Clinical Neurophysiology 161 (2024) 1–9

Contents lists available at ScienceDirect

Clinical Neurophysiology

journal homepage: www.elsevier.com/locate/clinph

Timing matters for accurate identification of the epileptogenic zone

Bartlomiej Chybowski [a], Petr Klimes [b], Jan Cimbalnik [c], Vojtech Travnicek [b,c], Petr Nejedly [b], Martin Pail [b,d,e], Laure Peter-Derex [f,g], Jeff Hall [h], François Dubeau [h], Pavel Jurak [b], Milan Brazdil [d,e], Birgit Frauscher [h,i,*]

[a] University of Edinburgh, School of Medicine, Deanery of Clinical Sciences, 47 Little France Crescent, EH164TJ Edinburgh, Scotland
[b] Institute of Scientific Instruments of the CAS, v. v. i., Královopolská 147, 612 00 Brno, Czech Republic
[c] International Clinical Research Center, St. Anne's University Hospital, Pekařská 53, 602 00 Brno, Czech Republic
[d] Brno Epilepsy Center, Department of Neurology, St. Anne's University Hospital, Member of ERN-EpiCARE, Faculty of Medicine, Masaryk University, Kamenice 5, 625 00 Brno, Czech Republic
[e] Behavioral and Social Neuroscience Research Group, CEITEC Central European Institute of Technology, Masaryk University, Žerotínovo nám 617/9, 601 77 Brno, Czech Republic
[f] Center for Sleep Medicine and Respiratory Diseases, Lyon University Hospital, Lyon 1 University, 103 Grande Rue de la Croix-Rousse, 69004 Lyon, France
[g] Lyon Neuroscience Research Center, CH Le Vinatier - Bâtiment 462 - Neurocampus, 95 Bd Pinel, 69500 Lyon, France
[h] Montreal Neurological Hospital, McGill University, 3801 Rue University, Montreal, QC H3A 2B4, Quebec, Canada
[i] Department of Neurology, Duke University Medical School and Department of Biomedical Engineering, Pratt School of Engineering, 2424 Erwin Road, Durham, NC, 27705, USA

HIGHLIGHTS

- Short segment of iEEG can achieve similar results in localization of the epileptogenic zone as a model based on long recordings.
- Random selection of short iEEG segments may give rise to inaccurate results.
- It is important that the analyzed segment is carefully and systematically selected, preferably from NREM sleep.

ARTICLE INFO

Article history:
Accepted 1 January 2024
Available online 18 February 2024

Keywords:
iEEG
Epilepsy
Seizure

ABSTRACT

Objective: Interictal biomarkers of the epileptogenic zone (EZ) and their use in machine learning models open promising avenues for improvement of epilepsy surgery evaluation. Currently, most studies restrict their analysis to short segments of intracranial EEG (iEEG).
Methods: We used 2381 hours of iEEG data from 25 patients to systematically select 5-minute segments across various interictal conditions. Then, we tested machine learning models for EZ localization using iEEG features calculated within these individual segments or across them and evaluated the performance by the area under the precision-recall curve (PRAUC).
Results: On average, models achieved a score of 0.421 (the result of the chance classifier was 0.062). However, the PRAUC varied significantly across the segments (0.323–0.493). Overall, NREM sleep achieved the highest scores, with the best results of 0.493 in N2. When using data from all segments, the model performed significantly better than single segments, except NREM sleep segments.
Conclusions: The model based on a short segment of iEEG recording can achieve similar results as a model based on prolonged recordings. The analyzed segment should, however, be carefully and systematically selected, preferably from NREM sleep.
Significance: Random selection of short iEEG segments may give rise to inaccurate localization of the EZ.

1. Introduction

Resective epilepsy surgery is the therapy of choice in patients with focal drug-resistant epilepsy (Jobst and Cascino, 2015). In epilepsy surgery, precise localization of the epileptogenic zone (EZ) is crucial. The current gold standard for EZ identification in more complex cases is the seizure-onset zone (SOZ) that is approximated from the ictal intracranial electroencephalogram (iEEG) (Thomschewski et al., 2019). Unfortunately, this approach leads to seizure freedom in only 40–50% of well selected candidates, and takes on average 1–2 weeks of recording and evaluation (Krucoff et al., 2017).

* Corresponding author.
E-mail address: birgit.frauscher@duke.edu (B. Frauscher).

https://doi.org/10.1016/j.clinph.2024.01.007
1388-2457/

The most promising new markers of the EZ, which are derived from interictal iEEG recordings, are interictal epileptic discharges (IEDs) with gamma activity (Ren et al., 2015; Thomas et al., 2022), high-frequency oscillations (HFOs) (Frauscher et al., 2017; Jacobs et al., 2018; Sarnthein et al., 2021), and more recently relative entropy (Travnicek et al., 2023) as well as other connectivity metrics and various combinations of these iEEG features in machine learning models (Cimbalnik et al., 2019; Klimes et al., 2016; Lagarde et al., 2018; van Mierlo et al., 2014). We demonstrated that a combination of these features outperformed models based on a single feature and that using short segments of iEEG recordings across different states of vigilance achieved the best scores in non-rapid eye movement (NREM) sleep (Klimes et al., 2019).

The influence of circadian and sleep cycles on interictal and ictal activity in epilepsy has been known for decades (Bercel, 1964; Karoly et al., 2021). The latest results suggest, that variations across the sleep-wake cycle in long recordings are not negligible, when assessing markers of the EZ. For example, the study by Spencer *et al.* shows that epileptiform activity has a strong 24 h periodicity and varies depending on the localization of the SOZ (Spencer et al., 2016). Baud *et al.* demonstrated that the phase information from circadian and multidien interictal epileptiform activity rhythms is a novel biomarker for determining relative seizure risk (Baud et al., 2018). Gliske *et al.* shows that HFO interpretation requires the analysis of prolonged recordings rather than isolated review of short data segments (Gliske et al., 2018). And finally, in our last paper, we provided evidence that epilepsy surgery outcome depends on strong and single sources of IEDs which are relatively stable over an at least 18-hour long period (Klimes et al., 2022). All these studies discourage us from using randomly selected short segments of EEG. However, the potential benefit of EEG feature variance over prolonged periods of time in machine learning models has not been systematically investigated.

In this study, we processed 2381 hours of iEEG data obtained from 25 patients with good postsurgical outcome and tested different iEEG features, feature selection strategies and machine learning models for interictal EZ localization in multiple, individually-selected 5-minute long segments across different sleep stages and interictal conditions. Then, we created a set of new features, calculated as the average and variance of iEEG features over a prolonged period, and tested whether these "variance features" outperform models based on features from single segments. We hypothesized that (i) the score of the models based on 5-minute segments varies across different segments; (ii) the model based on average and variance of iEEG features achieves a better score than a model based on average or variance alone; and that (iii) the model based on a well-selected 5-minute long segment, such as 5-minutes of NREM sleep, can achieve the same result as the model based on the data obtained from $\geq$ 24-hour recordings.

2. Methods

2.1. Patients

We analyzed all consecutive adult patients with drug-resistant focal epilepsy who underwent stereo-electroencephalography (SEEG) and subsequent resective surgery with a good post-surgical outcome (Engel I (Engel, 1993)) after a minimum follow-up period of $\geq$ 1 year at the Montreal Neurological Institute & Hospital (MNI) between 1/2010 and 12/2015 and at the St. Anne's University Hospital in Brno (SAUH) between 4/2017 and 10/2019. The patients undergoing SEEG prior to this period did not have additional scalp EEG with subdermal thin wire electrodes required for sleep staging. Inclusion criteria were: (a) high-resolution 3D magnetic resonance imaging (MRI) datasets; (b) availability of $\geq$ 24 hours of continuous SEEG recording; and (c) scalp EEG, electro-oculography, and electromyography or video for sleep scoring. The study was approved by the MNI and SAUH Ethics Review Board. All patients granted written informed consent.

2.2. Recordings

Standard clinical SEEG depth electrodes were inserted stereotactically using an image-guided system. SEEG was recorded at 2 kHz (MNI) and 5 kHz (SAUH). An average from all SEEG contacts with a confirmed location in brain structures was used as an SEEG reference signal. The white matter was not excluded from the analysis. Scalp EEG was obtained with subdermal thin wire electrodes (MNI) or scalp electrodes (SAUH) at F3, Fz, F4, C3, Cz, C4, P3, Pz, and P4 (Ives, 2005).

2.3. Data selection and pre-processing

For every patient, we selected the first 1–3 available days of continuous SEEG, ignoring 1 hour before and 1 hour after epileptic seizures. Electrode contact localization was determined according to methodology described by Frauscher *et al.* (Frauscher et al., 2018). The seizure-onset zone contacts were identified based on the earliest ictal changes at seizure onset irrespective of the fast activity content (Spanedda et al., 1997). Sleep was scored visually in 30 second epochs (Berry et al., 2017; Ives, 2005).

2.4. Segment selection

For every patient, we selected 16 different, 5-minute long segments of SEEG, detailed in Table 1. For the different states of vigilance, 5 minutes of continuous states were preselected. If not available, 5 minutes with at least 80% (4 minutes) of the defined state were selected. Every segment was carefully visually inspected for artefacts and excessively large IEDs which could affect the common reference signal. In case of Wake, NREM and random segments, we selected three different segments per state. We were not able to select more than three random, artefact-free segments for each patient, because two patients in the dataset had disturbed sleep by nocturnal awakenings or seizures which reduced the number of segments. In case of fragmented sleep, we were able to select only one random segment per patient.

2.5. Calculation of minimum and maximum IED rate segments

IEDs were detected using a validated detector on all SEEG channels irrespective of cortical or gray matter localization (Janca et al., 2015). To reduce false-positive detections caused by alpha rhythm or spindles, the algorithm was modified to ignore all detected events separated from each other by less than 300 msec. To reduce detections caused by artefacts, all detected events which appeared at the same time in $\geq$ 50% of SEEG signals were ignored (Klimes et al., 2022). A random sample of final detections was visually cross-checked for plausibility by a neurophysiologist. Further, we verified that the ranking of the top 20% of IED channels per patient corresponded to the localization of IEDs as per SEEG report.

Each recording was divided into 5-minute long segments with no overlap. For each segment, the average IED rate across all SEEG channels was calculated. The minimum and maximum values of IED rates were defined as the 5th and 95th percentiles from the whole distribution.

2.6. Calculation of minimum entropy segments

Each recording was divided into 5-minute long segments with no overlap. In every segment, IED rates for individual SEEG

Table 1
Selected time segments of stereo-electroencephalography (SEEG) recording. Legend: REM = rapid eye movement; NREM = non rapid eye movement; IED = interictal epileptic discharges.

Segment name	Definition	Number of analyzed segments
Random	Random interictal segment	3
Random Wake	Random interictal segment during Wake period	3
Random NREM	Random interictal segment during either NREM N2 or N3 sleep	3
Random N2	Random interictal segment during NREM N2 sleep	1
Random N3	Random interictal segment during NREM N3 sleep	1
Random REM	Random interictal segment during REM sleep	1
1st 5 minutes of N2	The first 5 minutes of interictal N2 sleep during the first night of the recording	1
Maximum IED rate[1]	Interictal segment with maximum IED rate, without any other pre-selection	1
Minimum IED rate[1]	Interictal segment with minimum IED rate, without any other pre-selection	1
Minimum entropy[2]	Interictal segment with minimum entropy, without any other pre-selection	1

[1] Detailed definition in "Calculation of minimum and maximum IED rate segments".
[2] Detailed definition in "Calculation of minimum entropy segments".

contacts were calculated and aligned by their anatomical location (agnostic to inter-electrode distance/electrode coordinates), resulting in an IED spatial distribution (Klimes et al., 2022). Then, the IED rates across SEEG contacts were normalized in the range of `<0,1>`, and the entropy of the IED distribution was calculated as: `-sum((x*log2(x))/log2(n))`, where `x` is the normalized IED spatial distribution and n is the number of SEEG contacts. The minimum entropy value, representing the segment with the most distinctive and localized peak of elevated IED rates in IED spatial distribution (Klimes et al., 2022), was defined as the 5th percentile from the whole distribution. Please note that the IED rates were normalized in the range of `<0,1>` just for computation of the entropy. For further analysis, not normalized rates were used.

2.7. Feature calculation

Multiple univariate, bivariate and event-related EEG features were calculated for each SEEG contact (or pair of SEEG contacts) in every 5-minute segment separately using our open-source python library EPYCOM (Cimbalnik and SweetVlad, 2020), as previously used in (Cimbalnik et al., 2019; Klimes et al., 2019). In case of univariate features, the features were calculated for a single SEEG contact. The univariate features were: spectral power, power spectral entropy, phase-amplitude coupling, frequency-amplitude coupling, low-frequency ratio, Hjorth complexity, phase synchrony, phase consistency, and phase lag index. In the case of bivariate features, the features were calculated between adjacent SEEG contacts on the same depth electrode. The bivariate features were: linear correlation, coherence, and relative entropy. The features were calculated in the following frequency bands: delta (1–4 Hz), theta (4–8 Hz), alpha (8–12 Hz), beta (12–20 Hz), gamma (20–45 Hz), high-gamma (65–80 Hz), ripples (80–250 Hz), fast-ripples (250–600 Hz) and for the raw unfiltered recordings (1–500 Hz for MNI, 1–1200 Hz for SAUH). Event-related features were calculated for single SEEG contacts, namely IED rates (already detected by Janca *et al.* detector (Janca et al., 2015)), high-frequency oscillations rates in ripple (80–250 Hz), and fast-ripple (250–500 Hz) bands. High-frequency oscillations were detected by the validated detector by von Ellenrieder *et al.* (von Ellenrieder et al., 2016, 2012). Furthermore, for each patient, we calculated the average and variance of each feature across all 5-minute segments.

2.8. Feature selection

We conducted four different comparisons of models trained with (i) single-segment features, (ii) average of features across all segments (AVG), (iii) variance of features across all segments (VAR), and (iv) average and variance of features across all segments (VAR_AVG). We decided to run all four comparisons so that our results would be complete, exhaustive and leave no ambiguity. Each comparison required respective feature selection from appropriate sets.

To identify the best possible features, we utilized three of the most explainable feature-selection methods, namely Logistic Regression (LR) which achieves the best fit by adjusting the decision threshold, support vector regression (SVR) which minimizes the difference between the observed value and the estimated value, and ElasticNet (EN) which creates a simpler and more generic model by adding penalties for making it too specific. Ultimately, each function will eliminate those features that do not contribute to the prediction significantly. We adjusted the hyperparameters of all three models using RandomizedSearchCV to achieve the best-performing features for each.

In the next step, we selected the best features using Recursive Feature Elimination Cross-Validation (RFECV) and the previously prepared models. RFECV function created subsets of the best-performing features by repetitively removing the least relevant ones to the model. Then, based on the cross-validation score, it chooses the best group. All functions and models in further analysis are implemented with usage of the scikit-learn library (Pedregosa et al., 2011).

2.9. Training and validation of the models

Calculated features were further used in machine learning models for localization of the EZ, defined as resected SOZ contacts in patients with a good post-surgical outcome. We tested two robust and fully explainable algorithms: the Support Vector Classifier (SVC) and Logistic Regression. This step required training of 114 models in total (2 classification models × 3 feature selection models × 19 segments (16 for single-segment and one each for AVG, VAR, and VAR_AVG). We used a RandomizedSearchCV for hyperparameter tuning and Precision-Recall Area Under the Curve (PRAUC) as an evaluation metric. PRAUC was chosen over the F1-score, since it was shown to be more informative in imbalanced datasets (Powers and Ailab, 2011; Saito and Rehmsmeier, 2015). Each model was cross-validated using the leave-one-patient-out approach. We determined the baseline result for PRAUC by calculating the ratio of positive (`P, n=165`) and negative (`N, n=2508`) cases as `baseline=P/(P+N)`. For the purpose of the evaluation of the feature importance, we used coefficients of each model normalized between −1 (the least useful) and 1 (the most useful).

2.10. Statistical analysis of the results

To confirm the hypothesis, that the performance of individual segments is not stable, we compared results achieved by individual

segment models with each other. To test our second hypothesis, that the combination of average and variance of iEEG features achieves a better score than a simple average or variance, we compared models trained only on AVG, VAR, or VAR_AVG features. For the third hypothesis, that a well-selected short recording segment might be sufficient to predict the EZ, we chose the best performing VAR_AVG model and compared it with single-segment models.

For the purpose of the evaluation, we defined True Positives (TP) as electrode contacts marked by a model as a target which were SOZ and resected; False Positives (FP) as electrode contacts marked as target which were not SOZ and resected; True Negatives (TN) as electrode contacts not marked as target which were not SOZ and resected; and False Negatives (FN) as electrode contacts not marked as target which were SOZ and resected. We used the PRAUC metric for all the comparisons and checked if the differences were statistically significant using the Hanley & McNeil test (Hanley and Mcneil, 1982). We used Bonferroni multiple-comparison correction to minimize *Type 1* error. Statistically significant (<0.05) and highly statistically significant (<0.001) p-values after correction for 16 comparisons equal 0.003 and 0.0000625 respectively. The code used for analysis is publicly available (Chybowski, 2023).

3. Results

3.1. Patient demographic characteristics

In the study periods at both sites, a total of 50 patients (42% of total patients undergoing surgery after SEEG) achieved a good surgical outcome (Engel 1). Twenty-five of these patients were excluded because they had missing scalp EEG, electro-oculography, and electromyography or video as needed for sleep scoring (N = 11); missing high-resolution imaging for electrode co-registration (N = 9); or less than 24 hours of continuous recording available for this analysis (N = 5). The final patient cohort consisted of 25 consecutive patients fulfilling the selection criteria (Brno = 8, Montreal = 17). They consisted of 11 females with a median (IQR) age of 28 (13.5) years. A total of 24% had temporal epilepsy; 76% were lesional cases. The two most frequent pathologies were focal cortical dysplasia (60%) and gliosis (16%). For further information please see **Supplementary Table S1**.

3.2. Selected segments

Fig. 1 shows times of selected segments in circular plots. Random and minimum entropy segments do not show any distinct distribution pattern. Random wake segments were selected mostly from 6AM to 9PM. The first 5 minutes of N2 were selected in the beginning of a night, around 9PM. NREM sleep segments from 9PM to 3AM and REM sleep from midnight to 4AM. Maximum IED segments were selected mostly from 9PM to 5AM with a clear preference in the first half of the night. The minimum IED segment was selected predominantly between 5-6AM which might correspond to the last REM sleep cycle.

3.3. Comparison of the models based on individual segments

Fig. 2 shows the comparison of the performance of the individual segments. Sleep segments exhibit statistically significant differences compared to the baseline (i.e., the outcome obtained by chance and used as a reference point has a PRAUC of 0.062). The best PRAUC results for individual segments with an average result of 0.423 are presented on the bar plot.

3.4. Comparison of the models based on data from all segments

Fig. 3 presents the performance of models trained on average (AVG), variance (VAR), and average and variance (VAR_AVG) features. For both SVC and Logistic Regression, the VAR_AVG model is significantly better than the AVG and VAR model. SVC model exhibits slightly higher score in all three variants compared to Logistic Regression and therefore we decided to use this model in our further investigation.

3.5. Comparison of single segment model and AVG_VAR model

Fig. 4 shows the comparison between the result of the SVC model obtained with VAR_AVG features and the results of SVC single segment models. Overall, the VAR_AVG model achieved better scores than models based on short segments. However, this was true in only 12 out of 16 segments (75%). Among the single segment models, the NREM segments achieved the best scores, which were not statistically different to the VAR_AVG model (all ps > 0.05), except random NREM #1 (p = 0.011). The results are presented in Table 2.

The most important features positively correlated with the pathology of the VAR_AVG model were: Phase synchronization in Theta, Beta, Gamma, and High Gamma bands, Coherence in Theta, Gamma, High Gamma, and Ripple band, Spike rate, Power in Gamma band, Median linear correlation in Ripple band, and HFO rates in the Ripple band. In case of the single-segment models, the most important features varied between segments. For further information please see Supplementary Tables S2 and S3 and **Supplementary Figure S1**.

4. Discussion

In this study, we analyzed 25 consecutive adult patients with drug-resistant focal epilepsy from the Montreal Neurological Institute and the St. Anne's University Hospital in Brno, who underwent SEEG with subsequent resective surgery resulting in good post-surgical outcome. We processed 2381 hours of SEEG recordings, calculated multiple EEG features and showed that (i) the score of the models based on short segments varies significantly across various sleep stages and interictal conditions introducing uncertainty for identifying the EZ; (ii) the model based on the combined average and variance of iEEG features across segments is significantly better than the model based on variance or average alone; and (iii) the model based on a well-selected 5-minute long segment such as NREM N2 and N3 can achieve the same result as the model based on the prolonged recording.

4.1. Short recordings vary in their ability to identify the EZ

In this study, we evaluated short EEG segments across different sleep stages and interictal conditions. The average score across different 5-minute segments was 0.423, which is an acceptable result considering the average PRAUC baseline of 0.062. Nevertheless, one should be extremely careful when selecting short segments of data for analysis. Based on our results, the PRAUC varied between 0.323–0.493 across different segments, suggesting that when selecting a short segment of SEEG, one can get lucky or very unlucky. These results are new but not unexpected considering the significant influence of circadian and sleep cycles on interictal and ictal activity in epilepsy (Bercel, 1964; Gliske et al., 2018; Karoly et al., 2021). It is well described, that circadian rhythms influence the timing of seizures, with seizures occurring most frequently in the morning hours, while sleep cycles can influence both the frequency and intensity of seizures (Spencer et al., 2016). Further-

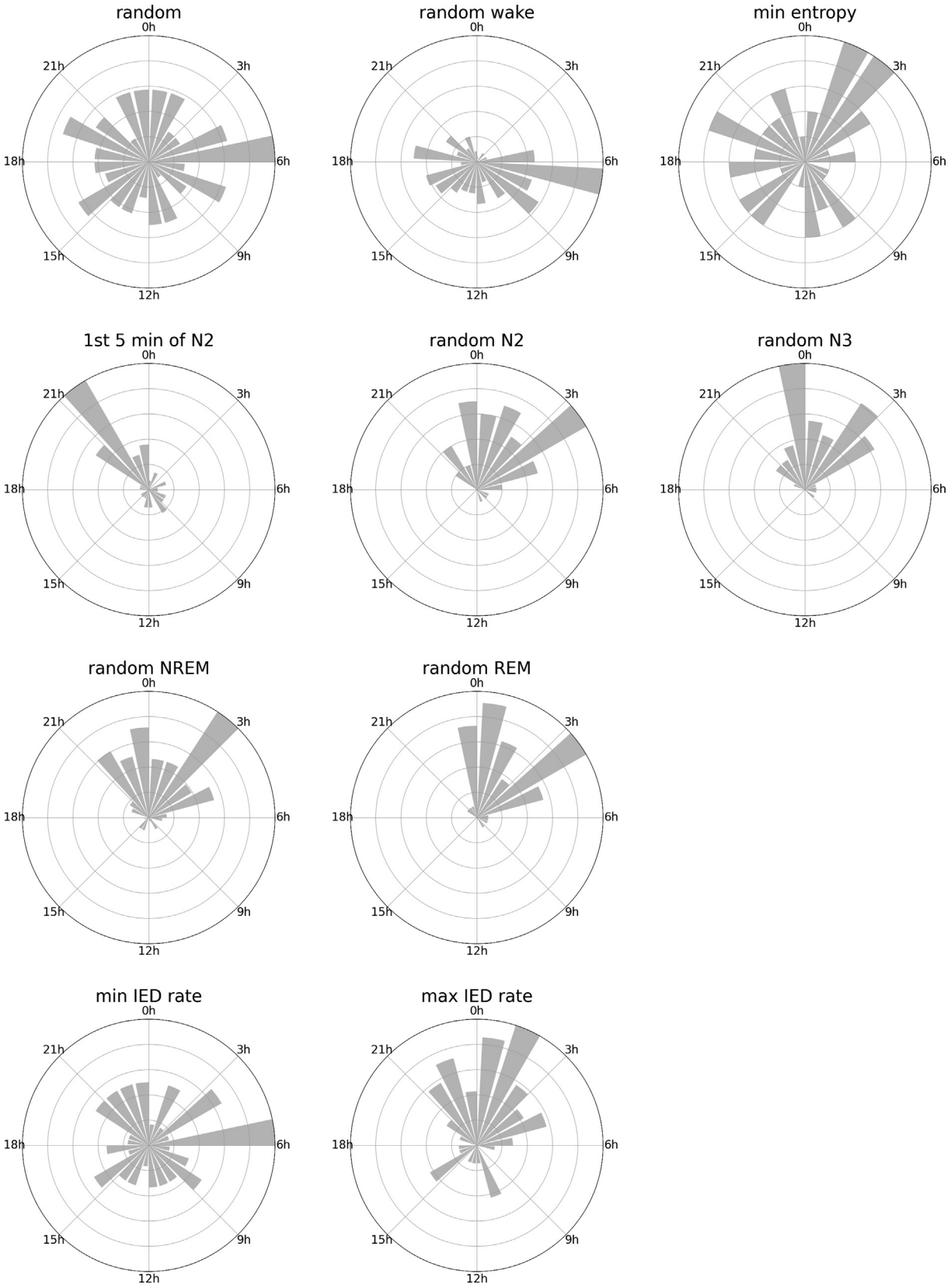

Fig. 1. Times of selected segments across all patients. The circular graphs show 24 h period with midnight at 0 h and midday at 12 h. The number of selected segments were counted in 1-hour bins.

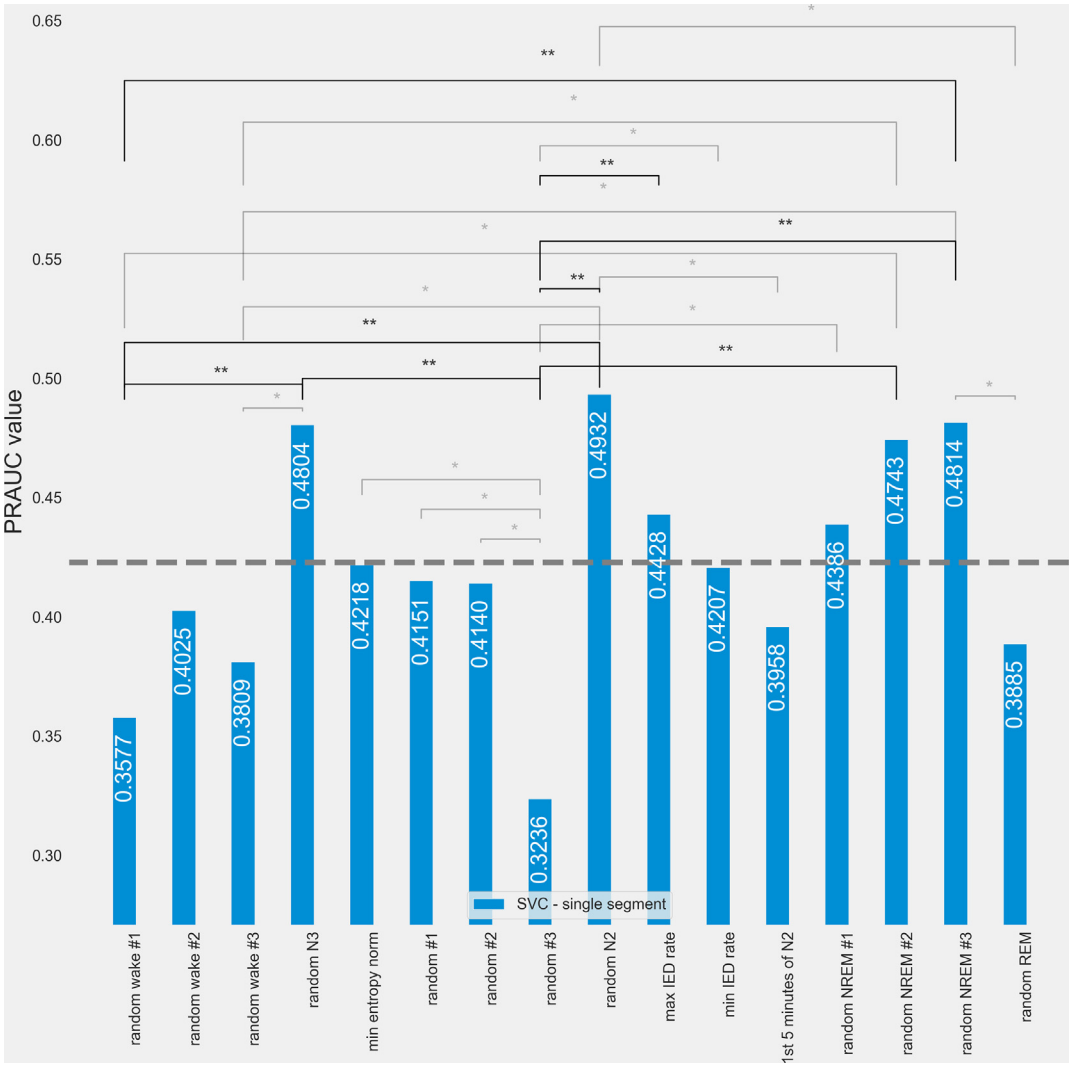

Fig. 2. Comparison of single-segment Area Under the Precision-Recall Curve (PRAUC) values for the Support Vector Classifier (SVC) model. The asterisks indicate statistically significant * (<0.05) and highly statistically significant ** (<0.001) results. The dashed line indicates the average result for single segment predictions.

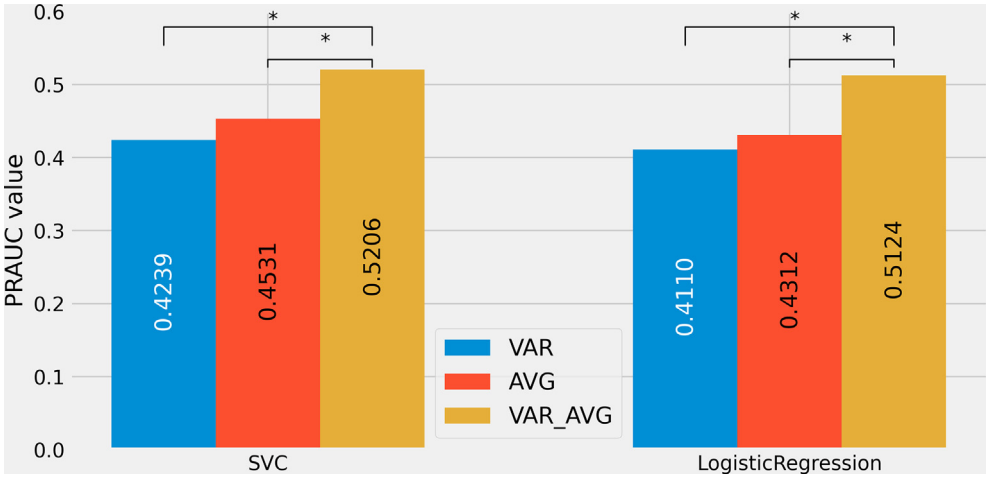

Fig. 3. Comparison of the Area Under the Precision-Recall Curve (PRAUC) values of Logistic Regression (LR) and Support Vector Classifier (SVC) models using average (AVG), variance (VAR), and both variance and average (VAR_AVG) features. The asterisks indicate statistically significant * (<0.05) differences. The p-value equals 0.01 and 0.001 when the LR VAR_AVG model is compared to AVG and VAR models respectively. In the case of the SVC model, p-values equal 0.04 and 0.003 for AVG and VAR respectively.

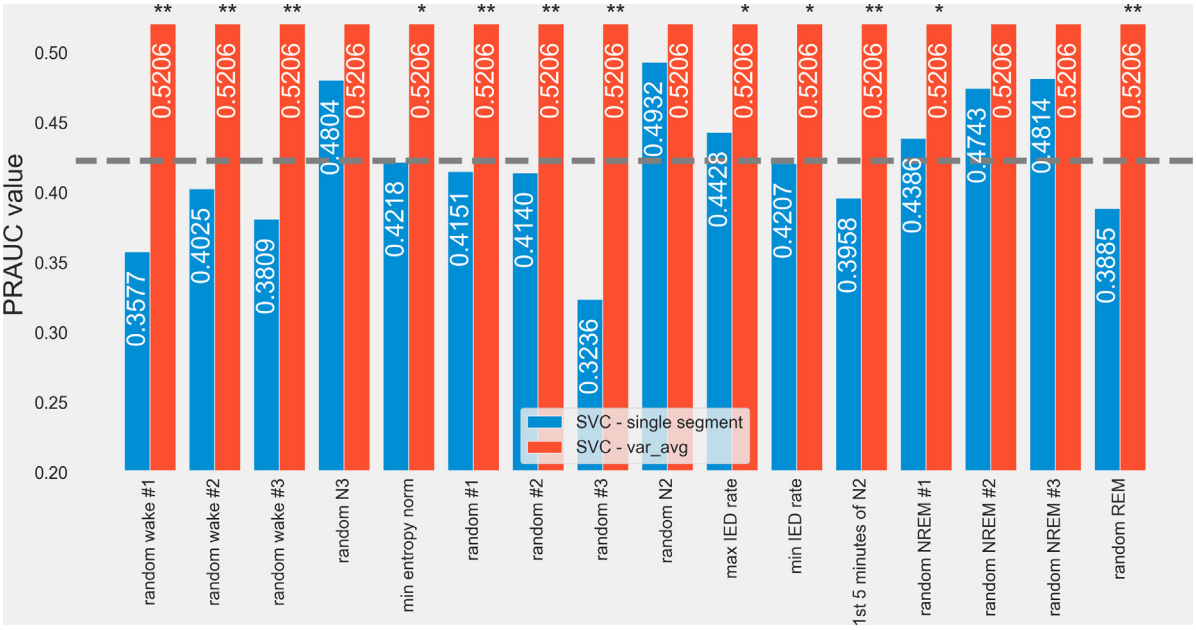

Fig. 4. Comparison of a single segment and variance and average (VAR_AVG) models. The asterisks indicate a statistically significant difference * (<0.05), ** (<0.001) between the single segment and the VAR_AVG model. The dashed line indicates the average result for single segment predictions.

Table 2
Statistical comparison of the Area Under the Precision-Recall Curve (PRAUC) between variance and average (VAR_AVG) model and single segment models. Highlighted p-values show statistically (bold) and strongly statistically (bold italic) significant difference. Legend: REM = rapid eye movement; NREM = non rapid eye movement; IED = interictal epileptic discharge.

Segment	PRAUC A (VAR_AVG)	PRAUC B (single segment)	standard error A	standard error B	difference: areaA - areaB	standard error of the difference	Z	P-value: directional (one-tailed)
random wake #1	0.521	0.358	0.024	0.020	0.163	0.031	5.274	***<0.001***
random wake #2	0.521	0.403	0.024	0.021	0.118	0.032	3.726	***<0.001***
random wake #3	0.521	0.381	0.024	0.021	0.140	0.031	4.463	***<0.001***
random N3	0.521	0.480	0.024	0.023	0.040	0.033	1.226	0.220
min entropy norm	0.521	0.422	0.024	0.022	0.099	0.032	3.080	**0.002**
random #1	0.521	0.415	0.024	0.022	0.106	0.032	3.310	***0.001***
random #2	0.521	0.414	0.024	0.022	0.107	0.032	3.342	***0.001***
random #3	0.521	0.324	0.024	0.019	0.197	0.030	6.525	***<0.001***
random N2	0.521	0.493	0.024	0.023	0.027	0.033	0.830	0.407
max IED rate	0.521	0.443	0.024	0.022	0.078	0.032	2.409	**0.016**
min IED rate	0.521	0.421	0.024	0.022	0.100	0.032	3.114	**0.002**
1st 5 minutes of N2	0.521	0.396	0.024	0.021	0.125	0.032	3.951	***<0.001***
random NREM #1	0.521	0.439	0.024	0.022	0.082	0.032	2.538	**0.011**
random NREM #2	0.521	0.474	0.024	0.023	0.046	0.033	1.418	0.156
random NREM #3	0.521	0.481	0.024	0.023	0.039	0.033	1.196	0.232
random REM	0.521	0.389	0.024	0.021	0.132	0.032	4.194	***<0.001***

more, a recent study found that IED spatial distribution fluctuated significantly over time, and a median of 12 sequential hours was required to capture 80% of this variability (Conrad et al., 2020). In our latest work, we showed that a minimum of 18 hours of continuous recording is needed for correct outcome prediction (Klimes et al., 2022). The above-mentioned studies are comprehensively describing the fluctuations in interictal activity. However, the variance of different EEG features in machine learning models designed for EZ localization over prolonged periods of time has not been systematically investigated, and this is the first study showing that combining average and variance features achieve the best performance.

4.2. Well selected short recordings have a similar value to identify the EZ as long recordings

Another finding of this paper is that albeit on average, short segments performed worse than long recordings, we found that NREM sleep periods had a similar value to identify the EZ as long recordings. This finding is in agreement with previous studies, that when selecting a short segment for automatic EZ localization, the NREM sleep stage is expected to provide the most accurate results (Bagshaw et al., 2009; Clemens et al., 2007; von Ellenrieder et al., 2017; Klimes et al., 2019; Sammaritano et al., 1991; Staba et al., 2004). However, to be on the safe side, the utilization of long recordings seems to be more reliable. Unfortunately, prolonged recordings are not always available, and even when they are, the processing times are considerably longer. Most retrospective studies, therefore select rather short segments of EEG. Given the results presented in this paper, the short segment strategy does not have to be always inferior. As long as the segment selection is not done randomly, but rather systematically, the probability of getting unreliable result is minimized.

4.3. Times of selected segments

Inspecting the circular graphs in Fig. 1, it is apparent that the random segments were evenly distributed around the clock except

for a peak around 6AM, which could be explained by the fact that patients around this time are usually well recorded as they are part of the night before patients often get disconnected for their morning routine. More distinct patterns were visible in timing of segments with maximum and minimum IED rate, which were distributed mainly from 9PM to 5AM and 5 to 6AM respectively as expected from previous studies (Conrad et al., 2023, 2020; Sammaritano et al., 1991). NREM sleep segments were picked mainly from night, but in some patients, few segments happened to be selected during the day time, suggesting that day time naps might be enough to record desired NREM activity for accurate EZ localization. This was however not the aim of this study and would require confirmation in further studies.

4.4. Considering the variance of EEG features is useful in long recordings

As a part of this study, we investigated the potential benefit of longer recordings and iEEG feature variance over time. The model based solely on the variance of features, calculated from at least 24-hour long segment achieved a score of 0.424, and the model based on average features achieved a score of 0.453. However, a significantly better result was achieved, when combining average and variance of features, reaching a score of 0.521. This clearly shows that the model based on the combination of average and variance achieves the best score in comparison to a model based on just variance or just average. The analysis of the features' importance shows that the variance and average of spike rate, HFO, coherence and phase synchrony in different frequency bands appear to contribute the most. The basic version (not variance or average) of features from VAR_AVG model are among the most contributing ones in the single segment models. This novel approach is not frequently used albeit tempting when aiming to analyze features that are known to have variability over time (Chen et al., 2021; Gliske et al., 2018). This is also in line with our previous work on the analysis of IED dynamics that showed that surgery outcome depends on strong, single and stable sources, where variance was one useful measure in the prediction model (Chybowski, 2023).

4.5. Selection of the PRAUC and not standard AUC

We decided to utilize the PRAUC instead of standard AUC as an evaluation metric. Large numbers of previous studies, including our past work, uses area under the receiver operating characteristics (ROC), as we refer to it as standard AUC (Abdallah et al., 2022; Cimbalnik et al., 2019; Klimes et al., 2016; Li et al., 2016; Murphy et al., 2017; Weiss et al., 2015). However, given the class imbalance inherent to our type of datasets with significantly lower number of channels inside than outside the EZ, we decided to use a metric that is the most informative and suitable for binary classification in datasets with such a distribution, focusing on pathology prediction. Unlike ROC, the precision-recall curve is less sensitive to true negative (TN) cases in the evaluation model. Given, that individual patients have between 100–150 SEEG contacts implanted, and the EZ is often not bigger than 5 contacts, all correctly classified non-EZ contacts might produce significant imbalance to metrics like specificity, which is the basis of ROC calculation. This imbalance would artificially increase the score of the model. On the other hand, precision and recall ignore TNs by their definition (Precision = TP/TP + FP; Recall = TP/TP + FN). In our case TP (true positive) is the correctly classified EZ, FP (false positive) is the incorrectly classified non-EZ, and FN (false negative) in the incorrectly classified EZ. PRAUC is, however, not completely independent on the TN ratio. It is important to relate results to the PRAUC baseline, which is the ratio of EZ contacts in the dataset (in our case 0.062). If, for example, this ratio is 50%, the result of PRAUC = 0.5 is basically a coin toss (similar to ROCAUC = 0.5).

4.6. Strengths and potential limitations

The main strength of the presented approach is its high potential to achieve a reduction in recording time. Every long recording is a risk for a patient. Reducing the recording time to only a few dozen minutes might be highly beneficial and significantly improve patient comfort and well-being, as well as lower the risks and costs. In this paper, we analyzed 25 patients with Engels I postsurgical outcomes gathered from 2 tertiary epilepsy centers. This sample size does unfortunately not allow to assess differences across pathologies regarding the type of segments with best performance. Furthermore, the data processed in this study are from 1-3 days and hence cannot assess if changes in medication might have contributed to changes in performance. This awaits clarification in complete SEEG investigations. Overall usefulness of these models has to be demonstrated not only for successful surgeries but also for patients with poor surgical outcomes, as done in other studies (for example Klimes et al. 2019). The scope of this study was to evaluate how the performance of these models vary over different states of vigilance and to do so we needed a correctly confirmed location of the EZ region, which is only possible in patients with a good postsurgical outcome.

5. Conclusions

Our study shows that models based on variability and average of iEEG features derived from $\geq$ 24-hour recordings achieve better scores than models based on short segments. However, this was not true in 25% of all short segments that we analyzed. This work was able to show that well selected 5-minute long segments, preferably from NREM N2 or N3 sleep, can achieve reliable results. Clinicians should hence be aware that timing matters when relying on short segments.

Conflict of interest

None of the authors have potential conflicts of interest to be disclosed.

Acknowledgments

This work was supported by The CAS project RVO:68081731; Ministry of Education, Youth and Sport of the Czech Republic (EATRIS-CZ, LM2023053); Supported by project nr. LX22NPO5107 (MEYS): Financed by European Union – Next Generation EU; Czech Science Foundation, project 22-28784S; Ministry of Health of the Czech Republic, project NU22-08-00278, and the Canadian Institutes of Health Research (project grant PJT-175056). B.F. is supported by a salary award ("Chercheur-boursier clinicien Senior") from the Fonds de Recherche du Québec - Santé 2021–2023).

Author contributions

Conception and design of the study: B.C, P.K. and B.F.

Acquisition and analysis of the data: B.C, P.K., J.C., V.T, P.N., M.P., L.P-D, J.H., F.D.

Drafting of the manuscript and preparing the figures: B.C, P.K. P. J, M.B, and B.F.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.clinph.2024.01.007.

References

Abdallah C, Hedrich T, Koupparis A, Afnan J, Hall JA, Gotman J, et al. Clinical yield of electromagnetic source imaging and hemodynamic responses in epilepsy: Validation with intracerebral data. Neurology 2022;98:e2499–511. https://doi.org/10.1212/WNL.0000000000200337.

Bagshaw AP, Jacobs J, LeVan P, Dubeau F, Gotman J. Effect of sleep stage on interictal high-frequency oscillations recorded from depth macroelectrodes in patients with focal epilepsy. Epilepsia 2009;50:617–28. https://doi.org/10.1111/j.1528-1167.2008.01784.x.

Baud MO, Kleen JK, Mirro EA, Andrechak JC, King-Stephens D, Chang EF, et al. Multi-day rhythms modulate seizure risk in epilepsy. Nat Commun 2018;9:88. https://doi.org/10.1038/s41467-017-02577-y.

Bercel NA. The periodic features of some seizure states. Ann N Y Acad Sci 1964;117:555–62. https://doi.org/10.1111/j.1749-6632.1964.tb48206.x.

Berry RB, Brooks R, Gamaldo C, Harding SM, Lloyd RM, Quan SF, et al. AASM scoring manual updates for 2017 (Version 2.4). J Clin Sleep Med 2017;13:665–6. https://doi.org/10.5664/jcsm.6576.

Chen Z, Grayden DB, Burkitt AN, Seneviratne U, D'Souza WJ, French C, et al. Spatiotemporal patterns of high-frequency activity (80–170 Hz) in long-term intracranial EEG. Neurology 2021;96:e1070–81. https://doi.org/10.1212/WNL.0000000000011408.

Chybowski B. bartlomiej-chybowski/ez_identification: release for the publication 2023. https://doi.org/10.5281/zenodo.7906344.

Cimbalnik J, Klimes P, Sladky V, Nejedly P, Jurak P, Pail M, et al. Multi-feature localization of epileptic foci from interictal, intracranial EEG. Clin Neurophysiol 2019;130:1945–53. https://doi.org/10.1016/j.clinph.2019.07.024.

Cimbalnik J, SweetVlad. ICRC-BME/epycom: EPYCOM-beta 2020. https://doi.org/10.5281/zenodo.4030570.

Clemens Z, Mölle M, Erőss L, Barsi P, Halász P, Born J. Temporal coupling of parahippocampal ripples, sleep spindles and slow oscillations in humans. Brain 2007;130:2868–78. https://doi.org/10.1093/brain/awm146.

Conrad EC, Revell AY, Greenblatt AS, Gallagher RS, Pattnaik AR, Hartmann N, et al. Spike patterns surrounding sleep and seizures localize the seizure-onset zone in focal epilepsy. Epilepsia 2023;64:754–68. https://doi.org/10.1111/epi.17482.

Conrad EC, Tomlinson SB, Wong JN, Oechsel KF, Shinohara RT, Litt B, et al. Spatial distribution of interictal spikes fluctuates over time and localizes seizure onset. Brain 2020;143:554–69. https://doi.org/10.1093/brain/awz386.

von Ellenrieder N, Andrade-Valença LP, Dubeau F, Gotman J. Automatic detection of fast oscillations (40–200 Hz) in scalp EEG recordings. Clin Neurophysiol 2012;123:670–80. https://doi.org/10.1016/j.clinph.2011.07.050.

von Ellenrieder N, Dubeau F, Gotman J, Frauscher B. Physiological and pathological high-frequency oscillations have distinct sleep-homeostatic properties. Neuroimage Clin 2017;14:566–73. https://doi.org/10.1016/j.nicl.2017.02.018.

von Ellenrieder N, Frauscher B, Dubeau F, Gotman J. Interaction with slow waves during sleep improves discrimination of physiologic and pathologic high-frequency oscillations (80–500 Hz). Epilepsia 2016;57:869–78. https://doi.org/10.1111/epi.13380.

Engel J. Update on surgical treatment of the epilepsies: summary of the second international palm desert conference on the surgical treatment of the epilepsies (1992). Neurology 1993;43:1612. https://doi.org/10.1212/WNL.43.8.1612.

Frauscher B, Bartolomei F, Kobayashi K, Cimbalnik J, van 't Klooster MA, Rampp S, et al. High-frequency oscillations: The state of clinical research. Epilepsia 2017;58:1316–29. https://doi.org/10.1111/epi.13829.

Frauscher B, von Ellenrieder N, Zelmann R, Doležalová I, Minotti L, Olivier A, et al. Atlas of the normal intracranial electroencephalogram: neurophysiological awake activity in different cortical areas. Brain 2018;141:1130–44. https://doi.org/10.1093/brain/awy035.

Gliske SV, Irwin ZT, Chestek C, Hegeman GL, Brinkmann B, Sagher O, et al. Variability in the location of high frequency oscillations during prolonged intracranial EEG recordings. Nat Commun 2018;9:2155. https://doi.org/10.1038/s41467-018-04549-2.

Hanley JA, Mcneil B. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology 1982;143:29–36. https://doi.org/10.1148/radiology.143.1.7063747.

Ives JR. New chronic EEG electrode for critical/intensive care unit monitoring. J Clin Neurophysiol 2005;22:119–23. https://doi.org/10.1097/01.WNP.0000152659.30753.47.

Jacobs J, Wu JY, Perucca P, Zelmann R, Mader M, Dubeau F, et al. Removing high-frequency oscillations: A prospective multicenter study on seizure outcome. Neurology 2018;91:e1040–52. https://doi.org/10.1212/WNL.0000000000006158.

Janca R, Jezdik P, Cmejla R, Tomasek M, Worrell GA, Stead M, et al. Detection of interictal epileptiform discharges using signal envelope distribution modelling: application to epileptic and non-epileptic intracranial recordings. Brain Topogr 2015;28:172–83. https://doi.org/10.1007/s10548-014-0379-1.

Jobst BC, Cascino GD. Resective epilepsy surgery for drug-resistant focal epilepsy: a review. JAMA 2015;313:285–93. https://doi.org/10.1001/jama.2014.17426.

Karoly PJ, Rao VR, Gregg NM, Worrell GA, Bernard C, Cook MJ, et al. Cycles in epilepsy. Nat Rev Neurol 2021;17:267–84. https://doi.org/10.1038/s41582-021-00464-1.

Klimes P, Cimbalnik J, Brazdil M, Hall J, Dubeau F, Gotman J, et al. NREM sleep is the state of vigilance that best identifies the epileptogenic zone in the interictal electroencephalogram. Epilepsia 2019;60:2404–15. https://doi.org/10.1111/epi.16377.

Klimes P, Duque JJ, Brinkmann B, Van Gompel J, Stead M, St Louis EK, et al. The functional organization of human epileptic hippocampus. J Neurophysiol 2016;115:3140–5. https://doi.org/10.1152/jn.00089.2016.

Klimes P, Peter-Derex L, Hall J, Dubeau F, Frauscher B. Spatio-temporal spike dynamics predict surgical outcome in adult focal epilepsy. Clin Neurophysiol 2022;134:88–99. https://doi.org/10.1016/j.clinph.2021.10.023.

Krucoff MO, Chan AY, Harward SC, Rahimpour S, Rolston JD, Muh C, et al. Rates and predictors of success and failure in repeat epilepsy surgery: A meta-analysis and systematic review. Epilepsia 2017;58:2133–42. https://doi.org/10.1111/epi.13920.

Lagarde S, Roehri N, Lambert I, Trebuchon A, McGonigal A, Carron R, et al. Interictal stereotactic-EEG functional connectivity in refractory focal epilepsies. Brain 2018;141:2966–80. https://doi.org/10.1093/brain/awy214.

Li Y-H, Ye X-L, Liu Q-Q, Mao J-W, Liang P-J, Xu J-W, et al. Localization of epileptogenic zone based on graph analysis of stereo-EEG. Epilepsy Res 2016;128:149–57. https://doi.org/10.1016/j.eplepsyres.2016.10.021.

van Mierlo P, Papadopoulou M, Carrette E, Boon P, Vandenberghe S, Vonck K, et al. Functional brain connectivity from EEG in epilepsy: seizure prediction and epileptogenic focus localization. Prog Neurobiol 2014;121:19–35. https://doi.org/10.1016/j.pneurobio.2014.06.004.

Murphy PM, von Paternos AJ, Santaniello S. A novel HFO-based method for unsupervised localization of the seizure onset zone in drug-resistant epilepsy. In: 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). p. 1054–7. https://doi.org/10.1109/EMBC.2017.8037008.

Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine Learning in Python. J Mach Learn Res 2011;12:2825–30.

Powers D, Ailab.. Evaluation: From precision, recall and F-measure to ROC, informedness, markedness & correlation. J Mach Learn Technol 2011;2:2229–3981. https://doi.org/10.9735/2229-3981.

Ren L, Kucewicz MT, Cimbalnik J, Matsumoto JY, Brinkmann BH, Hu W, et al. Gamma oscillations precede interictal epileptiform spikes in the seizure onset zone. Neurology 2015;84:602–8. https://doi.org/10.1212/WNL.0000000000001234.

Saito T, Rehmsmeier M. The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. PLoS One 2015;10:e0118432.

Sammaritano M, Gigli GL, Gotman J. Interictal spiking during wakefulness and sleep and the localization of foci in temporal lobe epilepsy. Neurology 1991;41:290. https://doi.org/10.1212/WNL.41.2_Part_1.290.

Sarnthein J, Jacobs J, Zijlmans M. Editorial: High-frequency oscillations in the hippocampus as biomarkers of pathology and healthy brain function. Front Hum Neurosci 2021;15.

Spanedda F, Cendes F, Gotman J. Relations between EEG seizure morphology, interhemispheric spread, and mesial temporal atrophy in bitemporal epilepsy. Epilepsia 1997;38:1300–14. https://doi.org/10.1111/j.1528-1157.1997.tb00068.x.

Spencer DC, Sun FT, Brown SN, Jobst BC, Fountain NB, Wong VSS, et al. Circadian and ultradian patterns of epileptiform discharges differ by seizure-onset location during long-term ambulatory intracranial monitoring. Epilepsia 2016;57:1495–502. https://doi.org/10.1111/epi.13455.

Staba RJ, Wilson CL, Bragin A, Jhung D, Fried I, Engel Jr J. High-frequency oscillations recorded in human medial temporal lobe during sleep. Ann Neurol 2004;56:108–15. https://doi.org/10.1002/ana.20164.

Thomas J, Kahane P, Abdallah C, Avigdor T, Zweiphenning WJEM, Chabardes S, et al. A subpopulation of spikes predicts successful epilepsy surgery outcome. Ann Neurol 2022. https://doi.org/10.1002/ana.26548.

Thomschewski A, Hincapié A-S, Frauscher B. Localization of the epileptogenic zone using high frequency oscillations. Front Neurol 2019;10.

Travnicek V, Klimes P, Cimbalnik J, Halamek J, Jurak P, Brinkmann B, et al. Relative entropy is an easy-to-use invasive electroencephalographic biomarker of the epileptogenic zone. Epilepsia 2023.

Weiss SA, Lemesiou A, Connors R, Banks GP, McKhann GM, Goodman RR, et al. Seizure localization using ictal phase-locked high gamma: A retrospective surgical outcome study. Neurology 2015;84:2320–8. https://doi.org/10.1212/WNL.0000000000001656.