolina - Columbia (United States)

Introduction

Temporal lobe epilepsy (TLE) is recognized as the most prevalent form of epilepsy (Téllez-Zenteno et al., 2012). Although 60-70% of patients with drug-resistant TLE achieve seizure freedom (SF) during the first year after surgery, the likelihood of remaining SF decreases to 56-67% over 10 years (Keleman et al., 2006). In contrast to short-term seizure recurrence, which often reflects inaccurate localization or incomplete resection, late seizure relapse can occur despite complete removal of the epileptic focus (Petrik et al., 2021). Moreover, its underpinning is assumed to reflect a stage of epileptogenesis associated with the generation of a new epileptic focus in the brain (Najm et al., 2013). Clinical and demographic predictors of late relapse are inconsistent, and few studies have examined MRI-based network biomarkers. Although the precise mechanisms of epileptogenesis remain unclear, persistent residual epileptic activity may eventually emerge clinically or even induce a new focus with independent onset (Thompson, 2023). Our study focuses on patients who achieved SF in the first year after surgery, but subsequently experienced seizure recurrence at least two years postoperatively. This strengthens the assumption that the epileptic focus was sufficiently removed surgically and enables investigation of how preoperative connectivity relates to later seizure development in the brain. We hypothesized that late relapsers would show greater preoperative integration with the broader brain network than long-term SF patients.

Methods

We prospectively collected preoperative resting-state fMRI and diffusion-weighted MRI (dMRI) from patients at six epilepsy centers diagnosed with drug-resistant unilateral TLE before they underwent surgery and followed their outcomes over time (n = 175). The inclusion criterion for patients in this analysis was ILAE = 1 in the first year after surgery (n = 107). Patients were then stratified by long-term maintenance of SF (SF-SF; ILAE = 1 at least two years after surgery) versus late seizure relapse (SF-NSF; ILAE = 2–6 at least two years after surgery). We only included patients who had not undergone a subsequent surgical procedure after relapse. We

obtained the same imaging protocol in a large multi-center cohort of healthy controls (n = 552). Figure 1 displays an overview of processing steps. For both imaging modalities, we computed hubness metrics: betweenness centrality (BC), eigenvector centrality (EC), and participation coefficient (PC) across 374 ROIs (combining cortical Glasser and subcortical Desikan-Killiany atlases). As counting PC relies on a split into modular communities of the ROIs, we individually divided them into eight communities: seven resting-state functional networks (Yeo et al., 2011) and a community of the ROIs from the delineated resection. We considered an ROI 'resected' if at least 25% of its volume overlapped the normalized resection mask. For each patient, we z-scored each hubness metric of an ROI using the normative distribution of ROI-level hubness from healthy controls.

First, we verified whether the hubness metrics of resected ROIs differed between the groups. For each metric (BC, EC, PC) within each modality, we built a logistic regression with seizure outcomes (SF-SF vs. SF-NSF) as the dependent variable and the mean value of the metric across resected ROIs as an independent variable. We additionally conducted 5000 iterations: (i) shuffling outcome labels while preserving covariates and hubness metrics, and (ii) generating size-matched null resections per patient (same number of ROIs), refitting the full adjusted model each iteration.

Results

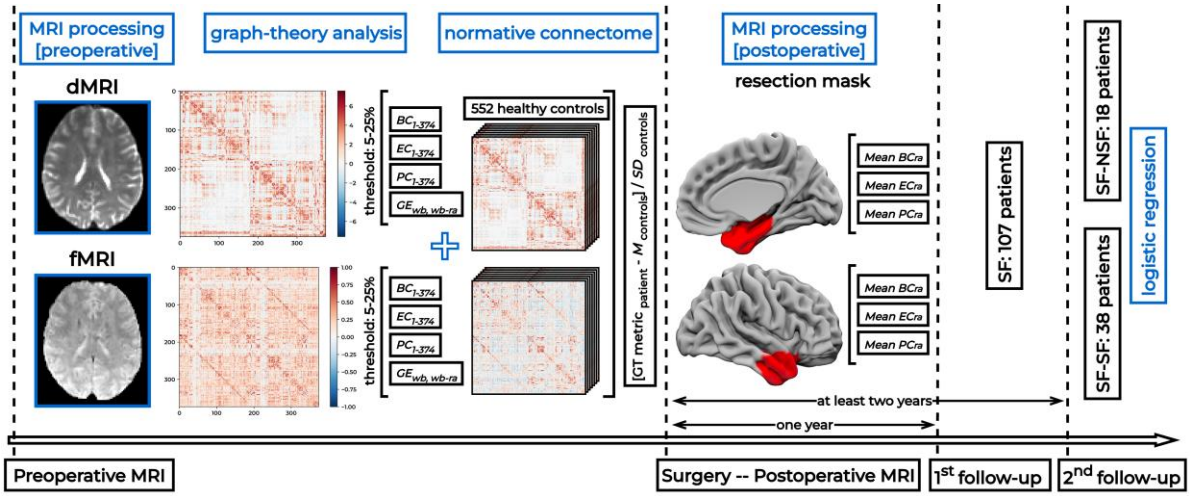
In total, 56 patients met our strict inclusion criteria for analysis, of which 38 sustained long term seizure freedom (SF-SF), and 18 were late relapsers (SF-NSF). For each hubness metric, we fit logistic regressions with seizure outcomes as the dependent variable and the metric as an independent variable (average values across the surgically-removed ROIs). We did not find significant associations between BC and EC in either modality. Greater mean fMRI-based PC of the resected ROIs was significantly associated with SF-NSF outcome (Est = 1.9, SE = 0.78, $t(21) = 2.4$, $p = 0.015$). Both permutation procedures – (1) random shuffling of outcome labels (SF-SF, SF-NSF) among the patients and (2) random shuffling of the same number of ROIs as in the real resection – confirmed the association with $Pperm = 0.002$ and $Pperm = 0.035$, respectively. Figure 2 shows the spatial distribution of fMRI-based z-scored PC of the resected ROIs.

We post hoc trained an Elastic Net Logistic Regression model within a nested 4-fold cross-validation framework to classify ROIs between the groups based on fMRI-based PC. The model also included age of onset, epilepsy duration, postsurgical duration, lesions, as well as gray-, white-matter, and total lesion volumes, and were each evaluated using 100 iterations, with AUC as the performance metric. To avoid overfitting, feature selection involved LASSO regression on training folds PC of the resection based on fMRI predicted late seizure relapse with a mean AUC of 0.80 (SD = 0.07) and significantly outperformed a null model ($t(198) = 41.6$, $Pperm < 0.001$).

Conclusions

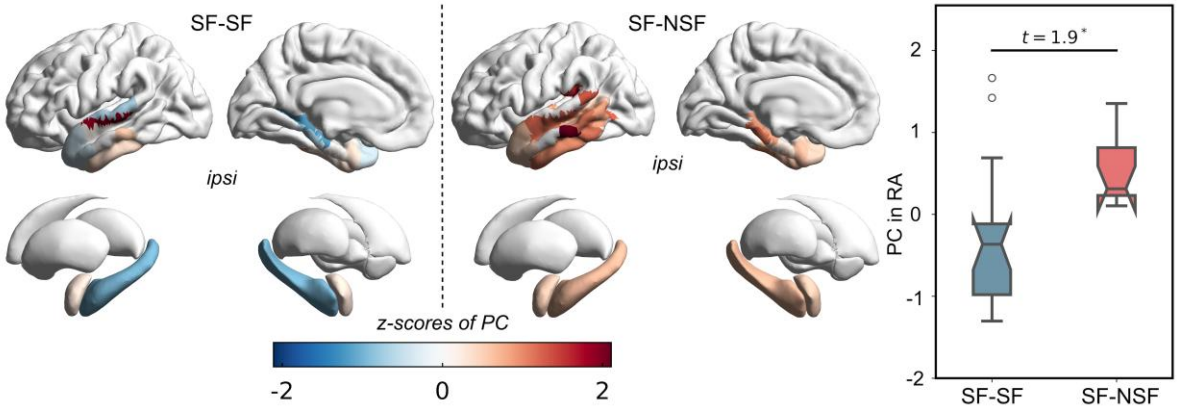
Our findings indicate that preoperative brain network integration of the epileptogenic focus contributes to the risk of late seizure relapse in TLE. Specifically, we showed that patients whose epileptic focus was more integrated with the remaining brain network, as reflected by higher participant coefficient values, were more likely to experience late seizure relapse after initial postoperative SF in the first year. This supports the view that highly integrated foci may facilitate secondary focus formation over time. Our results suggest that long term outcomes depend not only on accurate focus delineation but also on the focus network integration.

Overview of MRI processing



https://secure.key4events.com/onsite/api.aspx?e=1960&img=Figure+1.png&ai=74168&op=getabstractimg&dirN=0

Spatial distribution of fMRI-based z-scored PC of



https://secure.key4events.com/onsite/api.aspx?e=1960&img=Figure+2.png&ai=74168&op=getabstractimg&dirN=1

P51
An Implantable Device that Converses with Patients and Learns to Co-Manage Epilepsy

Z. Goldblum[1], S. Shi[1], Z. Xu[1], W.K.S. Ojemann[1], C.A. Aguila[1], K. Xie[1], K.C. Nix[1], K. Walsh[1], E. Chang[1], K.A. Davis[1], N. Sinha[1], L.H. Hammer[1], E.C. Conrad[1], B. Litt[1]

[1]University Of Pennsylvania - Philadelphia (United States)

Introduction

Approximately one-third of the 70 million people living with epilepsy suffer from medication-resistant seizures. While implantable devices offer a promising alternative, current technologies are limited by their inability to share information with patients in real-time or adapt rapidly to individual needs. Existing devices operate largely as "black boxes," missing the opportunity to leverage patient insights for algorithm training or to inform patients of immediate risks. We propose a novel paradigm: a device platform that utilizes large language models (LLMs) to converse freely with patients, interpreting physiological data to co-manage disease.

Methods

We developed a platform linking real-time scalp and intracranial EEG (iEEG) to a HIPAA-compliant cloud environment and a patient-facing smartphone application. The system integrates a secure AI agent (powered by GPT-4o-mini) that ingests biomarkers, including seizure probability, sleep staging, interictal spikes, and medication load, to communicate bidirectionally with patients. To ensure patient safety, we implemented a multi-tiered framework comprising strict AI guardrails and automated content moderation APIs, with a dashboard for real-time oversight of patient-AI interactions. We recruited 13 adult patients undergoing inpatient video-EEG monitoring to test this platform. The AI agent proactively messaged patients regarding clinical events and trends, while patients could query their own neurophysiological data and provide text-based annotations of symptoms. These patient annotations were used to fine-tune seizure detection algorithms (SPaRCNet and EEGWaveNet) up to every three hours, and supplemented a clinician-in-the-loop review framework.

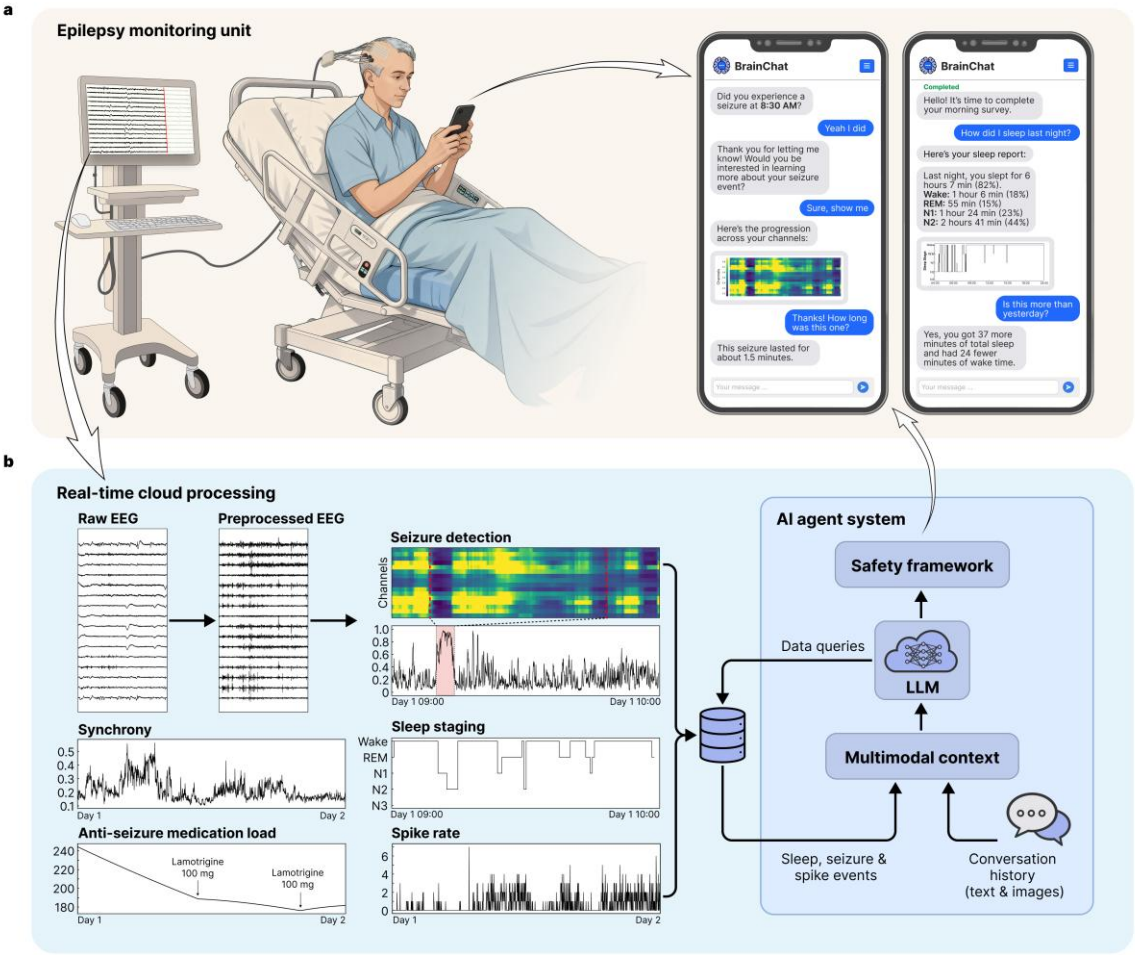
Results

Patients demonstrated high engagement, exchanging a collective total of 1,307 messages with the system over a median duration of 4.8 days. Use of the platform was intuitive, achieving a mean System Usability Scale score of 83/100 ("excellent"). From a safety perspective, expert review of conversation logs identified no instances of harmful content or inappropriate medical advice, confirming the robustness of the LLM guardrails. The system successfully facilitated real-time co-management; for example, seizure notifications elicited patient response rates of 60% within 30 minutes. Crucially, fine-tuning of seizure detection models using patient feedback significantly improved performance. Compared to baseline, this patient-in-the-loop approach reduced the false alarm rate by 82.4% ($p=0.002$) and increased the F1 score by 250% ($p=0.016$) without expert intervention.

Conclusions

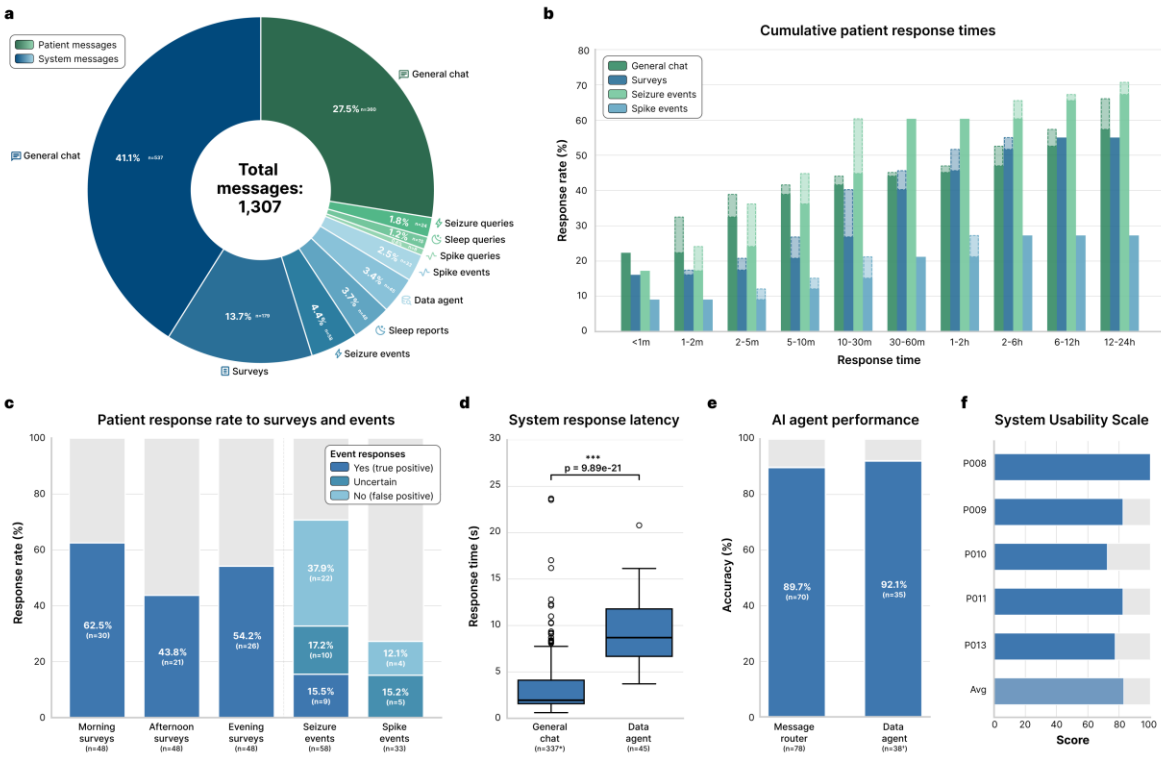
This study demonstrates the first proof-of-principle for a medical device that utilizes generative AI to converse with patients and improve its own performance through real-time interaction. By closing the loop between physiological data and patient experience, this platform significantly enhances seizure detection precision and patient engagement. This approach represents a scalable model for next-generation intelligent devices capable of adapting to patient needs and co-managing epilepsy and other chronic neurological conditions.

Overview of the Patient-AI interface.



https://secure.key4events.com/onsite/api.aspx?e=1960&img=fig1.png&ai=78613&op=getabstractimg&dirN=0

Patient engagement and AI technical performance.



https://secure.key4events.com/onsite/api.aspx?e=1960&img=fig2.png&ai=78613&op=getabstractimg&dirN=1

P52

Automated phenotyping of epilepsy and seizure types with pre-trained language models

E. Chang[1], K. Xie[1], D. Zhou[1], J. Korzun[1], B. Litt[1], C. Ellis[1]

[1]University Of Pennsylvania - Philadelphia (United States)

Introduction

Epilepsy affects more than 60 million people worldwide and encompasses a heterogeneous set of disorders with wide variation in etiology, seizure semiology, and treatment response. Accurate classification of epilepsy type and seizure type is essential for guiding therapy, counseling patients, and identifying candidates for surgical or device-based interventions. Misclassification can delay appropriate care and has direct safety implications, as generalized tonic-clonic seizures are the primary risk factor for Sudden Unexpected Death in Epilepsy (SUDEP). However, clinically meaningful phenotypic detail is poorly captured in structured electronic health record (EHR) fields and instead resides in unstructured clinical documentation, limiting the scale and generalizability of epilepsy research. Transformer-based natural language processing (NLP) models offer a promising solution by enabling automated extraction of complex, context-dependent clinical information from free text. While prior work has demonstrated the feasibility of using NLP to extract selected epilepsy outcomes, systematic mapping of fine-grained epilepsy syndromes and seizure types at health-system scale remains limited. In this study, we develop and evaluate automated models for epilepsy and seizure type phenotyping from unstructured clinical notes, benchmarked against expert epileptologist annotations, and deploy the best-performing model to generate a large-scale corpus of epilepsy phenotypes across an academic health system.

Methods

A subset of 309 epilepsy progress notes was expert-annotated by three board-certified neurologists for epilepsy type and seizure type using a granular (“all-way”) labeling scheme, with each note labeled by two annotators and disagreements adjudicated by consensus. Fine-grained labels were also mapped to binary, ternary, and six-way classification schemes for evaluation. Epilepsy type was treated as a single-label task, while seizure type was evaluated as single-label or multi-label depending on the scheme.

We developed and evaluated two transformer-based NLP approaches: a fine-tuned masked language model (BERT) and an autoregressive large language model (DeepSeek-R1) using zero-shot and few-shot prompting. BERT underwent domain adaptation on unannotated notes followed by supervised fine-tuning, with performance assessed via cross-validation. DeepSeek-R1 prompts were iteratively refined through error analysis and applied without model fine-tuning. Predictions from multiple runs were ensembled for robustness. Model performance was benchmarked against adjudicated expert labels using Matthews correlation coefficient (single-label tasks) and F1 score (multi-label task), with comparisons to inter-rater agreement and uncertainty estimated via bootstrap resampling.

The best-performing model was deployed across nearly 80,000 clinical notes from over 18,500 patients. Visit-level predictions were aggregated into patient-level epilepsy classifications using a majority-vote-based adjudication strategy. Longitudinal transitions between epilepsy types were

modeled using a continuous-time hidden Markov model to account for irregular follow-up. Extracted phenotypes were integrated with previously derived seizure frequency and seizure-freedom trajectories to assess differences in clinical outcomes across epilepsy types.

Results

Inter-rater agreement was higher for epilepsy type than seizure type classification (Cohen's κ = 0.77 for ternary epilepsy vs. κ = 0.49 and 0.40 for binary and ternary seizure type), reflecting greater consistency in documentation of epilepsy syndromes than seizure semiology. Across classification tasks, the large language model DeepSeek-R1 achieved performance comparable to or exceeding expert inter-rater agreement, particularly for ternary epilepsy classification (DeepSeek MCC [95% CI] = 0.85 [0.80-0.90] vs. human pairwise MCC = 0.77 [0.70-0.83]) and for binary, ternary, and multi-label seizure type classification (Figure 1). For multi-label all-way seizure phenotyping, DeepSeek achieved an F1 score of 0.72 [0.69-0.79], substantially exceeding expert agreement (F1 = 0.52 [0.31-0.48]). Additionally, DeepSeek consistently outperformed BERT on more granular classification tasks, while BERT demonstrated acceptable performance only on simpler binary and ternary tasks.

The best-performing DeepSeek configuration was deployed to 77,049 clinical notes from 18,566 patients, generating longitudinal epilepsy and seizure phenotypes at scale. Longitudinal modeling revealed high diagnostic stability once specific epilepsy syndromes were assigned, with greater transitions from initially unclassified or unspecified focal states. Seizure types frequently co-occurred within patients, particularly tonic-clonic, focal impaired-awareness, and focal aware seizures. Integration with seizure outcome measures showed similar maximum seizure-free durations between focal and generalized epilepsies (mean ~24 vs. 26 months), but higher tonic-clonic seizure burden in generalized epilepsy (83% vs. 57% of seizure-positive visits) and modestly higher monthly seizure frequencies (mean 8.6 vs. 7.0 seizures/month). Patients with Lennox-Gastaut syndrome exhibited substantially higher seizure burdens than other epilepsy subtypes, consistent with known clinical severity.

Conclusions

This study demonstrates that pre-trained language models can accurately phenotype epilepsy and seizure types from unstructured clinical notes at scale with performance comparable to expert epileptologists. In contrast to fine-tuned masked language models, autoregressive LLMs showed superior robustness for granular, ambiguous, and multi-label phenotypes, underscoring their advantage in settings where clinical documentation is heterogeneous and seizure semiology is inconsistently described. These findings reflect both the inherent variability in seizure documentation and the ability of LLMs to maintain internal consistency without being constrained by annotation noise.

Applying these models at health-system scale enabled longitudinal analyses that align with known clinical patterns of diagnostic evolution, seizure-type overlap, and syndrome-specific seizure burden. By unlocking fine-grained phenotypic information from unstructured clinical text, this approach enables population-level studies of epilepsy trajectories that were previously infeasible without manual abstraction. These methods support downstream applications including longitudinal SUDEP risk stratification, earlier identification of candidates for surgical or device-based therapies, automated cohort construction for treatment-effectiveness studies, and scalable clinical registry development. Overall, this work establishes a foundation for automated, expert-level epilepsy phenotyping in real-world EHR data, enabling more precise and longitudinally informed epilepsy research and care.

P53
Ictal Network Characteristics and Treatment Response in Epilepsy

K. Walsh[1], K. Knix[2], B. Scheid[1], W. Ojemann[1], E. Conrad[1, 3], N. Sinha[1, 3], B. Litt[1, 3], K. Davis[2, 3]

[1]Department Of Bioengineering, University Of Pennsylvania - Philadelphia (United States),
[2]Department Of Neuroscience, University Of Pennsylvania - Philadelphia (United States),
[3]Department Of Neurology, University Of Pennsylvania, Perelman School Of Medicine - Philadelphia (United States)

Introduction

Epilepsy affects over 70 million people worldwide, with approximately one-third of patients experiencing drug-resistant epilepsy (DRE) which does not respond to anti-seizure medication. Responsive neurostimulation (RNS) is an alternative therapy for patients who are not amenable to resection. RNS devices are surgically implanted in the brain to continuously monitor brain activity and deliver targeted electrical stimulation to suppress detected seizures. However, treatment response remains highly variable across individuals. While some patients experience a substantial reduction in seizures, nearly 30% show no meaningful improvement after three years. The underlying mechanisms of RNS efficacy remain poorly understood, limiting efforts to optimize therapy and improve patient selection.

Emerging evidence suggests that RNS may alter brain network dynamics over time, leading to long-term seizure control through indirect and progressive modulation of the seizure network. Given the invasive nature of RNS implantation and variable treatment response, identifying reliable pre-surgical predictors of efficacy is critical for appropriate patient selection. Retrospective analyses indicate that pre-implantation intracranial EEG (iEEG) and magnetoencephalography (MEG) network features may be associated with RNS response. We have collected iEEG recordings from 27 patients who subsequently received RNS implants, and we find that network features extracted from seizure periods improve the prediction of patient therapy response in our dataset as compared to clinical features alone.

Methods

We conducted a retrospective study of 27 patients from two epilepsy centers (UCSF, HUP) who underwent phase 2 monitoring with iEEG electrodes as part of standard clinical practice. All included patients were implanted with the NeuroPace RNS System following iEEG monitoring. Seizure onset and termination times were identified by a board-certified clinician for each patient. Ictal recordings were downsampled to 500 or 512 Hz, notch-filtered to remove line noise (60 Hz and 120/180 Hz harmonics), and common average referenced. Channels with substantial artifacts were excluded from analysis.

Time-varying functional connectivity networks were constructed from consecutive 1-second non-overlapping windows spanning the duration of each seizure event. Networks were represented as square adjacency matrices with multitaper coherence as edge weights, computed separately for multiple frequency bands (alpha, beta, low gamma, high gamma, and broadband). Global network

metrics (synchronizability, node strength dispersion, global efficiency, transitivity) and local network metrics (node strength, eigenvector centrality) were then calculated from each consecutive functional connectivity matrix to characterize temporal evolution of network topology during seizures. All ictal features were averaged across channels, time windows, and seizure clips for each patient.

To assess whether ictal network features from iEEG improve the prediction of future response to RNS therapy, we trained logistic regression models using three feature sets: 1) clinical features only (age at seizure onset, MRI findings, seizure lateralization, RNS lead localization); 2) ictal functional connectivity values; and 3) ictal network metrics. Response status was the binary outcome variable, with responders defined as patients achieving ≥50% reduction in seizure frequency compared to their pre-implant baseline. To evaluate model performance, we performed leave-one-out cross-validation (LOOCV), where each model was trained on n-1 patients and tested on the held-out patient. This procedure was repeated for all patients, and model performance was assessed using accuracy, AUROC, AUPRC, and maximum F1-score. All models used L2 regularization with balanced class weights to account for class imbalance. Feature importance was assessed using bootstrap confidence intervals (1000 stratified resamples) for regression coefficients.

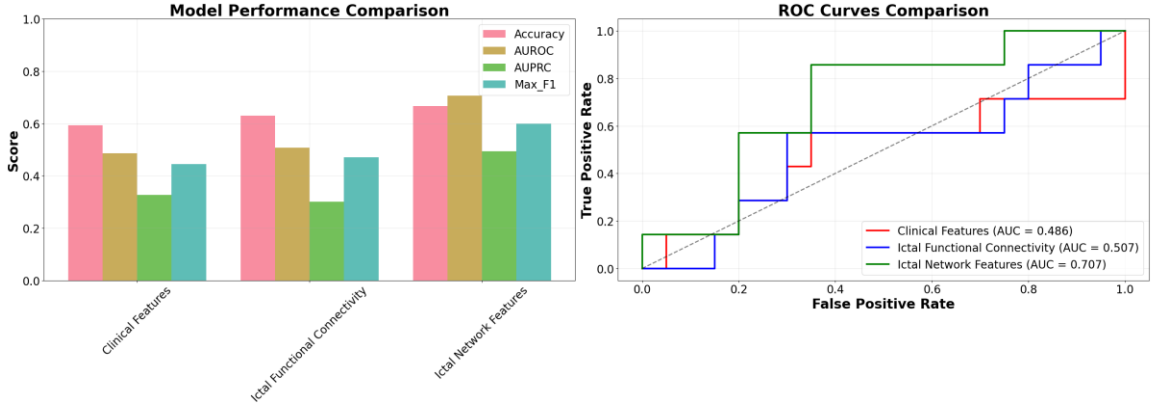
Results

Clinical features showed no significant associations with treatment response in univariate analyses (age at seizure onset: U=88, p=1.0; MRI findings: χ^2=4.64, p=0.20; seizure lateralization: χ^2=0.68, p=0.71; lead localization: χ^2=5.43, p=0.14). In predictive modeling with leave-one-out cross-validation, clinical features (AUROC=0.49) and ictal functional connectivity (AUROC=0.51) showed performance at chance level. Ictal network features demonstrated substantially better discrimination (AUROC=0.71, accuracy=67%, AUPRC=0.49) with 82% precision and 70% recall for responders. The five features with largest absolute coefficients were: (1) alpha synchronizability (β=0.76, 95% CI [0.15, 1.31]), (2) broadband synchronizability (β=0.59, 95% CI [-0.06, 1.13]), (3) high gamma global efficiency (β=0.48, 95% CI [-0.20, 1.22]), (4) alpha node strength dispersion (β=-0.44, 95% CI [-0.70, -0.17]), and (5) high gamma transitivity (β=0.43, 95% CI [-0.04, 0.88]).

Conclusions

Ictal network features from intracranial EEG improve prediction of RNS therapy response compared to clinical features alone, suggesting that seizure network architecture may inform treatment selection. Future work should validate these findings in larger cohorts and investigate whether network-based biomarkers can prospectively guide patient selection for responsive neurostimulation therapy.

Model Performance Comparison



 https://secure.key4events.com/onsite/api.aspx?e=1960&img=Abstract_fig1.png&ai=74276&op=getabstractimg&dirN=0

P54
Long-term EEG monitoring interpretation using artificial intelligence

D. Mansilla[1], J. Tveit[2], H. Aurlien[2], T. Avigdor[3, 4], C. Abdallah[4], P. Bosque-Varela[5, 6], M. Cacic Hribljan[7], V. Ros-Castelló[8], D. Schomer[9], J. Gotman[10], S. Beniczky[11, 12], B. Frauscher[3, 13]

[1]Neurophysiology Unit, Neurosurgery Institute Dr. Asenjo - Santiago (Chile), [2]Holberg Eeg - Bergen (Norway), [3]Department Of Neurology, Duke University Medical Center - Durham (United States), [4]Department Of Neurology And Neurosurgery, Montreal Neurological Institute And Hospital - Montreal (Canada), [5]Neuroscience Institute, Christian Doppler University Hospital, Centre For Cognitive Neuroscience - Salzburg (Austria), [6]Department Of Neurology, Neurocritical Care And Neurorehabilitation, Christian Doppler University Hospital, Member Of The European Reference Network Epicare, And Centre For Cognitive Neuroscience, Paracelsus Medical University - Salzburg (Austria), [7]Children's Hospital Srebrnjak, Zagreb And School Of Medicine, University Of Zagreb - Zagreb (Croatia), [8]Unidad De Epilepsia, Servicio De Neurología, Hospital De La Santa Creu I Sant Pau - Barcelona (Spain), [9]Department Of Neurology, Beth Israel Deaconess Medical Center, Harvard Medical School - Boston (United States), [10]Montreal Neurological Institute And Hospital, Mcgill University - Montreal (Canada), [11]Department Of Clinical Neurophysiology, Aarhus University Hospital - Aarhus (Denmark), [12]Department Of Clinical Neurophysiology, Danish Epilepsy Centre - Dianalund (Denmark), [13]Department Of Biomedical Engineering, Duke Pratt School Of Engineering - Durham (United States)

Introduction

Artificial intelligence (AI) models have demonstrated diagnostic accuracy comparable to that of human experts when assessing routine electroencephalograms (EEGs). The review and report of long-term EEG monitoring (LTM) is a time-consuming task, often requiring the expert to perform a visual analysis of many hours of EEG recordings. The automated interpretation of LTMs using AI holds the potential to reduce this workload significantly. We evaluated whether SCORE-AI/autoSCORE, an automated deep-learning tool that classifies routine EEGs as normal or into specific abnormal categories, achieves diagnostic accuracy comparable to expert readers for LTM studies. We further assessed whether SCORE-AI supervised by experts decreases the time required to interpret LTMs compared with conventional visual analysis, as currently done in clinical practice.

Methods

We assessed the diagnostic accuracy of a “fixed-and-frozen” AI model that has been approved for medical use (SCORE-AI/autoSCORE), using an independent single-center dataset and external gold standard, obtained from epilepsy monitoring unit discharge diagnosis of the same admission the LTM was registered. We benchmarked it against two panels of three experts each, blinded to all other data. Panel 1 reviewed the LTMs based on the SCORE-AI detections only (supervised SCORE-AI approach), and Panel 2 reviewed the raw recordings without SCORE-AI assistance (conventional approach). The dataset comprised 52 normal and 26 abnormal 25-electrode LTMs, which included four major classes of EEG abnormalities (focal epileptiform, generalized epileptiform, focal non-epileptiform, and diffuse non-epileptiform). Experts in Panel 1 were tasked with reviewing up to the fifteen SCORE-AI detections with the highest predicted probability within

each category; in case the LTM had no abnormality detections the experts did not need to review the recording. Experts in Panel 2 evaluated each LTM until they reached a diagnostic conclusion. For LTMs classified as normal, this entailed reviewing the full duration of the recording. Whenever experts disagreed on one or more LTM categories, consensus was defined as agreement between any two of the three experts. When multiple abnormalities co-occurred in the same LTM, we applied rules to avoid penalizing correct multi-label detections. If the gold standard indicated generalized epilepsy and any of the expert consensuses or the unsupervised SCORE-AI correctly identified generalized epileptiform activity, any additional labels (diffuse non-epileptiform or focal epileptiform) were counted as true positives. The same applied to focal epilepsy: once focal epileptiform activity was correctly detected, extra diffuse or focal non-epileptiform labels were also treated as true positives. This adjustment ensured that evaluations reflected the real-world scenarios in which overlapping features are common. To assess diagnostic performance, we computed sensitivity, specificity and overall accuracy of the AI model and the two expert panels against the external gold standard, and were accompanied by 95% confidence intervals (CI) derived using a non-parametric bootstrap procedure. Differences in performance metrics between groups were assessed using the Mann–Whitney test. We determined the interrater agreement among the three supervised SCORE-AI experts, and among the three conventional approach experts using Gwet's agreement coefficient (Gwet's AC), and we calculated 95% CIs using the bootstrap method. The time used to review the LTMs using both approaches (time to diagnostic decision) was also assessed. Differences in review times performance measures were also evaluated using the Mann–Whitney test.

Results

We analyzed 12 to 33-hour LTMs from 78 patients (53 females; median age 33.7 [range: 16-77.3] years). When comparing the interrater agreement between the two expert panels for the supervised SCORE-AI and conventional approach, we found that the supervised SCORE-AI approach yielded a significantly higher Gwet's AC across all categories ($p<0.001$), with the greatest improvement in agreement on identifying normal LTMs (supervised SCORE-AI Gwet's AC 0.83 [95% CI 0.71 – 0.93], conventional approach Gwet's AC 0.39 [95% CI 0.23-0.54], $p<0.001$). Supervised SCORE-AI achieved expert-level performance in LTM interpretation, showing no significant difference in sensitivity, specificity or accuracy for either the normal category (supervised SCORE-AI: sensitivity 0.75 [95% CI 0.64 – 0.86], specificity 0.96 [95% CI 0.87 - 1.00], accuracy 0.80 [95% CI 0.71 - 0.88]; conventional approach: sensitivity 0.78 [95% CI 0.67 - 0.88], specificity 0.92 [95% CI 0.81 -1.00], accuracy 0.82 [95% CI 0.74 - 0.89]; p-values >0.05) or any abnormal categories except for specificity and accuracy for the diffuse non-epileptiform category (overall abnormal classification: supervised SCORE-AI: sensitivity 0.96 [95% CI 0.87 – 1.00], specificity 0.75 [95% CI 0.64 – 0.86], accuracy 0.82 [95% CI 0.73 – 0.89]; conventional approach: sensitivity 0.92 [95% CI 0.81– 1.00], specificity 0.78 [95% CI 0.67 – 0.88], accuracy 0.82 [95% CI 0.74 – 0.91]; p-values >0.05). The performance of the unsupervised SCORE-AI approach had lower specificity in all abnormal categories (all p-values <0.001), as well as a lower sensitivity for the normal category compared to the supervised SCORE-AI and the conventional approach (unsupervised SCORE-AI sensitivity 0.46 [95% CI 0.32-0.59], vs. supervised SCORE-AI sensitivity 0.75 [95% CI 0.64-0.86], $p<0.001$, and vs. conventional approach sensitivity 0.78 [96% CI 0.67 – 0.88], $p<0.001$). The median time required for review (time to diagnostic decision) was 95.8% shorter when using the supervised SCORE-AI approach vs the conventional approach (supervised SCORE-AI approach 2.2 min [0.18 – 13.65], conventional approach 52.39 min [1.48 - 378], $p<0.001$).

Conclusions

Supervised SCORE-AI matched expert-level diagnostic accuracy for LTM interpretation. It also increased the expert interrater agreement while significantly lowering the median time spent

reviewing. Integration of SCORE-AI/autoSCORE in the interpretation of LTMs is promising to decrease the workload and improve efficiency and consistency in EEG interpretation.

P55

Prediction of Drug-resistance at Epilepsy Onset: An AI-driven, EEG-Informed Tool for Early Clinical Decision Support in Pediatric Epilepsy

N. Makaram[1], G. Sabino[1], F. Granados Monge[1], E. Tamilia[1]

[1]Boston Children's Hospital - Boston (United States)

Introduction

Early, objective identification of children that will develop for drug-resistant epilepsy (DRE) remains a critical unmet need. DRE develops in ~30% of patients with epilepsy, at even higher rates in children, and carries severe consequences including impaired quality of life, psychosocial comorbidities, financial burden, and elevated mortality. Despite this alarming prognosis, diagnostic and referral delays remain pervasive, extending the “DRE treatment odyssey” for many years. In children, timely DRE recognition is crucial: delays incur cognitive and developmental loss, reduce surgical success rates, and increase mortality. Because EEG is routinely obtained at epilepsy onset and captures core electrophysiologic features, it provides an objective, immediately available data source that, when combined with clinical phenotype, can support data-driven prognostication of disease trajectory.

This study aimed to validate a **scalable, pediatric-specific AI tool that uses brief routine EEGs integrated with clinical phenotype to estimate individualized DRE risk at epilepsy onset**, addressing the gap in early, objective triage for high-risk children.

Methods

We identified children with epilepsy who had EEG at Boston Children’s Hospital prior to starting anti-seizure medications. Each patient was classified as DRE or non-DRE following the consensus proposal by the *ILAE Task Force*. We retrieved artifact-free sleep EEG (19-24 channels) from each patient and computed five frequency-specific connectivity matrices (method: amplitude envelope correlation, AEC; Fig 1A). From these, we extracted quantitative network metrics (mean, variance, skewness, and kurtosis of connection strength, as well as modularity and efficiency) alongside power distribution (Fig. 1A). We also collected a comprehensive list of clinical variables that were previously linked to DRE risk: etiology, seizure type, MRI findings, baseline seizure burden, age at onset, developmental delay, and history of status epilepticus or febrile seizures.

We used *Chi-square* or *Wilcoxon rank-sum* test to assess the association of each clinical variable with DRE/non-DRE diagnosis. Two-way ANOVA was used to test group differences (DRE vs non-DRE) in EEG metrics across major clinical subgroups.

We then validated machine-learning (ML) algorithms with a 3-fold cross-validation scheme using either clinical-only input features (*clinical model*) or clinical+EEG features (*clinical+EEG model*, Figure 1B). All features entered a *genetic* feature selection algorithm that retained up to seven predictors per model; this upper limit was set to ensure at least 8 observations (patients) per predictor in each fold and thus minimize the risk of overfitting. Across multiple ML algorithms (Figure 1B) we identified the best performing one by assessing balanced accuracy and area-

under-the-ROC-Curve (AUC). To test generalizability, we also tested whether the ML models retained high AUC within major clinical subgroups.

Results

We included 94 drug-naïve children with epilepsy (35% DRE; mean age: 8 ± 5.4 years). From the analysis of the clinical variables (figure 2A, top bar charts), we observed that DRE children more often had symptomatic etiology ($p<0.001$), focal seizures ($p=0.038$), epileptogenic MRI ($p<0.001$), high seizure burden ($p<0.0001$), and younger age ($p=0.014$). When comparing EEG properties between DRE and non-DRE across clinical strata (2-way ANOVA, Figure 2A bottom), we found that lower network modularity, efficiency, and variability characterized DRE patients consistently across subgroups (symptomatic vs non-symptomatic etiology; focal vs non-focal seizures; high vs low seizure burden).

Across multiple ML algorithms, we report results from the KNN as it provided the highest performance. Figure 2B shows how **the integrated clinical+EEG model achieved AUC=0.85 and 81% accuracy** ($p<0.0001$, in orange) in predicting DRE at epilepsy onset, **outdoing the clinical-only model** (AUC=0.71, 75% accuracy, $p<0.0001$, in grey). Notably, EEG network efficiency and modularity were key predictors of later drug resistance, supporting our network-centric approach. Finally, **when tested across clinical subgroups, the integrated model retained high AUCs (0.83-0.86) within groups** (split by etiology, seizure type and burden), **while the clinical-only model dropped to 0.65-0.76 in several subgroups**. Figure 2B (bottom) shows how predicted probabilities from the integrated model tracked true outcomes both in patients with focal and non-focal seizures, unlike the clinical model that underperformed in non-focal seizure patients.

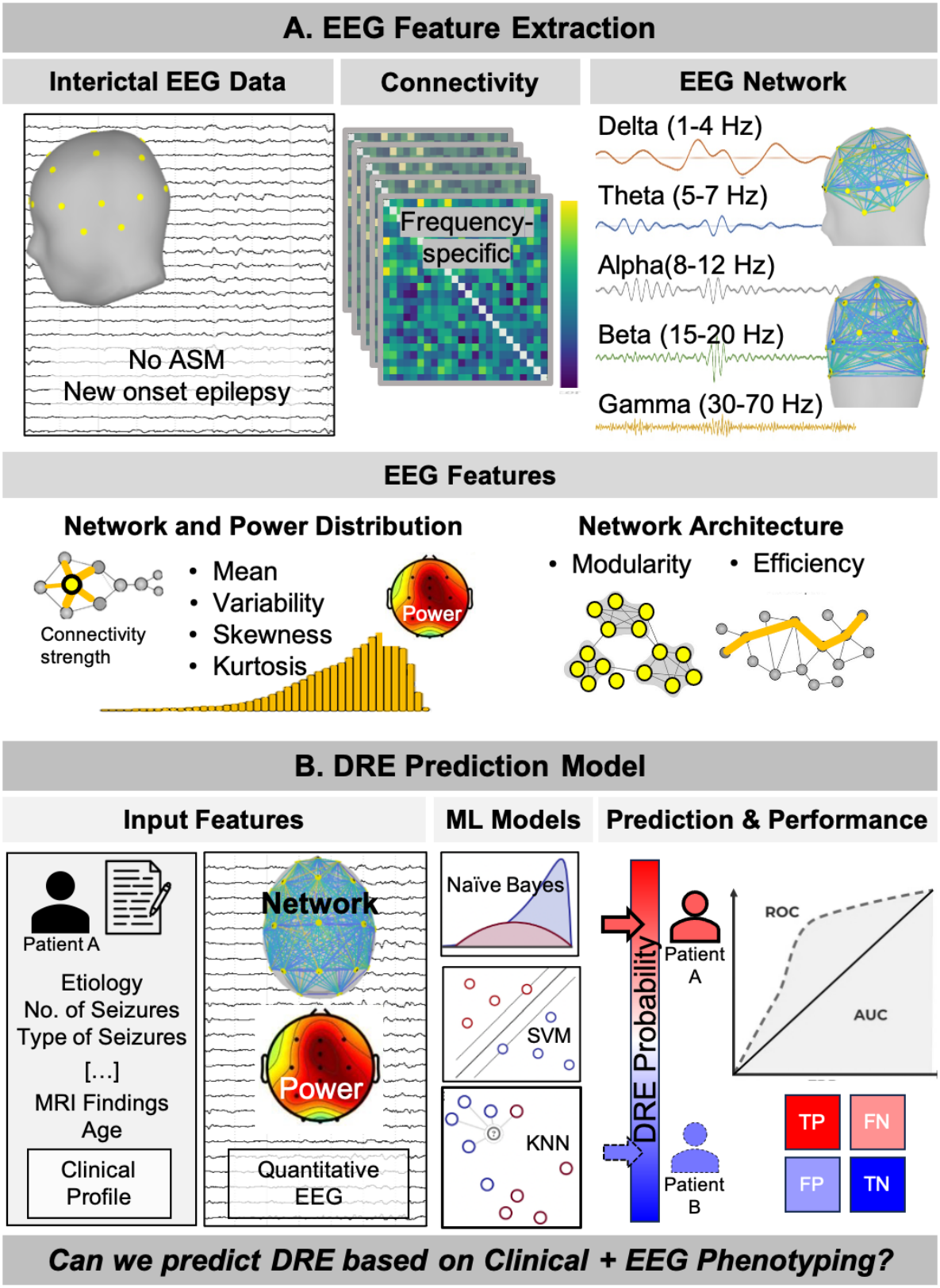
Conclusions

We present the first validation of an EEG-informed AI tool for individualized DRE prediction in children that leverages quantitative EEG and clinical data available at epilepsy onset.

Our findings pinpoint early robust EEG signatures of pediatric DRE, which significantly differed between our DRE and non-DRE groups across various clinical subgroups. The performance of the ML modeling suggests that integrating network-aware EEG features with clinical phenotype enables superior, generalizable early prediction of pediatric DRE compared to clinical data alone. Our consistent results across clinical strata support the potential for cross-subtype generalizability and suggest that network-level EEG patterns capture fine-grained epilepsy heterogeneity beyond clinical profile alone.

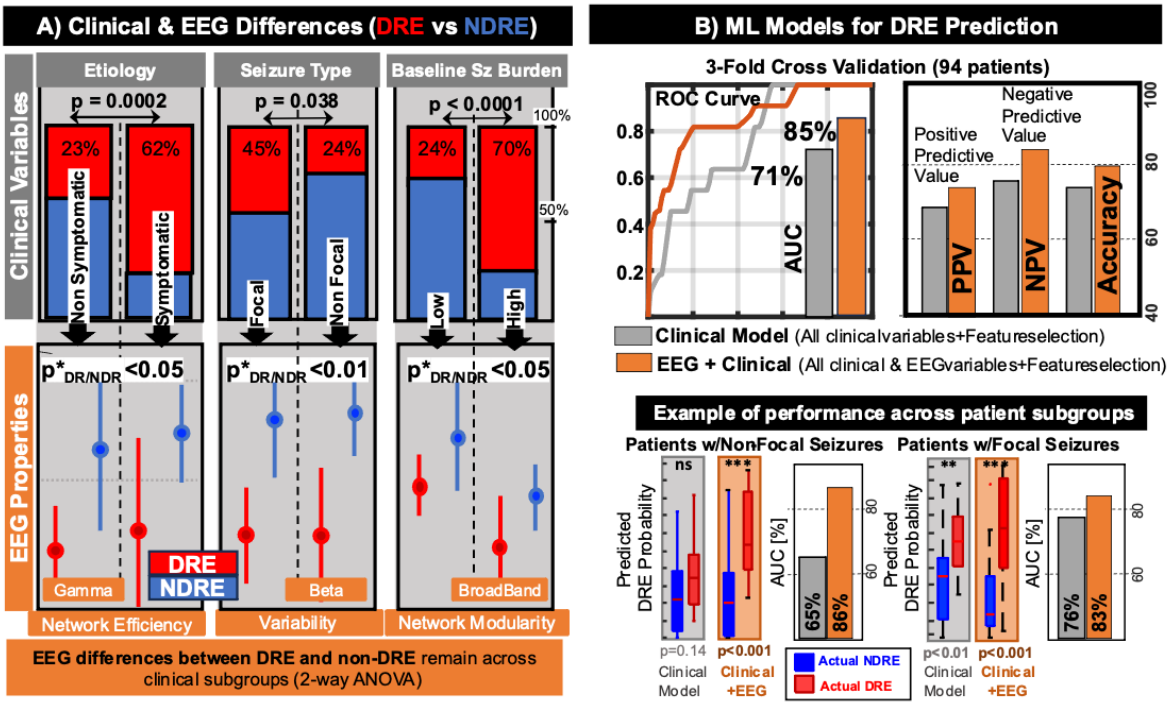
If further validated on a larger scale, the proposed tool can transform pediatric DRE care enabling: (1) earlier triage of high-risk patients and faster referrals for advanced diagnostics; (2) proactive family counseling and early data-informed planning; (3) streamlined escalation to surgical or device-based therapies once DRE is confirmed; (4) reduced cognitive decline and mortality via timely intervention within the optimal therapeutic window.

Figure 1. Methodology



https://secure.key4events.com/onsite/api.aspx?e=1960&img=Figure+1.png&ai=74284&op=getabstractimg&dirN=0

Figure 2. Results



https://secure.key4events.com/onsite/api.aspx?e=1960&img=Figure+2.png&ai=74284&op=getabstractimg&dirN=1

P56

Machine Learning–Driven MRI Lesion Mapping Tool for End-to-End Clinical Assessment in Drug-Resistant Temporal Lobe Epilepsy

J. Chen[1], K. Xie[1], E. Sahlas[1], I. Goodall-Halliwell[1], A. Ngo[1], S. Lariviere[2], R. Rodriguez-Cruces[1], Y. Weng[3], J. Dekraker[1], D. Schrader[4], Z. Zhang[3], L. Concha[5], A. Bernasconi[1], B. Neda[1], B. Bernhardt[1]

[1]Montreal Neurological Institute And Hospital - Montreal (Canada), [2]Université De Sherbrooke - Sherbrooke (Canada), [3]Jinling Hospital, Nanjing University School Of Medicine - Nanjing (China), [4]Bc Children's Hospital, University Of British Columbia - Vancouver (Canada), [5]Instituto De Neurobiología, Universidad Nacional Autónoma De México - Mexico City (Mexico)

Introduction

Epilepsy affects >60 million people worldwide, and for one-third of patients, remain a chronic syndrome that cannot be controlled by pharmacotherapies. These patients experience marked disruptions in psychosocial functioning, elevated psychiatric comorbidity, and significantly reduced quality of life, and will need to undergo further diagnostic investigations, such as multimodal resonance imaging (MRI), to delineate the underlying pathology and support surgical planning, the current gold-standard therapy. High-resolution, multimodal MRI is essential throughout the clinical work-up and treatment for pharmaco-resistant epilepsy patients; however, 30-40% will continue to experience seizures post-operatively, highlighting the need for more accurate diagnostic and prognostic tools to refine candidate selection and improve post-surgical seizure freedom rates. AI/ML approaches are poised to transform medicine by supporting the diagnostic and prognostic work-up of patients with pharmaco-resistant epilepsy. In temporal lobe epilepsy (TLE), the most common focal epilepsy in adults, AI/ML-driven models may support the following three main clinical challenges: (1) accurately subtyping TLE from other focal epilepsies (2) lateralizing TLE, and (3) predicting post-surgical seizure outcomes. Here, we introduce an opn-source ML-driven, patient-specific MRI lesion mapping tool integrating multimodal structural features to support clinical decision-making in TLE.

Methods

Participants and MRI preprocessing. 606 individuals from four independent sites (*MICs*: TLE/healthy controls/FCD = 65/100/17; *NOEL*: 72/42/28; Universidad Nacional Autónoma de México *(EpiC)*: 29/34; and Jinling Hospital (*Nanj*): 124/96; Overall: 290/272/44) underwent 3.0-Tesla T1-weighted (T1w) and diffusion MRI (**Table 1**). All patients underwent comprehensive multidisciplinary evaluation following ILAE guidelines to localize the seizure focus. Data was preprocessed using combat harmonization (https://github.com/Jfortin1/ComBatHarmonization)[1], *micapipe* (version 0.2.3; http://micapipe.readthedocs.io)[2], *hippunfold* (verson 1.3.0; https://hippunfold.khanlab.ca/en/latest/)[3] and mapped to surface points. Cortical and hippocampal surfaces were smoothed with 10mm and 5mm gaussian kernels respectively. Patients were compared against a normative range that account for biological covariates (i.e. linear regression model including age and sex) derived from healthy controls. *W*-scores were then calculated by taking the difference between the observed and predicted values from patients against this

normative range, with |*w*|<=1.96 considered to be significantly altered, allowing for individualized deviation mapping. Each patient's cortical and hippocampal surfaces were parcellated to 400- and 50-regions of interest (ROI)[3,4]. Right-sided TLE patients were then "flipped" for an ipsilateral/contralateral orientation. Cortical and hippocampal prevalence maps were derived by taking the proportion of significant alterations, providing a group-level summary of common deviation patterns.

Machine Learning (ML). Normalized w-scored features from the main cohort were used to train a support vector machine (SVM) classifier to address three key clinical tasks: 1) identifying TLE from other common focal epilepsy substrates (e.g. FCD); 2) lateralizing left- and right-TLE; and 3) predicting post-surgical seizure outcomes. Given the imbalanced cohorts between TLE and FCD patients, and seizure-free and non-seizure-free patients, we employ a weighted loss function relative to the ratio of the two groups. Classifiers were evaluated using five-fold cross-validation with bootstrapped participant sampling across multiple feature permutations (e.g., cortex-FA, hippocampus-ADC, cortex+hippocampus-FA), yielding ten feature sets per classification task. This design can ensure stable performance estimates and enabled comparison of feature-specific contributions, as well as facilitate more interpretable model outputs by clarifying how specific feature groups contribute to classification performance. Finally, each feature set was bootstrap (n=100) resampled to obtain performance distributions.

Results

When compared against healthy controls, TLE patients demonstrated diffuse structural deviations with strong localization to the ipsilateral temporal lobe (**Figure 1**). Across all feature types, cortical measures showed the highest deviations within the ipsilateral temporal lobe, with additional but lower-magnitude abnormalities observed in other neocortical regions. Hippocampal measures indicated volume loss in both the ipsilateral and contralateral hippocampi. Diffusion features differed in their spatial profiles: FA demonstrated mild ipsilateral deviations primarily localized to the hippocampal head, whereas ADC exhibited a strongly lateralized pattern with pronounced deviations only in the ipsilateral hippocampus. Across modalities, the temporal lobe exhibited the greatest proportion of significant deviations, yet for no feature did it surpass 50%, highlighting considerable heterogeneity within the TLE population. This heterogeneity underscores the need for personalized clinical assessment strategies and motivates evaluating whether patient-specific profiles can meaningfully enhance machine-learning classification.

Our SVM-based model demonstrated strong performance for both focal epilepsy subtyping and TLE lateralization, with more modest accuracy for post-surgical outcome prediction (**Table 2**). For distinguishing TLE from FCD, combined cortical and hippocampal features achieved the highest accuracy (0.80, 95% CI: 0.79–0.81), with cortical features alone performing similarly (0.76–0.78) and hippocampal features alone performing notably lower (0.61–0.63). Lateralization of TLE yielded the highest classification performance overall, with accuracies up to 0.84 (0.83–0.85) using combined features, driven largely by strong ipsilateral ADC signals; hippocampal ADC alone also performed well (0.82, 95% CI: 0.80–0.83). Prediction of postsurgical seizure freedom was more challenging, with combined features reaching 0.66 (95% CI: 0.61–0.71); diffusion features (FA/ADC) consistently outperformed thickness measures across both cortical and hippocampal regions.

Conclusions

These findings highlight the value of integrating multimodal, multi-structure MRI features to achieve reliable subtyping and lateralization performance in TLE. A key strength of our approach is the use of individualized lesion mapping via w-score deviation profiles, which not only captures patient-specific pathology but also enables clear interpretation of features driving model

performance. While the present study is limited by its retrospective design and internal cross-validation, the robustness of our findings suggests strong translational promise. External validation will be essential to establish generalizability when advancing towards prospective clinical deployment.

REFERENCES.

1. Fortin et al. Harmonization of cortical thickness measurements across scanners and sites. NeuroImage. 2018;167:104-120.
2. Rodriguez-Cruces et al. Micapipe: a pipeline for multimodal neuroimaging and connectome analysis. Neuroimage. 2022;263:119612.
3. DeKraker, Jordan, et al. "Evaluation of surface-based hippocampal registration using ground-truth subfield definitions." ELife 12(2023):RP88404.
4. Vos de Wael, Reinder, et al. "BrainSpace: a toolbox for the analysis of macroscale gradients in neuroimaging and connectomics datasets." Communications biology 3.1 (2020):103.

Table 1 (clinico-demographics) and 2 (ML results)

Table 1. Clinico-demographic information of training dataset.

Group	Age	Sex (M/F)	Laterality (L/R)	Age of onset	Disease duration	HS (%)	DRE (%)	Lesional/ Non-lesional	ENGEL I/II+
TLE (290)	32.5±10.3 (18-64)	135/155	151/139	17.9±10.7 (0.2-60)	14.5±11.0 (0-49)	196 (68)	221 (76)	--	74/25
HC (272)	30.1±9.4 (18-73)	151/147	--	--	--	--	--	--	--
FCD (44)	26.9±7.7 (18-54)	23/21	18/26	11.5±6.9 (0.5-27)	15.9±10.3 (0.5-48)	--	40 (88.9%)	15/19	25/10

Table 2. Support vector machine classification performance across feature permutations for focal epilepsy subtyping, TLE lateralization, and seizure-outcome prediction

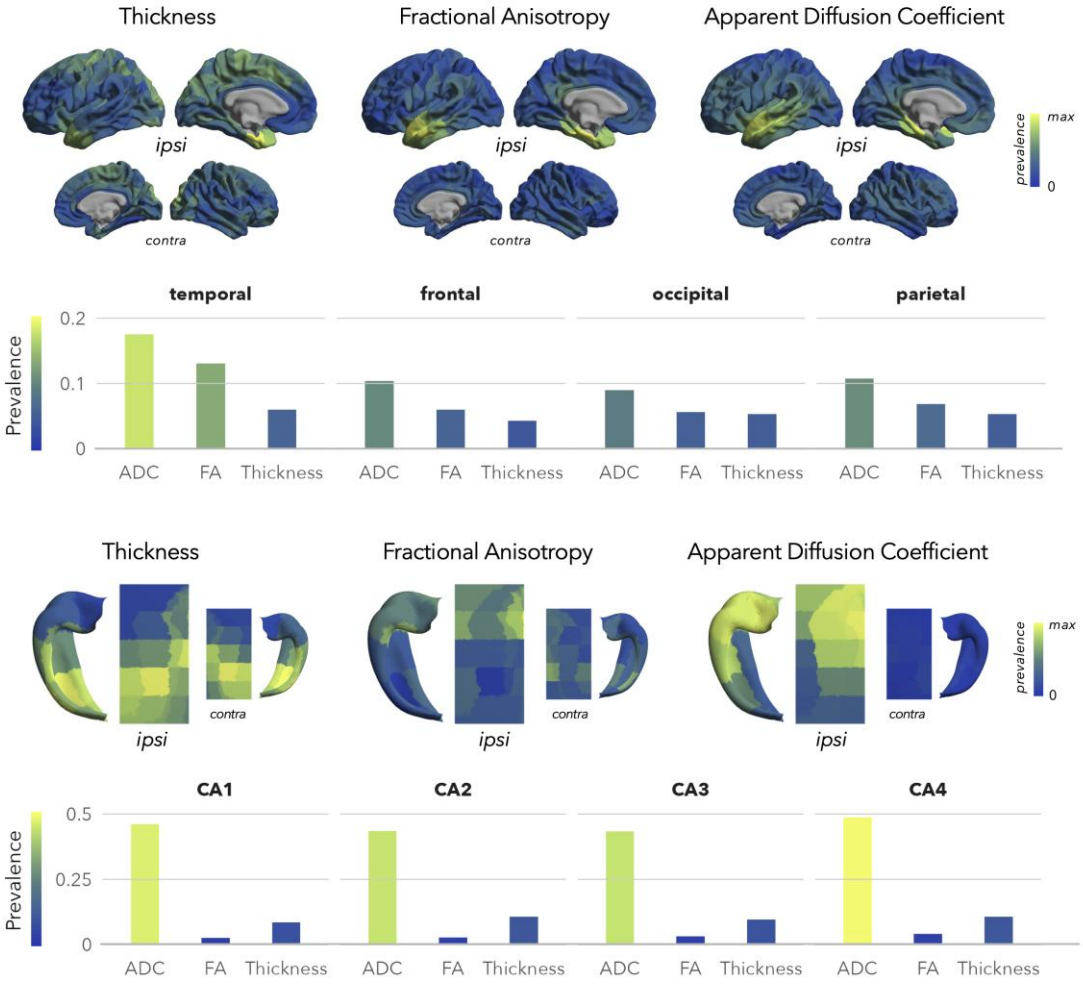
Structure	Feature	Mean Accuracy (95% CI)
TLE subtyping (*vs. FCD*)		
Cortex & hippocampus	**All**	**0.80 (0.79 - 0.81)**
	Thickness	0.78 (0.77 - 0.79)
	FA	0.78 (0.77 - 0.79)
	ADC	0.76 (0.75 - 0.77)
Cortex	**Thickness**	**0.78 (0.77 - 0.79)**
	FA	0.77 (0.76 - 0.78)
	ADC	0.76 (0.75 - 0.77)
Hippocampus	Thickness	0.61 (0.60 - 0.62)
	FA	**0.63 (0.62 - 0.63)**
	ADC	0.62 (0.60 - 0.63)
TLE lateralization (*left vs right*)		
Cortex & hippocampus	All	0.83 (0.81 - 0.83)
	Thickness	0.69 (0.68 - 0.70)
	FA	0.78 (0.76 - 0.78)
	ADC	**0.84 (0.83 - 0.85)**
Cortex	Thickness	0.69 (0.68 - 0.70)
	FA	0.73 (0.72 - 0.74)
	ADC	**0.79 (0.78 - 0.80)**
Hippocampus	Thickness	0.73 (0.72 - 0.74)
	FA	0.73 (0.72 - 0.74)
	ADC	**0.82 (0.80 - 0.83)**
Post-surgical seizure freedom status (*ENGEL I vs. ENGEL II+*)		
Cortex & hippocampus	All	0.66 (0.61 - 0.71)
	Thickness	0.53 (0.46 - 0.60)
	FA	**0.68 (0.60 - 0.75)**
	ADC	**0.68 (0.60 - 0.75)**
Cortex	Thickness	0.61 (0.55 - 0.66)
	FA	0.59 (0.51 - 0.66)
	ADC	**0.67 (0.62 - 0.72)**
Hippocampus	Thickness	0.63 (0.52 - 0.73)
	FA	**0.67 (0.62 - 0.71)**
	ADC	0.51 (0.45 - 0.58)

FA = fractional anisotropy; ADC = apparent diffusion coefficient; TLE = temporal lobe epilepsy; FCD = focal cortical dysplasia

https://secure.key4events.com/onsite/api.aspx?e=1960&img=Table_AIEND.png&ai=74311&op=getabstractimg&dirN=0

Figure 1. Region-specific prevalence maps of struc

Fig 1 | Region-specific prevalence maps of structural deviations in TLE



https://secure.key4events.com/onsite/api.aspx?e=1960&img=FIGURE_1_AIEND.png&ai=74311&op=getabstractimg&dirN=1

P57
Brain Entropy Index: A Modality-Invariant Biomarker of Neural Health

L. A. Sanchez[1], J. Y. Kang[2], C. U. Onyike[3], S. V. Sarma[1]

[1]Institute For Computational Medicine, Johns Hopkins University - Baltimore (United States), [2]Department Of Neurology, Johns Hopkins University School Of Medicine - Baltimore (United States), [3]Department Of Psychiatry And Behavioral Sciences, Johns Hopkins University School Of Medicine - Baltimore (United States)

Introduction

Modern medicine relies on quantitative biomarkers to assess organ function across health and disease. Troponin levels stratify cardiac risk, D-dimer reflects clotting dynamics, and viral loads guide infectious disease treatment. The brain lacks such a universal metric. Neurological disorders, including stroke, Alzheimer's disease, Parkinson's disease, and epilepsy, affect over three billion individuals worldwide and represent a leading cause of disability and death. Current assessment relies on detecting structural or functional changes after significant pathology has developed, while the preclinical phase remains largely inaccessible to existing methods.

We hypothesize that a healthy brain at rest maintains flexible, evenly distributed energy across regions. This balanced state, which maximizes thermodynamic entropy, can be quantified through five measurable neural properties: similar regional energy magnitudes, comparable distribution patterns across regions, independent regional fluctuations, low pathological synchrony, and stable temporal dynamics. We introduce the Brain Entropy Index (BEI), a physiological biomarker grounded in the principle that deviations from this maximum-entropy configuration signify neural dysfunction. By quantifying how closely the observed brain state approximates the theoretical equilibrium predicted by maximum entropy theory, BEI provides a disease-agnostic, modality-invariant measure of brain health.

Methods

We analyzed multimodal neuroimaging data from 525 participants spanning healthy controls (n=311) and nine diagnostic categories: Alzheimer's disease (AD, n=35), frontotemporal dementia (FTD, n=21), Parkinson's disease (PD, n=9), epilepsy (n=20), major depressive disorder (MDD, n=39), obsessive-compulsive disorder (OCD, n=27), schizophrenia (SCZ, n=36), attention-deficit/hyperactivity disorder (ADHD, n=16), and autism spectrum disorder (ASD, n=11). Data were acquired across seven repositories from multiple centers using scalp EEG, intracranial EEG, fMRI and MEG.

From the maximum entropy principle, we derived that regional energy distributions follow a multivariate lognormal form when the brain operates at thermodynamic equilibrium. We computed five energetic properties from preprocessed signals. Three emerge directly from maximum-entropy constraints: (1) Lognormality, quantifying adherence to maximum-entropy statistics via normalized log-likelihood; (2) Equal Energy Distribution (EED), measuring spatial uniformity via Jensen-Shannon divergence and Kolmogorov-Smirnov statistics; and (3) Coupling Ratio (CR), reflecting inter-regional statistical independence. Two additional properties derive from metabolic

homeostasis: (4) Energy Variance (VR), quantifying cross-regional energy dispersion; and (5) Brain Stability (BS), measuring temporal fluctuation magnitude through aggregate coefficient of variation. These five properties were integrated via binary logistic regression to yield a continuous BEI score representing the probability of neural health.

Results

Our analysis revealed consistent patterns supporting the existence of a maximum-entropy configuration in healthy brains. At the population scale, all five emergent properties converged on a coherent signature of neural health. Healthy participants clustered tightly around theoretical optima: regional energy conformed to multivariate lognormal distributions (normalized log-likelihood 0.9-1.0), energy allocation was remarkably uniform across regions (EED 0.8-1.0), coupling ratios remained low (0.2-0.4), variance was tightly constrained (<0.15), and stability was high (0.8-1.0). By contrast, neurological, neuropsychiatric, and neurodegenerative disorders exhibited systematic deviations across all metrics.

The integrated BEI demonstrated robust discrimination between health and disease (Fig 1). Healthy controls clustered near the upper entropy limit (median BEI=0.95, IQR=0.89-1.01), significantly exceeding all patient populations (Kruskal-Wallis H=412.3, df=9, p<0.0001). All healthy controls resided in the Healthy Zone (BEI>0.85), while 213 of 214 clinical cases fell below the diagnostic threshold (BEI=0.58), yielding 100% specificity and 99.5% sensitivity. External validation on six previously unseen diagnostic categories (tinnitus, stroke, traumatic brain injury, post-traumatic stress disorder, chronic pain, and bipolar disorder) confirmed complete separation between health and disease states, with all positive likelihood ratios exceeding 10.

Figure 1. Brain Entropy Index distinguishes healthy from pathological states across neurological, neuropsychiatric, and neurodegenerative disorders. (A) BEI values across diagnostic cohorts (n=525 subjects). Healthy controls exhibit significantly higher BEI (median=0.95) than all patient populations (Kruskal-Wallis H=412.3, p<0.0001). Three zones are demarcated: Healthy Zone (BEI>0.85), At-Risk Zone (0.58<BEI<0.85), and Disease Zone (BEI<0.58). (B) Likelihood ratios relative to healthy controls. Positive likelihood ratios (LR+) exceed 10 and negative likelihood ratios (LR-) approach 0.1 for all disorders, surpassing benchmarks for strong diagnostic evidence.

BEI exhibited robust correlation with clinical severity across four diagnostically independent cohorts spanning 93 patients (Fig 2). In OCD (n=27), BEI showed strong negative correlation with Yale-Brown Obsessive-Compulsive Scale scores (R^2=0.783, p=9.6×10^{-9}). In MDD (n=39), BEI decreased proportionally with Beck Depression Inventory severity (R^2=0.550, p=1.5×10^{-7}). Parallel gradients emerged in neurodevelopmental conditions: ASD (R^2=0.452, p=3.3×10^{-2}) and ADHD (R^2=0.743, p=7.4×10^{-5}). Statistical comparison of regression slopes revealed no significant heterogeneity across disorders, indicating a shared energetic trajectory despite distinct pathophysiologies.

Figure 2. BEI scales with disorder severity across independent diagnostic cohorts. Relationship between BEI (log scale) and validated clinical severity scales across four disorders: (A) OCD (Y-BOCS scores), (B) MDD (BDI scores), (C) ASD (ADOS-2 severity), and (D) ADHD (ADHD-RS scores). Shaded regions indicate 95% confidence and prediction intervals. BEI declines monotonically with worsening symptoms across all cohorts, revealing a shared energetic trajectory despite distinct pathophysiologies.

Longitudinal analysis in epilepsy patients (n=8) demonstrated treatment responsiveness. BEI values increased markedly upon establishment of effective anti-seizure medication regimens, shifting from disease or at-risk zones toward the healthy zone as seizure control was achieved. Group-level analysis confirmed significant elevation of BEI following therapeutic optimization (Wilcoxon signed-rank p=6.7×10^{-4}; Cohen's d=5.73), with all post-treatment values exceeding the safety-margin threshold. BEI remained robust across age and sex, with healthy controls showing

negligible age-related decline (Spearman's ρ=-0.17, p=0.15) and no significant sex differences (Mann-Whitney U test p=0.9).

Conclusions

Our work provides compelling evidence for a Brain Entropy Index that quantifies neural health through the lens of thermodynamic equilibrium. Using maximum entropy principles, we demonstrated that healthy brains operate near a high-entropy configuration characterized by lognormal energy distributions, statistical independence, equal energy allocation, low variance, and temporal stability. Pathological states represent systematic departures from this equilibrium along disorder-specific trajectories, yet converge on a shared endpoint of reduced entropy and constrained flexibility.

The BEI framework fundamentally shifts conceptualization from disease-specific structural or neurotransmitter abnormalities to a unified energetic signature of brain function. The near-perfect discrimination between health and disease (100% specificity, 99.5% sensitivity), robust scaling with clinical severity, generalization to unseen disorders, and treatment responsiveness establish BEI as a candidate universal biomarker for neural health assessment. By linking entropy maximization to measurable neural energy balance, BEI provides a quantitative physiological standard that transcends traditional nosological boundaries.

Future work will expand analysis to larger longitudinal cohorts, correlating BEI trajectories with disease progression and therapeutic outcomes across diverse patient populations. Prospective validation could establish new paradigms for early detection and preventive intervention, potentially enabling personalized treatment strategies grounded in thermodynamic principles. This research ultimately aims to translate computational network insights into clinical decision support tools that leverage the brain's natural energetic organization for improved diagnosis, monitoring, and therapeutic assessment.

P58
Dissociating physiological ripples and epileptiform discharges with vision transformers

D. Zhang[1], J. Kleen[1]

[1]University Of California, San Francisco - San Francisco (United States)

Introduction

Two frequently studied bursts of neural activity in the hippocampus are normal physiological ripples and abnormal interictal epileptiform discharges (IEDs). While they are different waveforms, IEDs are notoriously picked up as false positives when using typical automated ripples detectors which are prone to sharp edge artifacts. This has created challenges for studying ripples and IEDs independently. We leveraged recent advances in computer vision AI on time-frequency feature representations to enable more comprehensive and objective dissociation of these phenomena.

Methods

We retrospectively evaluated human intracranial recordings from 46 hippocampal depth electrode sites among 17 patients with focal epilepsy, the majority of whom had a seizure-onset zone/network involving the hippocampus. We implemented a common human ripple detection algorithm and broadband spectrograms of all detected “ripple candidates” were projected into low-dimensional space. We segmented them using k-means to infer pseudo-labels for probable ripples and probable IEDs. Independently, human expert IED labels were manually annotated for comparison. State-of-the-art vision transformer models were implemented on individual spectrograms to approach ripple vs. IED dissociation as an image classification problem.

Results

We detected 31,847 ripple/IED candidates, and a median 3.9% per patient (range: 0-47.2%) were IEDs based on expert label overlap. Low-dimensional projection of spectrograms separated canonical IEDs vs. ripples better than raw or ripple-filtered waveforms. Canonical ripple and IED candidates emerged at opposite poles with a continuous landscape of intermediates in between. This spectrotemporal feature space was similar across electrodes localized to the CA1, CA3, and DG subfields. A binary vision transformer model trained on expert-labeled IED vs. non-IED candidate spectrograms with 5-fold cross-validation showed a mean area under the curve (AUC) of 0.970 and mean precision-recall curve of 0.694, both significantly above chance. To evaluate generalizability, we implemented a leave-one-patient-out cross-validation approach, in which training on pseudo-labels and testing on expert-labeled data demonstrated near-expert performance (mean AUC 0.966 across patients, range 0.892-0.997). Transformer-derived attention maps revealed that models were tuned to triangle-like edge artifact spatial features in the spectrograms. Model-derived probabilities (i.e. of being an IED) for all candidates demonstrated a continuous transitional distribution between ripples vs. IEDs, mirroring the low-dimensional projection gradients.

Conclusions

The delineation between ripples and IEDs appears best represented as a gradient (i.e. not binary) due to physiological ripple features overlapping with sharpened and/or high frequency pathophysiological IED features. ViTs nevertheless perform virtually at human expert levels in dissociating these phenomena by leveraging time-frequency spatial features enabled by neural data spectrograms. Such tools applied to spectrotemporal representations may augment comprehensive investigations in cognitive neurophysiology and epilepsy signal biomarker optimization for closed-loop applications.

P59

Leveraging Machine Learning to Address Epilepsy Treatment Gaps in Rural Primary Care: A Case Report

C. Saad[1], A. Ravisankar[1]

[1]University Of Illinois College Of Medicine Peoria - Peoria, Il (United States)

Introduction

The use of artificial intelligence (AI) in the workforce is increasing year after year. In late 2025, up to 45% of U.S. employees use AI “a few times per year,” whereas 10% use it daily. One sector where AI has been implemented is healthcare, evident in tools such as sepsis alerts for admitted patients and mobile health applications. Epilepsy remains an area of medicine with a significant treatment gap, with an estimated one-third of newly diagnosed individuals remaining untreated for up to three years after diagnosis. This discussion explores the concept of utilizing machine learning as a tool to bridge the treatment gap for epilepsy patients within primary care in rural America.

Methods

A case report was analyzed to illustrate barriers to epilepsy care and identify opportunities for AI-driven interventions.

Results

A 44-year-old male presented to the emergency department via stretcher after a witnessed bilateral tonic-clonic seizure in a hospital setting. The patient was visiting a family member when the event occurred while he was asleep in a recliner. He was promptly treated with a loading dose of levetiracetam. Postictal symptoms included generalized muscle soreness and headache. A head CT was performed, which was normal. The patient disclosed nonadherence to prescribed levetiracetam (750 mg twice daily) due to insurance and affordability issues. He reported experiencing seizures approximately every three months for the past two years, does not drive, and his employment status is unclear. After returning to baseline, he was scheduled for neurology follow-up in three months but did not attend. He resides 90 minutes from the scheduled appointment and one hour from the nearest neurology office, highlighting geographic and socioeconomic barriers to care.

Conclusions

This case underscores the persistent treatment gap in epilepsy management, particularly in rural populations where access to specialty care is limited. Machine learning–based tools integrated into primary care settings could help identify high-risk patients, predict nonadherence, and facilitate timely interventions through telehealth or automated alerts. Such approaches have the

potential to reduce delays in treatment, improve medication adherence, and ultimately decrease seizure-related morbidity. Future research should focus on developing and validating AI-driven models tailored to rural healthcare environments to enhance equity and continuity of care.

P60

Deep phenotyping with unsupervised clustering reveals new subtypes of temporal lobe epilepsy

P. Klimes[1, 2], P. Bosque-Varela[1, 3], I. Dolezalova[4], L. Wadi[1], M. Moye[1], K. Jaber[1, 5], Y. Zhu[1], J. Stuart[6], T. Aung[7], J. Li[7], J. Gotman[8], D. Southwell[9], D. Carlson[10], B. Frauscher[1, 5]

[1]Analytical Neurophysiology Lab, Department Of Neurology, Duke University Medical Center - Durham (United States), [2]Institute Of Scientific Instruments, Czech Academy Of Sciences - Brno (Czechia), [3]Department Of Neurology, Christian Doppler University Hospital, Centre For Cognitive Neuroscience, Member Of The European Reference Network Epicare, Paracelsus Medical University Of Salzburg - Salzburg (Austria), [4]St Anne's University Hospital - Brno (Czechia), [5]Department Of Biomedical Engineering, Duke Pratt School Of Engineering - Durham (United States), [6]Department Of Neurology, Duke University Medical Center - Durham (United States), [7]Department Of Neurology, University Of Pittsburgh Medical Center - Pittsburgh (United States), [8]Montreal Neurological Institute, Mcgill University - Montreal (Canada), [9]Department Of Neurosurgery, Duke University Medical Center - Durham (United States), [10]Department Of Biostatistics And Bioinformatics, Duke University - Durham (United States)

Introduction

Temporal lobe epilepsy (TLE) is the most common focal epilepsy type, yet the current subtype labels - Mesial, Neocortical, Mixed, and TLE-Plus - capture only coarse clinical distinctions and insufficiently reflect underlying variability in disease biology, epileptogenic network organization, and treatment response. The heterogeneity within these categories complicates surgical decision-making and contributes to inconsistent postoperative outcomes. Deep phenotyping approaches, integrating high-dimensional multimodal data to derive data-driven subtypes, offer an opportunity to refine TLE classification and discover more homogeneous patient groups. Here, we present proof-of-concept data from a single-center cohort demonstrating that multimodal, feature-level deep phenotyping produces clinically meaningful clusters with better feature homogeneity and stronger alignment with markers of focality than traditional TLE subtypes.

Methods

We analyzed 51 consecutive adult patients with drug-resistant TLE who underwent standard presurgical evaluation at Duke University Hospital (2023–2025) and subsequent neurosurgical intervention guided or unguided by stereo-EEG (SEEG). Each patient was characterized by ~140 multimodal features summarizing clinical background, seizure semiology, electrophysiology, neuroimaging, and neuropsychology.

Clinical, demographic, and neuropsychological variables were extracted from the ANPHY lab RedCap research database, which aggregates structured presurgical evaluation data, operative reports, and seizure conference summaries. Electrophysiological features were derived from both interictal and ictal SEEG recordings. Interictal biomarkers such as spectral power, entropy, interictal spike rates, ripple-spikes, and high-frequency oscillation (HFO) rates were computed

from a 30-minute NREM sleep segment selected at least two hours away from any seizures. Ictal features were quantified using a seizure fingerprint model applied to representative seizure segments (Grinenko et al. 2018). Neuroimaging findings were encoded using structured radiology reports and validated lesion descriptors. Hippocampal asymmetry was assessed using MAP18 automated morphometry, producing volumetric deviation indices that were integrated into the multimodal feature set.

Because the feature set comprised mixed data types, we applied Factor Analysis of Mixed Data (FAMD) to obtain a unified low-dimensional representation of the multimodal space. Prior to dimensionality reduction, missing values were addressed using median imputation with missingness indicators, allowing the model to retain information about whether a value was originally observed or imputed. The optimal number of FAMD components was determined using an elbow-based procedure that selected the dimensionality yielding the best silhouette score after clustering while retaining a parsimonious set of components. K-means clustering was then performed on the resulting FAMD latent space. To enable direct comparison with the four traditional clinical TLE subtypes, we constrained the algorithm to form four data-driven clusters. Cluster validity was quantified using the silhouette coefficient (0-1; higher score means higher separability).

To evaluate neurobiological relevance, we analyzed the 5-SENSE score, a validated marker of the focality of the epileptogenic zone (range 0–100; higher values indicate highly localized epileptogenic networks, likely associated with favorable surgical outcomes) (Astner-Rohracher et al. 2022). We compared 5-SENSE score distributions across (i) the four clinical TLE subtypes and (ii) the four data-driven clusters using the Kruskal–Wallis test, and computed eta squared (η^2) to quantify explained variance. Robustness was assessed via a 10,000-iteration permutation test in which cluster labels were randomly reassigned to generate a null distribution for η^2.

Results

The four data-driven clusters showed markedly improved internal structure compared to the existing TLE labels (Figure 1). The silhouette score for the data-driven solution was 0.403, whereas the silhouette score for the clinical subtypes was –0.073, indicating that traditional categories do not reflect meaningful separability in the multimodal feature space. The most distinguishing features were derived from electrophysiology, neuroimaging, and seizure semiology.

In terms of the 5-SENSE score, traditional clinical TLE subgroups showed substantial overlap and limited separability (Figure 2). Mesial TLE exhibited the highest focality (median = 76.9, IQR = 34.7), followed by Neocortical (median = 66.4, IQR = 35.3), Mixed (median = 46.8, IQR = 30.1), and TLE-Plus (median = 22.3, IQR = 57.9). There were no statistically significant differences between traditional TLE subgroups (Kruskal–Wallis $p = 0.126$).

In contrast, data-driven clusters demonstrated sharper separation between high- and low-focality patients and reduced overlap between intermediate groups ($p = 0.018$), explaining 28% of the variance ($\eta^2 = 0.28$). Cluster #1 showed the highest focality (median = 84.7, IQR = 32.2), with Cluster #2 (median = 70.8, IQR = 49.6) and Cluster #3 (median = 60.9, IQR = 34.1) forming intermediate phenotypes, and Cluster #4 capturing the lowest-focality group (median = 22.3, IQR = 60.2).

Permutation testing confirmed robustness: the probability of obtaining an η^2 as large as the observed value under random label assignment was $p = 0.0017$, indicating that the association between the new clusters and epileptogenic focality is highly unlikely to occur by chance.

Conclusions

Multimodal deep phenotyping produces significantly more homogeneous and better-separated TLE subtypes than current clinical classifications. The improved silhouette score and the strong clustering of patients by spatial focality collectively demonstrate that data-driven phenotypes better reflect underlying disease biology and may enable more precise prognostication and personalized treatment selection.

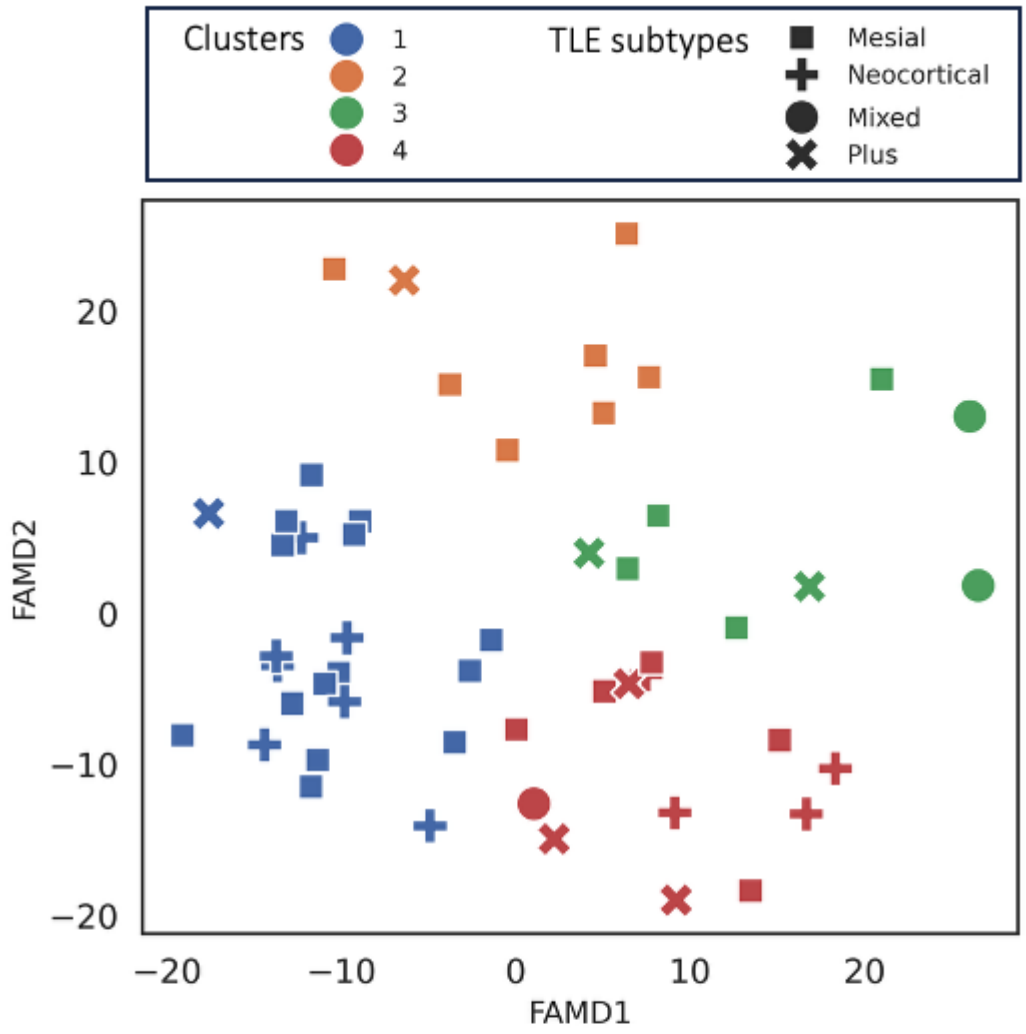


Figure 1. Cluster components showing the distribution of 51 TLE patients in low-dimensional feature space. Symbols denote the clinical TLE subtypes. Colors denote the four data-driven phenotypes. These new phenotypes exhibit clear separation and compactness, reflected by a silhouette score of 0.403, indicating substantially greater internal coherence than the existing TLE classification (silhouette = –0.073). These results demonstrate that deep phenotyping uncovers meaningful structure not captured by clinical subtypes.

https://secure.key4events.com/onsite/api.aspx?e=1960&img=Klimes_AIepilepsy_Fig1.png&ai=75289&op=getabstractimg&dirN=0

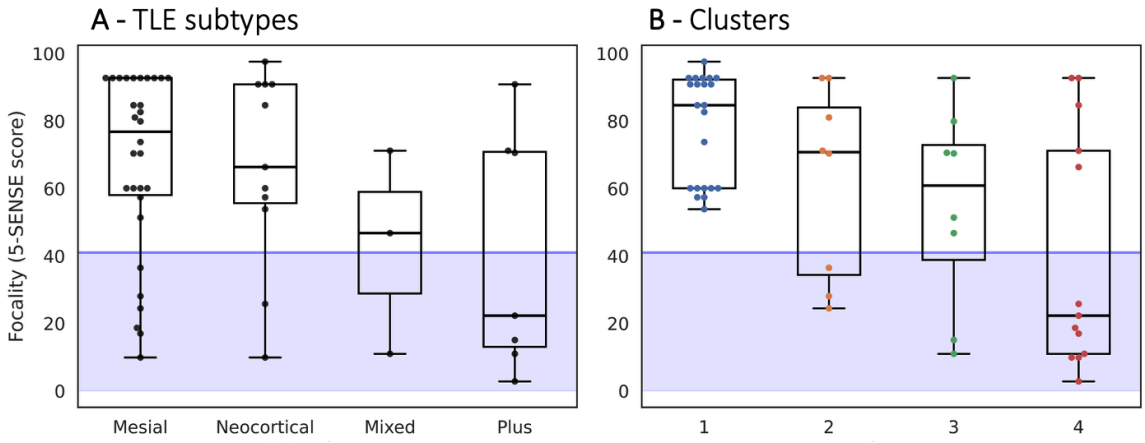


Figure 2. Boxplots showing the distribution of the 5-SENSE score across the original clinical TLE subtypes (A) and the new data-driven phenotypes (B). Clinical labels do not significantly differentiate focality ($p = 0.126$). Deep phenotyping reveals markedly distinct focality profiles ($p = 0.018$; $\eta^2 = 0.28$), and permutation testing confirmed that this effect is unlikely to occur by chance ($p < 0.0017$). These findings indicate that data-driven phenotypes more accurately capture neurophysiological distinctions relevant to surgical candidacy. The blue line indicates the 5-SENSE score focality threshold.

https://secure.key4events.com/onsite/api.aspx?e=1960&img=Klimes_AIepilepsy_Fig2.png&ai=75289&op=getabstractimg&dirN=1

P61
Biomarker Attenuation Following Virtual Resection Validates Clinically Defined Epileptogenic Networks

D. Wang[1], E. Pereira[2], G. Kamali[1], V. Boerwinkle[3], J. Gonzalez-Martinez[4], J.Y. Kang[1], S. Sarma[1]

[1]Johns Hopkins University - Baltimore (United States), [2]Texas Tech University - Lubbock (United States), [3]University Of North Carolina At Chapel Hill - Chapel Hill (United States), [4]University Of Pittsburgh - Pittsburgh (United States)

Introduction

Epilepsy affects over 60 million people worldwide. Approximately one-third of these individuals suffer from medically refractory epilepsy (MRE), defined as inadequate seizure control despite trials of two or more appropriately chosen antiseizure medications. For these patients, surgical resection of the epileptogenic zone (EZ) represents the most effective treatment option. However, outcomes remain highly variable, with only 30–70% of patients achieving long-term seizure freedom, primarily due to challenges in accurately localizing the EZ.

Standard presurgical evaluation typically relies on prolonged intracranial EEG (iEEG) monitoring to capture spontaneous seizures. While interictal biomarkers like High-Frequency Oscillations (HFOs) have been proposed, their clinical utility is limited by inconsistent performance and lack of network-level insight. Network-based computational approaches view epilepsy as a disorder of connectivity. Specifically, patient-specific Dynamic Network Models (DNMs) estimated from interictal data can capture directional influences among brain regions.

We propose a framework based on the Leading Eigenvector (LEV) of these models. In linear systems theory, the LEV represents the steady-state distribution of network activity. Previous work indicates that LEV components correlate with the Source-Sink Index (SSI), identifying regions that persistently accumulate influence. Virtual resection—the simulation of removing brain regions in silico—offers a way to test surgical hypotheses. We hypothesized that if the clinically defined EZ is accurate, its virtual resection should significantly attenuate the network's dominant eigenmode.

Methods

Patient Cohort and Data Acquisition:

We analyzed a retrospective cohort of 44 adult patients with MRE treated at four centers: Cleveland Clinic, Johns Hopkins Hospital, National Institutes of Health, and University of Pittsburgh Medical Center. All participants underwent sEEG monitoring and subsequent resection (n=29) or ablation (n=15). Outcomes were classified as Engel Class 1 (success, n=21) or Engel Class 2–4 (failure, n=23) at one year postoperatively.

Dynamic Network Modeling:

A single randomly selected interictal epoch was extracted for each patient. Signals were preprocessed (0.5--300 Hz bandpass, 60 Hz notch, common average reference). We estimated

linear time-varying (LTV) models comprising sequences of linear time-invariant (LTI) models from 500 ms windows:

$$x(t+1) = Ax(t)$$

where A is the state-transition matrix. Only models with $R^2 > 0.6$ were retained.

Virtual Resection:

We computed the LEV for each matrix A. Virtual resection was simulated by setting the columns of A corresponding to clinically annotated EZ channels to zero ($A_{post}(:, j) = 0$), eliminating their outgoing influence. We quantified the change in LEV magnitude:

$$\Delta\mathrm{LEV} = |\mathrm{LEV}_{pre}| - |\mathrm{LEV}_{post}|$$

We defined the "EZ tail proportion" as the fraction of elements in the top 5% of the ΔLEV distribution that originated from EZ channels.

Results

Biomarker Attenuation:

Virtual resection produced distinct LEV attenuation patterns correlated with surgical outcome. Patients achieving seizure freedom (Engel Class 1) exhibited a mean EZ tail proportion of 0.449, whereas patients with unfavorable outcomes (Engel Class 2–4) displayed markedly lower proportions. In Figure 1, we see the EZ tail proportion stratified by Engel outcome score via box plots showing the distribution of values within each Engel outcome class. Here, the EZ tail proportion is defined as the fraction of elements in the top 5% of the ΔLEV distribution that originate from clinically annotated EZ channels; Engel Class 1 patients (seizure free) exhibit significantly higher EZ tail proportions compared to Engel Class 2–4 patients (ongoing seizures). This indicates that the clinically targeted regions exerted strong influence on the network's steady-state dynamics. In contrast, failed surgeries showed significantly lower EZ tail proportions, suggesting the resected regions were not critical network sinks.

Outcome Prediction:

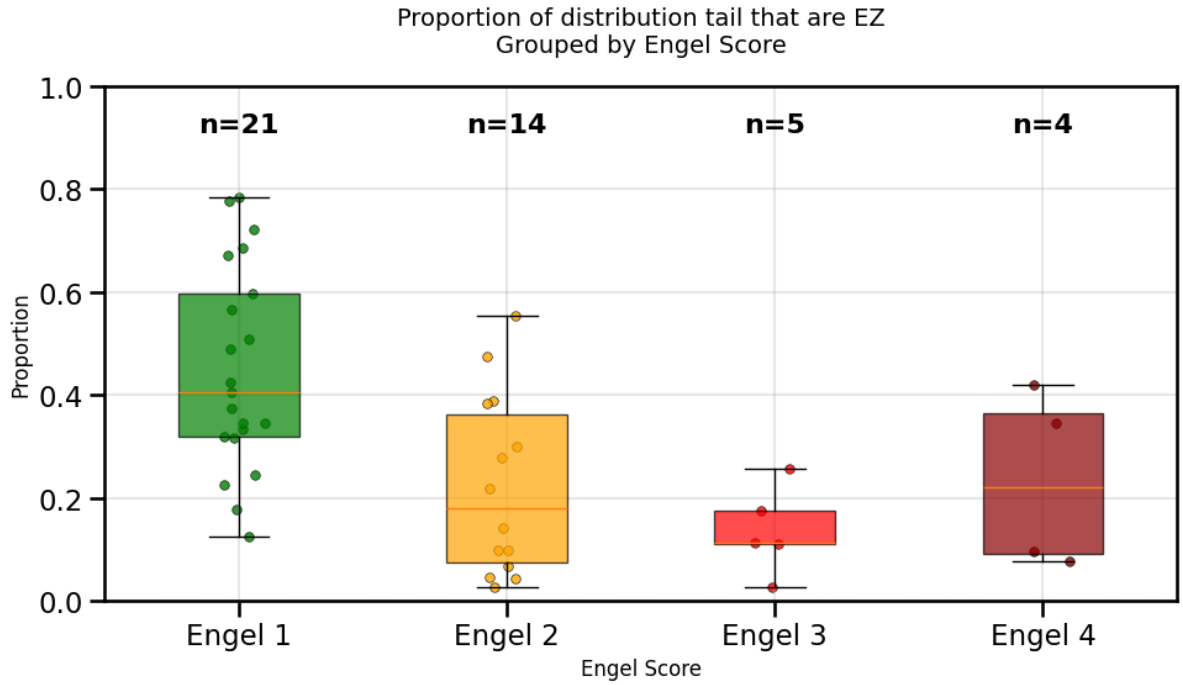
To validate the clinical utility of this biomarker, we trained a logistic regression classifier using the EZ tail proportion as the sole predictor. In leave-one-out cross-validation (LOOCV), the model achieved an Area Under the Curve (AUC) of 0.8, with 71% accuracy, 74% specificity, and 67% sensitivity. Figure 2 depicts the machine learning model performance for surgical outcome prediction, evaluating the logistic regression classifier using EZ tail proportion. Panel (A) shows box plots of EZ tail proportion values in the training cohort (n=44) stratified by outcome. Panel (B) shows the Receiver Operating Characteristic (ROC) curve from leave-one-out cross-validation (LOOCV), demonstrating discrimination between Engel Class 1 and Engel Class 2–4 outcomes (AUC = 0.8). Panel (C) displays the confusion matrix for the test set showing classification performance: overall accuracy = 71%, specificity = 74%, sensitivity = 67%. Higher EZ tail proportions were significantly associated with surgical success ($p=0.0014$).

Conclusions

This study validates LEV attenuation as a biomarker for virtual resection efficacy in a multicenter cohort. Unlike previous approaches requiring ictal data or computationally intensive neural mass models, our framework leverages abundant interictal data and efficient linear systems theory. The results demonstrate that surgical success depends on disrupting the dominant network mode

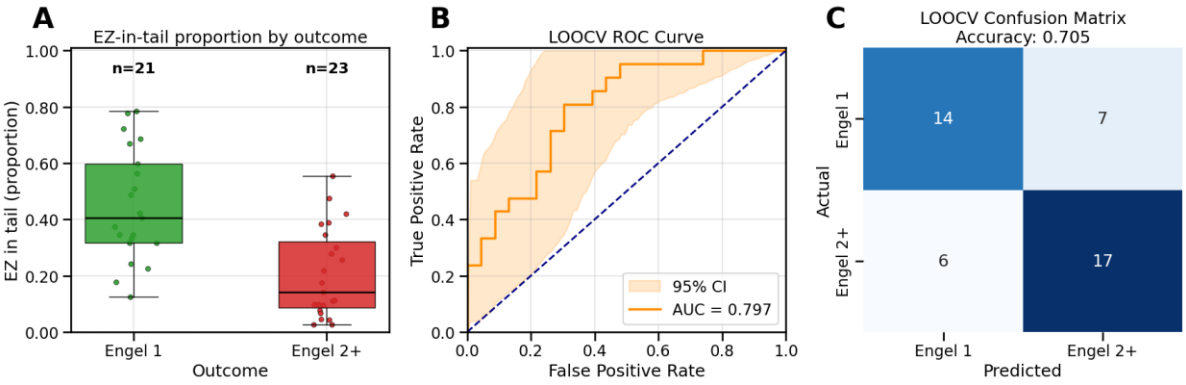
stabilized by the EZ. This LEV-based virtual resection framework offers a scalable, mechanism-based tool to evaluate presurgical hypotheses and potentially improve outcomes in drug-resistant epilepsy.

EZ Tail Proportion Stratified by Engel Score



https://secure.key4events.com/onsite/api.aspx?e=1960&img=proportion_by_engel.png&ai=76018&op=getabstractimg&dirN=0

Machine Learning Model Performance



https://secure.key4events.com/onsite/api.aspx?e=1960&img=ml_model.png&ai=76018&op=getabstractimg&dirN=1

P62

Bridging Clinical Criteria and Deep Learning: An Interpretable model for Automated IED Detection in Scalp EEG

L.C. Martinez Ruiz[1], A. Guerrero Aranda[2, 3], H. Vélez Pérez[1], R. Romo Vázquez[1], O. Paredes[4, 5]

[1]Dept Of Translational Bioengineering, University Of Guadalajara - Guadalajara (Mexico), [2]Dept Of Health Sciences, University Of Guadalajara - Guadalajara (Mexico), [3]Epilepsy Clinic, Hospital Country 2000 - Guadalajara (Mexico), [4]Universidad De Guadalajara - Guadalajara (Mexico), [5]Biodigital Innovation Lab - Guadalajara (Mexico)

Introduction

Interictal epileptiform discharges (IEDs) are brief abnormal transients observed on scalp electroencephalography (EEG) between seizures and serve as biomarkers for epilepsy diagnosis and focus localization. In clinical practice, IEDs are identified through expert visual review, a process that is time-consuming, experience-dependent, and subject to limited inter-rater agreement. IED-like waveforms may overlap with artifacts or normal variants, motivating the development of accurate, clinically usable automated detection methods. The International Federation of Clinical Neurophysiology (IFCN) proposed six operational criteria for defining IEDs: spiky morphology, duration difference from background, waveform asymmetry, after-going slow wave, background disruption, and plausible physiological field distribution. However, these qualitative criteria may contribute unequally to discrimination. This study bridges detection performance and interpretability by combining frequency-aware latent representations from Variational Autoencoders (VAEs) with Kolmogorov-Arnold Network (KAN) classifiers, enabling both accurate IED detection and explainable feature analysis aligned with clinical criteria

Methods

Data. The pipeline was developed using an in-house annotated routine scalp EEG dataset from Hospital Country 2000 in Guadalajara, Mexico (Dataset A) comprising 90 records (36 epilepsy, 54 non-epilepsy; 40 males, 50 females; average duration 18.3 min) containing 1,512 single-channel IED epochs. External validation used an independent dataset of 13-second EEG segments with record-level epilepsy labels (Dataset B) comprising 100 records (54 epilepsy, 46 non-epilepsy; 40 males, 60 females), no records from this dataset were used to train any stage of the model.

Preprocessing. Records were harmonized by retaining 19 standard 10-20 channels, resampling to 256 Hz, notch filtering (60 Hz for Dataset A; 50 Hz for Dataset B), average re-referencing, converting amplitudes to μV, and discarding segments exceeding 400 μV.

Epochs generation. From Dataset A, IEDs were extracted as 1-second single-channel epochs centered on annotated IED peaks, yielding 4,536 IED epochs after data augmentation (jittering ±200 ms). Non-IED epochs were randomly sampled from non-epilepsy records (750 epochs per record; 36,000 total). Epochs were split using stratified random sampling into train/validation/test sets (70/15/15).

VAE Training. Dataset A records were split patient-wise (80 training, 10 testing) to prevent data leakage. For each patient, 500 1-second single-channel segments were randomly extracted

(40,000 training epochs, 5,000 testing epochs). Each epoch was filtered into five frequency bands: delta (0.5-4 Hz), theta (4-8 Hz), alpha (8-13 Hz), low beta (13-20 Hz), and high beta (20-30 Hz). A separate VAE was trained per band to reconstruct band-limited signals using MSE + Kullback-Leibler divergence loss. Latent dimensions were z=[8,10,12,10,15] (delta to high beta); band latents were concatenated into a 55-dimensional vector. Reconstruction quality was quantified using Pearson correlation.

KAN Training. The KAN received 55-dimensional latent vectors and predicted IED vs. non-IED. Training used an 8:1 non-IED:IED ratio, binary cross-entropy with class weighting (1/8 penalty for non-IED), and LBFGS optimization with spline functions (grid G=10; order K=3) and regularization (λ=0.01, λ_ent=10). Performance was evaluated using accuracy, precision, recall, F1-score, and balanced accuracy.

External Validation and Interpretability. For Dataset B, each record was segmented into 1-second single-channel epochs with 50% overlap, encoded with trained VAEs, and classified with the trained KAN; records were labeled epilepsy if any epoch was predicted as IED. Interpretability was assessed by inspecting learned KAN spline functions and weights, and through sensitivity analysis: varying individual latent dimensions within observed ranges, decoding resulting signals, and measuring changes in KAN output to identify features corresponding to IFCN criteria.

Results

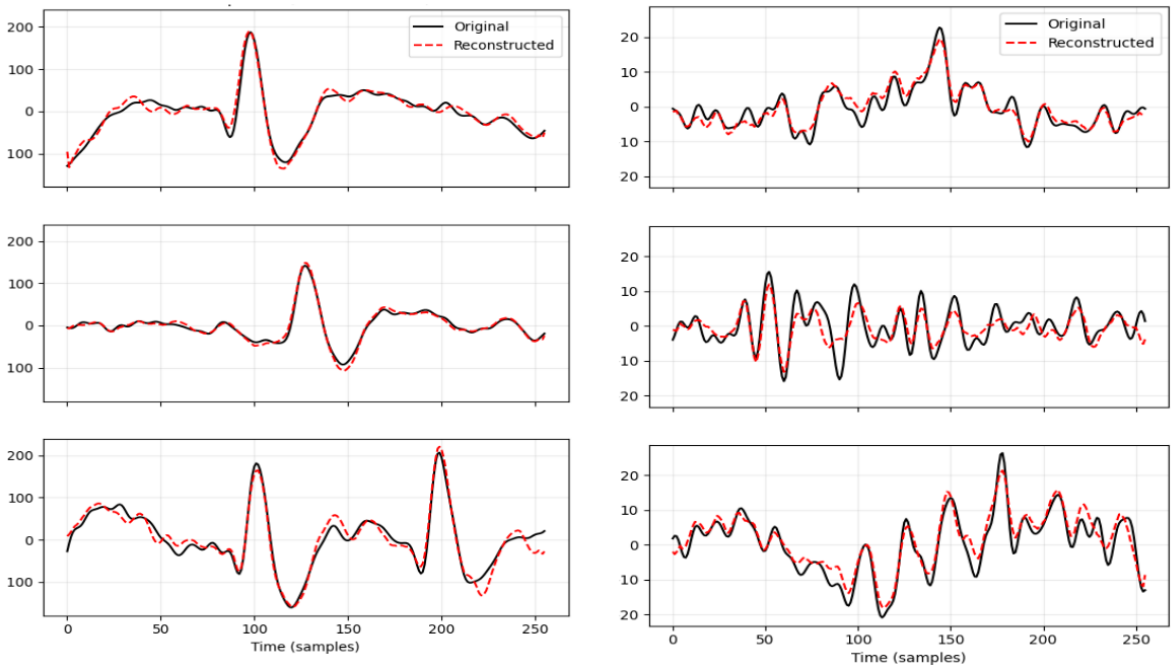
The VAE achieved high reconstruction fidelity with Pearson correlation coefficients of 0.98 ± 0.01 for IED-containing epochs and 0.93 ± 0.12 for non-IED epochs. On Dataset A, the KAN classifier demonstrated accuracy of 0.95, sensitivity of 0.94, specificity of 0.95, precision of 0.71, F1-score of 0.81, and balanced accuracy of 0.94. External validation on Dataset B showed accuracy of 0.62, sensitivity of 0.86, specificity of 0.35, precision of 0.60, F1-score of 0.71, and balanced accuracy of 0.61.

Feature importance analysis revealed that theta band features contributed most heavily (31.73%), followed by alpha (26.09%) and low beta (18.84%). However, individual feature analysis showed that two low beta features (features 31 and 34) were among the most influential overall, while theta band importance was distributed across multiple features. Band-specific analysis indicated that IFCN criteria for pointed peaks and slow waves were predominantly captured by theta-alpha band combinations.

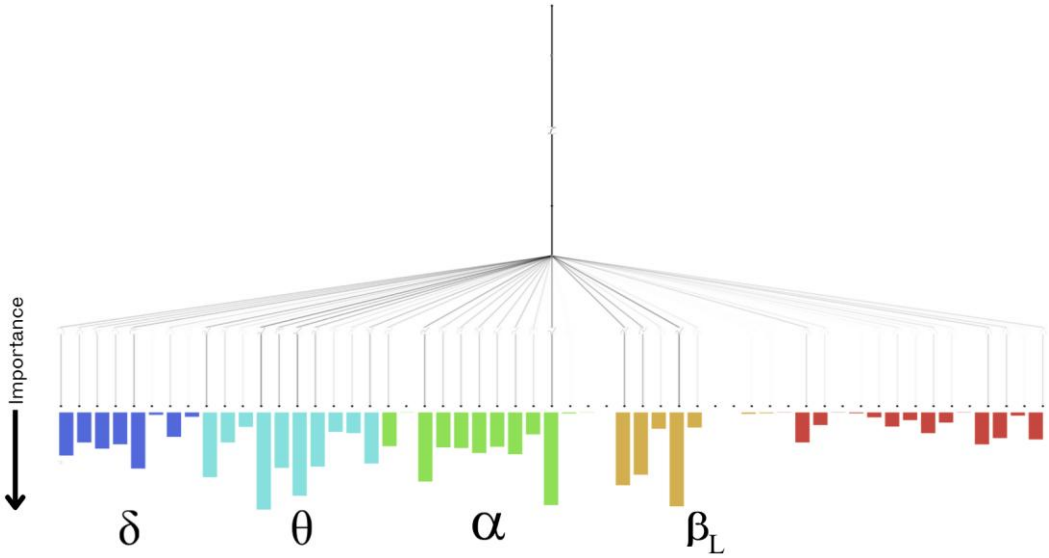
Conclusions

The VAE+KAN architecture demonstrates promising performance for interpretable IED detection. The VAE's high reconstruction fidelity, even for IED patterns not explicitly selected during training, suggests its capacity to capture signal relationships in low-dimensional space. The KAN's direct weight accessibility enables transparent identification of decision-making features, with theta-alpha bands emerging as key contributors to canonical IED morphology criteria. However, reduced performance on external validation indicates the need for more diverse training data encompassing varied IED morphologies and challenging mimics (e.g., blink artifacts, sharp transients, wicket spikes). Future work should expand training datasets and perform detailed latent space analysis to characterize discriminative features for IED versus common confounders.The VAE+KAN architecture demonstrates promising performance for interpretable IED detection. The VAE's high reconstruction fidelity, even for IED patterns not explicitly selected during training, suggests its capacity to capture signal relationships in low-dimensional space. The KAN's direct weight accessibility enables transparent identification of decision-making features, with theta-alpha bands emerging as key contributors to canonical IED morphology criteria. However, reduced

performance on external validation indicates the need for more diverse training data encompassing varied IED morphologies and challenging mimics (e.g., blink artifacts, sharp transients, wicket spikes). Future work should expand training datasets and perform detailed latent space analysis to characterize discriminative features for IED versus common confounders.



https://secure.key4events.com/onsite/api.aspx?e=1960&img=reconstruction.png&ai=78579&op=getabstractimg&dirN=0



https://secure.key4events.com/onsite/api.aspx?e=1960&img=kan_features.png&ai=78579&op=getabstractimg&dirN=1

P63
Automated EEG Screening with a Transformer Model in Shadow Operation in Kenya

W. Lehn-Schiøler[1, 2], M.G. Pedersen[1], A.M. Storgaard[1]

[1]Braincapture - Lyngby (Denmark), [2]Dtu Health Tech - Lyngby (Denmark)

Introduction

In many emerging markets, including Kenya, Indonesia, the Philippines, Malaysia, and Vietnam, the shortage of trained neurophysiologists means that a large number of people live with undiagnosed epilepsy. This is known as the Epilepsy Treatment Gap. While portable EEG solutions, like the BrainCapture (BC-1) system, have increased data collection capacity, the bottleneck remains the expert interpretation of these recordings. Artificial Intelligence (AI), particularly the use of self-supervised learning (SSL) and transformer-based models offers a scalable solution to assist technicians and clinicians. This study evaluates the real-world performance of a transformer-based EEG screening model operating in a "shadow" capacity within a clinical workflow in Kenya, aiming to provide automated triaging and decision support in resource-limited settings.

Methods

We deploy an EEG transformer model into the BrainCapture infrastructure to analyze routine EEG recordings uploaded from more than 20 clinical sites in Kenya. The model flow was designed to automatically screen every recording upon upload. During this "shadow" phase, the AI results were generated in the background without influencing the primary clinical decision making.

The transformer model, inspired by LaBraM [1] utilizes an attention-based architecture to capture long-range temporal and spatial dependencies across EEG recordings. The model was pretrained on more than 2.500 hours of data and finetuned on a collection of 2130 EEG recordings from the BrainCapture operation during the period from November 2023 to October 2025.

Data was preprocessed using an evolved version of the pipeline described in [2]. The model was finetuned on a binary classification task in triaging abnormal and normal EEG recording, using 60-second windows. In this work, an abnormal EEG is a recording that contains interictal findings in the form of _focal or generalized spike (or polyspike) and wave activities, focal or generalized sharp waves, burst suppression patterns, focal spike and wave activities, and focal or diffuse slowing_, inspired by the SCORE framework [3].

To evaluate the model's real-world utility, we compared model outputs to the established clinical workflow. In this workflow, recordings are first annotated and reported by specialized neurophysiology assistants, with the final clinical reports validated and signed off by a senior neurophysiologist. A dataset of 347 recordings from the BrainCapture operation collected in the period from November to December 2025 was used to assess the level of agreement between the automated AI screening and the official clinical reports.

[1] W.-B. Jiang, L.-M. Zhao, and B.-L. Lu, “Large Brain Model for Learning Generic Representations with Tremendous EEG Data in BCI.” [Online]. Available: https://arxiv.org/abs/2405.18765

[2] A. Gjølbye, L. Skerath, W. Lehn-Schiøler, N. Langer, and L. K. Hansen, “Speed: Scalable Preprocessing of EEG Data for Self-Supervised Learning,” in 2024 IEEE 34th International Workshop on Machine Learning for Signal Processing (MLSP), IEEE, Sept. 2024, pp. 1–6. doi: 10.1109/mlsp58920.2024.10734749.

[3] S. Beniczky et al., “Standardized computer-based organized reporting of EEG: SCORE – Second version,” Clinical Neurophysiology, vol. 128,no. 11, pp. 2334–2346, 2017, doi: 10.1016/j.clinph.2017.07.418.

Results

The transformer model demonstrated high diagnostic utility in the Kenyan clinical environment. For the online detection of abnormal EEG recordings, the transformer model achieved a sensitivity of 84%, a specificity of 65%, an F1-score of 79%, and an AUROC of 83%. These results suggest a clear potential in reduction of expert workload, and preliminary feedback from EEG technicians indicates that the "screening" flag could serve as a valuable quality-control tool during the interpretation process.

Conclusions

The successful deployment of a transformer-based AI model in a shadow operation demonstrates the feasibility of automated EEG screening in emerging markets. Our results indicate that AI can achieve the diagnostic performance necessary to assist technicians and prioritize recordings for expert review, effectively narrowing the diagnostic gap. This work paves the way for a hybrid diagnostic model where AI-assisted screening empowers local healthcare workers, ensuring that specialized neurological expertise is directed where it is most needed. Future research will focus on the clinical impact of moving the model from shadow operation to active decision support.

P64
Energy-Efficient Epileptic Seizure Prediction Using Spiking Neural Networks

A. Brady[1], H. Daoud[1]

[1]Augusta University - Augusta (United States)

Introduction

Accurate prediction of epileptic seizures from scalp electroencephalography (EEG) remains a critical challenge in epilepsy management, particularly for enabling timely intervention and improving patient quality of life. While recent Deep Learning (DL) approaches have demonstrated strong seizure prediction performance, their high computational and energy demands limit practical deployment in wearable and edge-based systems. Spiking Neural Networks (SNNs), which process information using sparse, event-driven spike activity, offer a biologically-inspired alternative with the potential for significantly reduced energy consumption. The cost of this reduction in energy, however, tends to be a loss in prediction accuracy. This study investigates whether an SNN-based architecture can achieve high patient-specific seizure prediction accuracy from scalp EEG, comparable to that of leading DL models.

Methods

A patient-specific SNN model was developed and evaluated using the CHB-MIT scalp EEG database. Continuous EEG recordings were segmented into non-overlapping 5-second windows and labeled as pre-ictal or inter-ictal using a one-hour pre-ictal horizon. Class imbalance was addressed through equal sampling of pre-ictal and inter-ictal segments. The proposed architecture integrates two spiking convolutional layers with Leaky Integrate-and-Fire (LIF) neurons for spatial feature extraction across EEG channels, a spiking recurrent Long Short-Term Memory (LSTM) layer to capture temporal dependencies, followed by fully-connected layers and LIF neurons for producing spike trains. Surrogate-gradient backpropagation was employed to enable end-to-end training of the spiking components. Model performance was evaluated using seizure-wise leave-one-out cross-validation across seven subjects, with accuracy, sensitivity, specificity, F1-score, and area under the receiver operating characteristic curve (AUROC) reported.

Results

Across the seven-patient cohort, the proposed SNN achieved a mean accuracy of 93.60%, sensitivity of 88.91%, specificity of 98.29%, F1-score of 91.85%, and AUROC of 0.9360 for pre-ictal state classification. Training and validation curves demonstrated stable convergence without evidence of overfitting. When compared with recent DL and SNN-based seizure prediction methods evaluated on the same dataset, the proposed model achieved competitive predictive performance.

Conclusions

The results demonstrate that SNNs can achieve high patient-specific seizure prediction performance from scalp EEG while offering architectural and computational advantages relevant to low-power deployment. The combination of spiking convolutional feature extraction and temporal modeling enables effective spatiotemporal EEG representation with fewer operations than conventional DL models. This work supports the feasibility of SNN-based seizure prediction for future wearable and edge-based systems. Ongoing work will focus on deployment on neuromorphic hardware, direct energy consumption measurements, and validation on larger and more diverse patient cohorts.

P65

Network Analysis of the Seizure Onset Zone in Epilepsy Surgery: A Multi-Institutional Study

K. Nix[1], K. Walsh[1], O. Feys[1], W. Ojemann[1], B. Scheid[1], E. Conrad[1], N. Sinha[1], A. Khambati[2], J. Kleen[2], B. Litt[1], V. Rao[1], D. Friedman[3], K. Davis[1]

[1]University Of Pennsylvania - Philadelphia (United States), [2]University Of California San Francisco - San Francisco (United States), [3]New York University - New York (United States)

Introduction

Surgical interventions for drug-resistant epilepsy, including responsive neurostimulation (RNS), resection, and laser ablation, show variable treatment outcomes, with 50-70% of patients achieving seizure reduction. Identifying biomarkers that predict treatment response is critical for patient selection and therapeutic optimization. We investigated whether ictal network properties of the seizure onset zone (SOZ) electrodes differentiate treatment responders from non-responders.

Methods

We analyzed ictal intracranial EEG data from patients with drug-resistant epilepsy who underwent RNS implantation at three institutions: Hospital of the University of Pennsylvania, University of California San Francisco, and New York University. Using coherence-based functional connectivity during seizures, we computed five graph-theoretical metrics from SOZ electrodes: synchronizability, global transitivity, global efficiency, node strength, and eigenvector centrality.

To address network size bias, we performed a three-phase analysis:

Phase 1: Tested whether metrics correlate with network size by resampling non-SOZ interictal electrode activity at ten fixed network sizes (4-25 electrodes, 100 resamples per size).

Phase 2: Standardized ictal SOZ networks by resampling all patients to four electrodes (100 iterations) for size-matched comparisons.

Phase 3: Evaluated full ictal SOZ networks if Phase 1 showed minimal size effects.

We compared metrics between groups using Mann-Whitney U tests and utilized Random Forest classification with nested leave-one-out cross-validation to predict treatment response. Model performance was assessed via ROC analysis with bootstrap confidence intervals. We used DeLong's test to compare network-based versus clinical models and permutation testing to assess whether models performed better than chance.

Results

Phase 1: Analysis is ongoing. If Phase 1 analysis reveals significant network size dependence, size-standardized comparisons will be essential for valid inference.

Phase 2: In SOZ subnetworks resampled to four electrodes, node strength was most variable across resamples (mean coefficient of variation (CV) 0.511), while global efficiency (mean CV 0.153) and eigenvector centrality (mean CV 0.252) were more stable.

Phase 3 (N=36, 21 responders, 15 non-responders): Mann-Whitney U tests show no difference in SOZ ictal network metrics between responders and non-responders, suggesting that ictal network topology from the SOZ itself may not contain prognostic information for RNS therapy. In this initial cohort, Random Forest Classifier the model with ictal network metrics features showed modest classification ability (AUROC = 0.610) compared to the model with clinical features alone, whose performance was limited by sample size or near-chance (AUROC = 0.333, 95% CI [0.000, 0.750], p = 0.776).

Conclusions

Preliminary results suggest that ictal network topology of the SOZ alone does not strongly distinguish RNS responders from non-responders. However, multivariate models using network features show modest predictive value and may complement clinical variables. Plans include expansion to patients undergoing surgical resection and laser ablation in future analyses. Ongoing analyses will test generalizability across institutions and assess whether integrating network and clinical data improves prognostic accuracy for personalized epilepsy surgery planning.

P66

Towards an EEG you can talk to: Enabling Large Language Models to natively process EEGs

D.D. Kranz[1], R. Terziev[1], C. Meisel[1]

[1]Berlin Institute Of Health And Department Of Neurology, Charité – Universitätsmedizin Berlin - Berlin (Germany)

Introduction

Clinical EEG interpretation is a high-dimensional reasoning task: clinicians integrate evolving rhythms, reactivity, artifacts, vigilance, and intermittent pathological patterns into a narrative report with differential diagnoses and actionable recommendations. Most EEG AI systems, however, collapse this complexity into narrow classifiers (e.g., seizure yes/no), limiting trust, generalization, and clinical usefulness. We introduce EEG2Vec, a representation-first approach that learns structured latent embeddings from large-scale unlabeled EEG as a prerequisite for interactive, language-based clinical assistants.

Methods

EEG2Vec is a wav2vec2-style self-supervised encoder for EEG Continuous recordings are segmented into 30-second windows that yield 240 latent tokens per window, each represented as a 512-dimensional embedding. The encoder is trained with masked modeling and a contrastive objective (NT-Xent), encouraging contextual representations that discriminate true masked targets from negatives while remaining robust to common EEG variability. After pretraining, we assess the structure of the learned representation space using lightweight downstream probes. These probes are not intended as task-optimized classifiers, but as diagnostic readouts that quantify how readily clinically relevant information can be accessed from the frozen EEG2Vec tokens. We therefore select a set of routine annotation labels (eyes open vs. eyes closed, photic stimulation, hyperventilation, movement, and swallow events) as convenient, well-defined signals that are known to modulate EEG in stereotyped ways.

All probes are trained on a frozen encoder with minimal capacity, such that performance reflects the degree to which relevant information is already linearly or shallowly separable in the pretrained token space, rather than the expressive power of the probe itself.

Results

We trained EEG2Vec on 3 million 30-second EEG segments from routine clinical recordings. Using the frozen encoder and a lightweight probe as an information readout, we observed a balanced accuracy of 0.49 across six routine conditions after 30 epochs of self-supervised pretraining. This indicates that EEG2Vec tokens make multiple forms of clinically relevant structure

readily accessible without task-specific fine-tuning, rather than reflecting optimization for any particular downstream label.

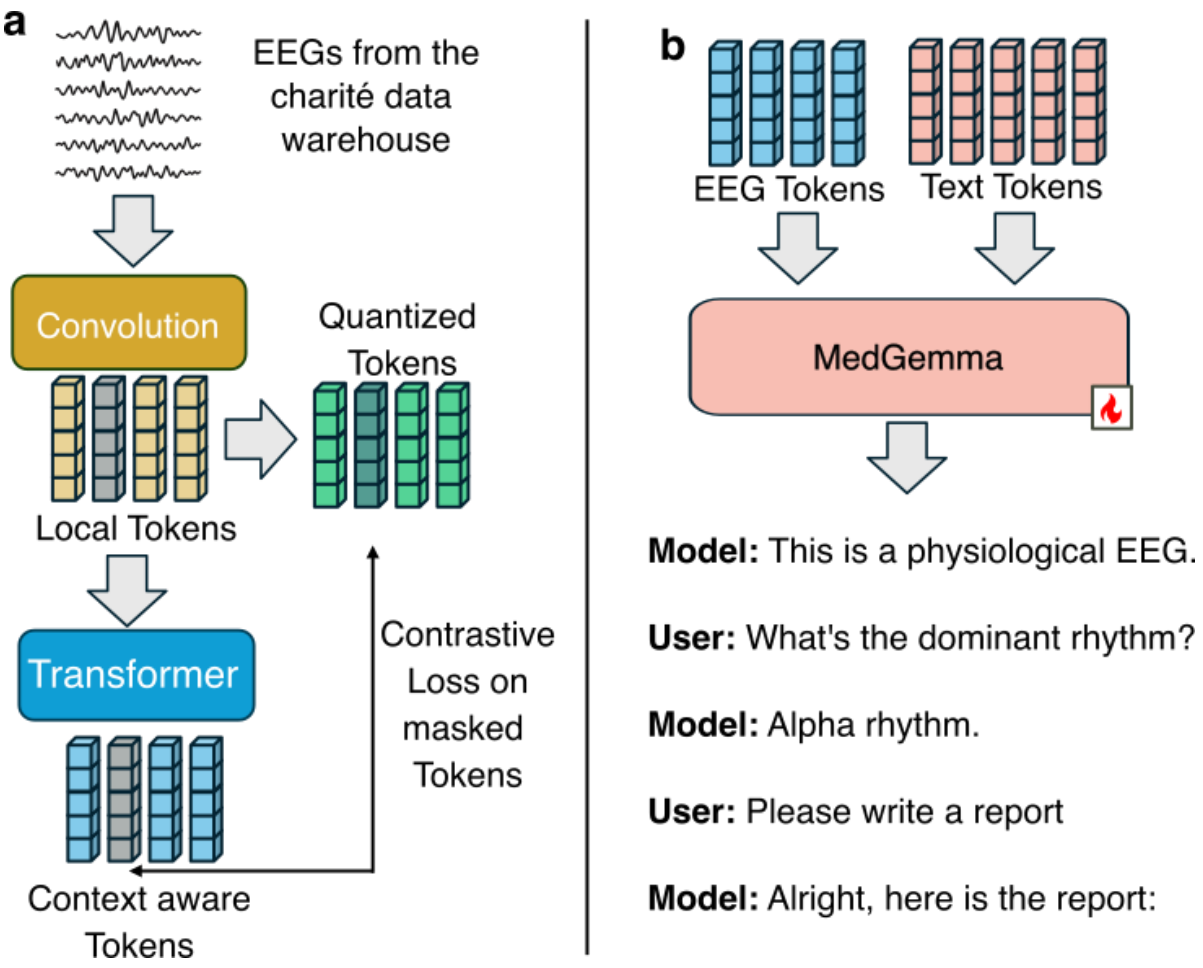
Interestingly, downstream discriminability improved substantially with longer pretraining even though the contrastive loss plateaued early. Balanced accuracy increased from 0.33 after 5 epochs to 0.45 after 15 epochs and 0.49 after 30 epochs, suggesting that continued training refines the information content of the contextual tokens despite only minor changes in the self-supervised objective.

Conclusions

Our results establish a scalable foundation for "EEG-native" AI: a self-supervised encoder that converts continuous clinical EEG into structured token sequences that preserve morphology and temporal context. This representation can serve as the modality adapter required to connect EEG to modern generative models.

As an outlook, we aim to extend the "LLM-to-VLM" paradigm to clinical neurophysiology. Analogous to how vision-language models enable an LLM to reason over images by introducing a learned visual tokenizer/encoder, we envision a Clinical Copilot that enables an LLM to reason over EEG by consuming EEG2Vec tokens alongside text. Such a system would support interactive questioning (e.g., dominant rhythm, reactivity, artifact vs. pathology, state changes) with explicit provenance to time-stamped EEG evidence, and would draft conventional EEG reports as a natural downstream product. While this copilot layer is future work, EEG2Vec provides the key prerequisite: a robust, language-like EEG representation on which interactive, evidence-grounded clinical assistance can be built.

Fig. 1 Clinical Copilot overview. (a) Self-supervised pretraining: raw EEG recordings from the Charite data warehouse are converted into local tokens via a ´ convolutional frontend, contextualized by a Transformer into context-aware tokens, and discretized into quantized tokens. Training uses a contrastive loss on masked tokens to learn robust, physiologically meaningful representations. (b) Outlook: EEG tokens are combined with text tokens as input to an LLM (here illustrated with MedGemma), enabling interactive, evidence-grounded dialogue about the recording and automated generation of structured EEG reports.



https://secure.key4events.com/onsite/api.aspx?e=1960&img=g4.png&ai=78420&op=getabstractimg&dirN=0

P67
New epilepsy insights from leveraging AI and LLMs

M. Richardson[1], J. Winston[1], J. Davies[1], Y. Gan[1], S. Barlow[1], J. Teo[1]

[1]King's College London - London (United Kingdom)

Introduction

Our research team in the School of Neuroscience at King's College London has pioneered the use of cutting-edge, generative Large Languagew Models (LLMs) to develop new methods to extract key epilepsy data from Electronic Health Records (EHRs). In turn, we have used that data to develop a prototype prognostic tool that can assist clinicians in deciding what antiseizure medications (ASMs) are best for an individual patient. Epilepsy affects about one per cent of the population. About one in every three people with epilepsy do not respond to ASMs -- the main form of treatment -- and are therefore regarded as refactory. In the United Kingdom, more than 30 ASMs are available for prescription, so the potential number of combinations of ASMs is very large. Our project applies innovative machine learning and AI methods to investigate epilepsy patient records at scale with the ultimate goal of discovering ASM combinations that can either eliminate or reduce seizures for refactory patients. This abstract consists of an overview of the project and our key achievements.

Methods

Our research project, which commenced in 2023, consists of two parts. The first part involves utilising new LLMs to extract key epilepsy data from text-based, unstructured outpatient clinic letters, which are written by clinicians and nurses. The essential problem with these letters is that they are completely unstructured, following no template, and frequently omit key information. So far we have analysed about 90,000 letters for more than 10,000 individual epilepsy patients, from two London hospitals, King's College Hospital and Guy's and St Thomas' Hospital. Soon we will add data from a third hospital, University College London Hospital. Key data for each patient we collect includes: epilepsy type and seizure type, current ASMs the patient is taking, if the patient is having seizures, and if so, the frequency of seizures, or whether the patient is seizure free and for how long, comorbidities, and symptoms associated with taking ASMs. Over the past two years in particular we have continued to refine our LLM models by using newer (and often larger) open-source models as they become available. LLMs we have used include Meta's Llama models, Mistral's LLMs, and Qwen, among others. We have experimented with various merhods, including prompting, chain-of-thought reasoning, and fine-tuning LLMs. Our fine-tuning and evaluation of LLMs was made possible by annotating manually thousands of ourpatient letters. A more recent innovation of ours involved generating synthetic outpatient letters -- by using OpenAI's GPT5 -- to fine-tune the LLMs. We found that by using the synthetic letters either as a substitute or complement to the real letters impoved the performance of the fine-tuned LLMs in some instances. The second part of our peroject involves developing a multivariate prediction model that utilises the structured key data that we have extracted from the outpatient letters. We train the prediction model by using several classical machine learning algorithms. As a result, we are able to identify combinations of ASMs that are more likely to assist refactory patients in improving their seizure status.

Results

For the extraction of key epilepsy data using LLMs we achieved the best results for relatively straightforward data: epilepsy type, seizure type, and current ASMs. For these we achieved F1 scores of 0.9 or more. However, we found it significantly more difficult to fine-tune LLMs for seizure freqency and seizure freedom. This was because only a minority of outpatient letters contain information about these cartegories, and because the letter writers (especially clinicians) are often vague or imprecise about the nature of these seizures. More recently, we experimented with generating synthetic letters to fine-tune LLMs for seizure frequency and seizure freedom and to date have achieved F1 scores of 0.8 or more. Comorbities and associated symptoms are even more problematic due to the paucity of data in the underlying dataset. For the multivariate prediction model, we found that three tree-based prediction models outperformed Supporrt Vector Machine; they were XGBoost, Random Forest, and AdaBoost. Our top F1 score was 0.78 with a 95% confidence interval of 0.68 to 0.87. Over time, with beter quality data extracted with newer, more improved LLM models, we expect to improve the performance of our prediction model.

Conclusions

Prediction models built on data that originates from free-text, unstructured EHRs and which is extracted by new, generative LLMs show promise in being able to determine whether a person with epilepsy is likely to be a treatment responder based on the ASMs they are taking. With further development, we are aiming to put our prediction model in a real-world clinical setting for beta testing some time in 2026.

Authors index :

A

A. Sanchez L. P57
Abdallah C. P54
Abend N.S. P36
Adenekan T. P06
Aguila C.A. P51
Aguila C. P06, P45
Ajay P. P09
Al-Haj Mustafa S. P21, P22
Alsammani A. P44
Ambroise C. P33
Anand S. P36
Arkoun G. P10
Armenta Salas M. P37
Asuncion J. P06
Atienza Alonso D. P43
Aung T. P60
Aurlien H. P54
Avigdor T. P54

B

Bagić A. P50
Bardyn C.E. P43
Barlow S. P67
Baudouin F. P33
Baumgartner C. P43
Bebarova M. P15
Beniczky S. P43, P54
Benninger F. P29, P42
Bernasconi A. P56
Bernhardt B. P56
Bernini A. P41, P43
Beuchat I. P43
Boerwinkle V. P61
Bolton J. P30, P32
Bolus H. P09
Bonilha L. P50
Bosque-Varela P. P54, P60
Bou Assi E. P19, P31
Brady A. P64
Brazdil M. P04, P12, P15
Brichtova E. P15

C

Cacic Hribljan M. P54
Cao T. P14
Carlson D. P60
Carrasco-Gómez M. P24
Carreño M. P46

Casadio M. P40
Castelnovo A. P40
Centeno M. P46
Chambon S. P33
Chang E. P16, P51, P52
Charles A.S. P18
Chellappa R. P11
Chen J. P56
Cho J. P13
Chu C. P07
Chybowski B. P05, P08
Cimbalnik J. P15, P17, P27
Cohen A. P42
Concha L. P56
Conde E. P46
Conrad E.C. P51
Conrad E. P06, P45, P53, P65
Cordani R. P40
Cronk M. P11

D

D'onofrio G. P31
Daniel P. P12
Daoud H. P64
Daraie A.H. P18
Das S. P44
Davies J. P67
Davis K. P06, P50, P53, P65
Davis K.A. P51
Dekraker J. P56
Dessevres E. P10
Devinsky O. P25
Dolezalova I. P12, P60
Donaire A. P46
Drane D. P50
Dubeau F. P27
Dugan P. P50

E

Eagles J. P39
Ellis C. P52
Escudero J. P05, P08

F

Fabricius M.E. P43
Fallah A. P01
Fan J. P16
Feys O. P65
Fisher K. P08
Frauscher B. P27, P54, P60
Friedman D. P65

G

Galindo Lazo D.A. P31
Gan Y. P67
Gao W. P48
Gao F. P48
Geller A. P28
Gliske S. P44
Gloaguen O. P33
Goffinet L. P43
Goldberg I. P29, P42
Goldblum Z. P51
Gonzalez Sulser A. P05
Gonzalez-Martinez J. P47, P61
González-Martínez J.A. P18
Gonzalez-Sulser A. P08
Goodall-Halliwell I. P56
Gotman J. P54, P60
Gotshalk A. P47
Gould L. P25
Granados Monge F. P55
Guerrero Aranda A. P62
Gupta A. P36, P37, P38

H

Hadjinicolaou A. P01
Hahn C. P36
Halamek J. P15
Hammer L.H. P51
Harikar M. P25
Hays M. P47
Helmstaedter C. P21, P22
Hendrych M. P15
Hopp J. P47
Hops_collaborators H. P01
Hribkova H. P15
Hrtonova V. P12
Huberfeld G. P10

I

Imbach L. P43

J

Jaber K. P60
Jahani A. P31
Jahromi S. P30, P32
Jang H.R. P13
Jansen A. P21, P22
Jiang T. P48
Jing J. P06
Johnson S. P47
Jordan B. P08

Juan O. P09
Juarez-Martinez E. P39
Juliano C. P36
Jung K.Y. P26
Jurak P. P12, P15
Jürgensen L. P03

K

Kamali G. P47, P61
Kamousi B. P36, P37, P38
Kang J.Y. P18, P49, P61
Kang J. P20
Karpychev V. P50
Karunakaran S. P37
Khambati A. P65
Khambhati A. P16
Khoshkhou M. P49
Kim H. P13
Kim S.S. P26
Kim J. P06
Kimchi E. P39
Kleen J. P58, P65
Klimes P. P12, P15, P17, P27, P60
Knix K. P53
Kogan A. P29
Kolajova M. P15
Kook M.H. P13
Kopytek-Beuzen M. P10
Koren J.P. P43
Korzun J. P52
Kossoff E. P11
Kramer D. P28
Kramer M. P07, P35
Kranz D.D. P66
Krüger B. P21, P22
Krumm L. P34
Kuhlmann L. P25
Kunert P. P10
Kuzniecky R. P50

L

La Vega-Talbott M. P36
Lahtoo S. P43
Laperriere S.O. P01
Lariviere S. P56
Laxpati N. P50
Le Van Quyen M. P10
Lehn-Schiøler W. P63
Lemoine E. P19
Lesser R.P. P49
Lhatoo S. P24
Li J. P60
Litt B. P06, P45, P51, P52, P53, P65

Little T.Y. P38
Lo Bianco G. P33
Lotfi N. P24

M

Madsen J.R.M. P30
Makaram N. P55
Mansilla D. P27, P54
Martinez Ruiz L.C. P62
Matouskova B. P12
Medina N. P46
Meisel C. P03, P34, P66
Meuwly S.C. P41
Milan B. P17
Millan Sandoval J.P. P31
Min J. P39
Miranda R. P47
Moon J.H. P26
Moreno J. P10
Moro M. P40
Mosher J. P24
Moye M. P60
Müller P. P03

N

Natarajan P. P05
Neda B. P56
Nejedly P. P12, P15, P27
Ngo A. P56
Nguyen D.K. P19, P31
Nhu D. P25
Nix K. P65
Nix K.C. P51
Nobili L. P40

O

Odone F. P40
Ojemann W.K.S. P51
Ojemann S. P28
Ojemann W. P06, P53, P65
Oliver D. P09

P

Pail M. P12, P17
Pandiaraju S. P20
Papadelis C. P30, P32
Parashos A. P50
Pardoe H. P50
Paredes O. P62
Partamian H. P30, P32
Parvizi J. P37, P38

Pazuelo J. P10
Pearl P.L. P30, P32
Pedersen M.G. P63
Peh I. P05, P08
Pendelio-Boyer M. P33
Pereira E. P61
Perry M.S. P30, P32
Pielas M. P10
Pijackova K. P27
Proserpio P. P40
Provini F. P40
Puranik T. P36, P37, P38
Pyrzowski J. P10

R

Ramot M. P42
Rao V. P16, P65
Ravisankar A. P59
Razmara A. P28
Ren G. P47
Reyes J.N. P10
Rheims S. P43
Rhode L. P11
Ribière L. P33
Richardson M. P67
Rivera G. P11
Robert M. P31
Robin A. P31
Roccaforte Z. P28
Rodriguez-Cruces R. P56
Romeo A. P09
Romo Vázquez R. P62
Ros-Castelló V. P54
Roth R. P50
Rowling J. P05
Ryvlin P. P41, P43

S

Saad C. P59
Saber R. P19, P31
Sabino G. P55
Saboo K. P16
Sahlas E. P56
Said R.R. P36
Sanchez L.S. P18
Sarma S.S. P20
Sarma S. P47, P61
Sarma S.V. P18, P49
Sarzo Wabi I. P31
Sassi F. P40
Satzer D. P28
Scheid B. P53, P65
Schomer D. P54

Schrader D. P56
Schuele S. P39
Segura-Quijano F. P10
Shen J. P02
Shetty J. P05, P08
Shi S. P51
Shimshoni I. P29
Sido Bouzan N. P21, P22
Singh A. P36
Sinha N. P45, P51, P53, P65
Sivathamboo S. P25
Smith L. P05
Southwell D. P60
Spahr A. P41, P43
Sperling M. P47
Stacey W. P44
Stastna D. P23
Steininger M. P21, P22
Stone S.S.D. P30, P32
Storgaard A.M. P63
Streett M.R. P47
Stuart J. P60
Sullivan J.E. P36
Sun Q.Q. P14
Surges R. P21, P22
Svecova O. P15
Szuromi M. P07

T

Tadayon E. P16
Tamilia E. P30, P32, P55
Tassi L. P40
Tauste Campo A. P46
Teo J. P67
Terziev R. P03, P66
Thompson J. P28
Tozzi L. P36, P37
Travnicek V. P27
Tveit J. P54
Tyner K. P28

U

U. Onyike C. P57
United Kingdom Infantile Spasms Trust U. P05
Uscátegui A.M. P10

V

V. Sarma S. P57
Vakilna Y. P24
Vakilna Y.S. P43
Valderrama M. P10
Van Gompel J. P23

Vélez Pérez H. P62
Venegas M.A. P10
Vila-Vidal M. P46
Von Wrede R. P21, P22

W

Wadi L. P60
Walia A. P47
Walsh K. P51, P53, P65
Wang D. P61
Weil A. P01
Weng Y. P56
Westover M. P06
Winston J. P67
Wu H. P45
Wyeth D. P47

X

Xie K. P51, P52, P56
Xu Z.C. P45
Xu Z. P51

Y

Y. Kang J. P57
Yun H.J. P13
Yune I. P13

Z

Zamansky A. P29
Zatloukalova E. P15
Zhang Z. P56
Zhang C. P14
Zhang D. P58
Zhou D. P52
Zhu Y. P60