Received: 14 May 2024 | Revised: 22 September 2024 | Accepted: 25 September 2024

DOI: 10.1002/epi4.13073

Epilepsia Open® Open Access

ORIGINAL ARTICLE

ECG-based epileptic seizure prediction: Challenges of current data-driven models

Sotirios Kalousios[1] | **Jens Müller**[2] | **Hongliu Yang**[2] | **Matthias Eberlein**[2] | **Ortrud Uckermann**[1,3] | **Gabriele Schackert**[1] | **Witold H. Polanski**[1] | **Georg Leonhardt**[1]

[1]Department of Neurosurgery, Faculty of Medicine and University Hospital Carl Gustav Carus, Technische Universität Dresden, Dresden, Germany

[2]TU Dresden, Faculty of Electrical and Computer Engineering, Institute of Circuits and Systems, Dresden, Germany

[3]Division of Medical Biology, Department of Psychiatry and Psychotherapy, Faculty of Medicine and University Hospital Carl Gustav Carus, Technische Universität Dresden, Dresden, Germany

Correspondence
Georg Leonhardt, Department of Neurosurgery, University Hospital Carl Gustav Carus, Fetscherstrasse 74, Dresden 01307, Germany.
Email: georg.leonhardt@ukdd.de

Funding information
Else Kröner Fresenius Center (EKFZ) for Digital Health

Abstract

Objective: Up to a third of patients with epilepsy fail to achieve satisfactory seizure control. A reliable method of predicting seizures would alleviate psychological and physical impact. Dysregulation in heart rate variability (HRV) has been found to precede epileptic seizures and may serve as an extracerebral predictive biomarker. This study aims to identify the preictal HRV dynamics and unveil the factors impeding the clinical application of ECG-based seizure prediction.

Methods: Thirty-nine adult patients (eight women; median age: 38, [IQR = 31, 56.5]) with 252 seizures were included. Each patient had more than three recorded epileptic seizures, each at least 2 hours apart. For each seizure, one hour of ECG prior to seizure onset was analyzed and 97 HRV features were extracted from overlapping three-minute windows with 10s stride. Two separate patient-specific experiments were performed using a support vector machine (SVM). Firstly, the separability of training data was examined in a non-causal trial. Secondly, the prediction was attempted in pseudo-prospective conditions. Finally, visualized HRV data, clinical metadata, and results were correlated.

Results: The mean receiver operating characteristic (ROC) area under the curve (AUC) for the non-causal experiment was 0.823 (±0.12), with 208 (82.5%) seizures achieving an improvement over chance (IoC) classification score ($p < 0.05$, Hanley & McNeil test). In pseudo-prospective classification, the ROC-AUC was 0.569 (±0.17), and 86 (49.4%) seizures were classified with IoC. Off-sample optimized SVMs failed to improve performance. Major limiting factors identified include non-stationarity, variable preictal duration and dynamics. The latter is expressed as both inter-seizure onset zone (SOZ) and intra-SOZ variability.

Significance: The pseudo-prospective preictal classification achieving IoC in approximately half of tested seizures suggests the presence of genuine preictal HRV dynamics, but the overall performance does not warrant clinical application at present. The limiting factors identified are often overlooked in non-causal

study designs. While current deterministic prediction methods prove inadequate, probabilistic approaches may offer a promising alternative.

Plain Language Summary: Many patients with epilepsy suffer from uncontrollable seizures and would greatly benefit from a reliable seizure prediction method. Currently, no such system is available to meet this need. Previous studies suggest that changes in the electrocardiogram (ECG) precede seizures by several minutes. In our work, we evaluated whether variations in heart rate could be used to predict epileptic seizures. Our findings indicate that we are still far from achieving results suitable for clinical application and highlight several limiting factors of present seizure prediction approaches.

KEYWORDS

ECG, HRV, preictal, seizure prediction, warning system

1 | INTRODUCTION

Epilepsy affects approximately 1% of the general population,[1] and individuals may experience severe neurobiological, psychological, and social consequences associated with this condition.[2] The unpredictability of seizures exposes people with epilepsy (PWE) to social stigma and fear of loss of control. As a result, PWE are more susceptible to anxiety and mood disorders[3] and often face risks from seizure-related injuries or sudden unexpected death in epilepsy (SUDEP).[4]

Despite advances in medical therapy, up to a third of PWE do not achieve satisfactory seizure control.[5,6] Secondary treatment options include resective or ablative surgery and neurostimulation therapy; however 25% of PWE remain therapy-refractory.[7] According to patients and caregivers, a reliable, non-invasive, and non-stigmatizing seizure prediction method could alleviate their burden.[8,9] Additionally, a warning system could promote the development of innovative therapeutic or preventive strategies for refractory patients.[10,11]

To this aim, it is imperative to identify a biomarker for impending seizures that aligns with patients' preferences and meets the requirements for the intended application scenario. Several studies have already described the occurrence of preictal tachycardia, seconds prior to seizure onset.[12] In contrast, alterations in heart rate variability (HRV) may precede seizures by several minutes, when compared to a patient's interictal baseline.[13] HRV is considered to reflect the activity of the autonomic nervous system (ANS).[14] Therefore, such alterations in the dynamic balance between the sympathetic and the parasympathetic system may be used as extracerebral predictive biomarkers for epileptic seizures. A series of wearable ECG devices are already available,[15] making ECG an easily amenable modality for seizure prediction.

Key points

- Better than chance pseudo-prospective preictal classification for seizures suggests the presence of genuine preictal HRV dynamics.
- The group-level results do not yet support clinical application.
- ECG-based seizure prediction is challenged by the non-stationarity, the variability in preictal duration, and HRV dynamics.
- These critical factors are often overlooked in non-causal study designs.
- The parallels with EEG-based seizure prediction suggest a broader neurocardiac information flow.

Novak et al. used retrospective data and were first to propose a method utilizing time-frequency mapping to predict seizures several minutes in advance.[16] This sparked considerable interest in the predictive potential of HRV, and subsequent studies followed. A fuzzy clustering algorithm implemented by Kerem and Geva on short-term recordings detected preictal alterations in 18 out of 21 seizures, up to 11 min before seizure onset.[17] In a further attempt, Fujiwara et al. presented an approach with 91% sensitivity, with a false-positive (warning) rate per hour (FP/h) of 0.7.[18] Pavei et al. demonstrated that early warning was possible in 94.1% of seizures tested, with a FP/h of 0.49 for a five-minute preictal interval.[19] Another study maintained a high sensitivity (89.06%), but extended the average prediction time to 13.7 min.[20] Despite promising results on retrospective data, a real-world application of

an ECG-based seizure prediction system has not yet been implemented.

The aim of this study was to assess the feasibility of patient-specific ECG-based seizure prediction, utilizing preictal short-term HRV dynamics, and unveil factors that impede its real-world application. To accomplish this, we compared results obtained through non-causal analyses, which verified preictal short-term dynamics in our cohort's ECG signals, with the results of pseudo-prospective, causal experiments. Finally, we correlated our results with clinical and neurophysiological factors, seeking to interpret the performance of current data-driven seizure prediction models.

2 | MATERIALS AND METHODS

This study was conducted at the neurosurgical epilepsy unit, University Hospital Carl Gustav Carus, Dresden, Germany, between September 2021 and December 2022. ECG data were obtained from PWE during monitoring with non-invasive (scalp) video-electroencephalography (V-EEG) and PWE undergoing presurgical evaluation with intracranial EEG (iEEG). Inclusion criteria for participants were: (i) 18 years of age or older, (ii) more than three recorded epileptic seizures during monitoring, each at least 2 h apart. Exclusion criteria were: (i) vagus nerve stimulator, (ii) cardiac pacemaker, (iii) cardiac arrhythmia. Seizure onset was determined by a senior epileptologist based on the first appearance of either ictal EEG activity or clinical manifestation, whichever occurred first. Data were acquired using a Nihon Kohden Neurofax EEG-1200 system, with a sampling rate of 200 Hz and 500 Hz for non-invasive and invasive patients, respectively.

Thirty-nine PWE were included, comprising 8 women and 31 men. The median age was 38 years (IQR = 31, 56.5). A total of 252 seizures were eligible for analysis, with patients experiencing a median of five seizures (IQR = 4, 8). Additional clinical metadata, including demographics, epilepsy history, and seizure-related information, were collected from electronic health records and discharge notes.

The study was approved by the Ethics Committee of the Technische Universität Dresden (BO-EK-398082021 and BO-EK-116022021).

2.1 | ECG processing and feature extraction

For each seizure, an 1-h segment of ECG signal preceding seizure onset was analyzed. The ECG signals were preprocessed with a fifth-order Butterworth bandpass filter (low cut-off: 1.5 Hz, high cut-off: 30 Hz). R-peaks were detected utilizing a newly developed algorithm based on the biopeaks' peakfinder,[21] incorporating heart rate (HR) adaptive temporal criteria to identify R-peaks exploiting their semi-periodicity. The quality of peak detection was evaluated in consecutive 10-s segments, categorizing them as either physiologic or artifact. The implemented method, originally proposed by Orphanidou et al., was adapted for zero-error tolerance.[22]

Features were extracted from overlapping three-minute windows with 10-s stride. Windows containing 10-s segments previously flagged as artifact were excluded from further analyses. The extracted features belong to the time and frequency domain, with a considerable part obtained by non-linear analyses, including recurrent quantification analysis. A table listing all 97 features can be found in the supplementary material.

2.2 | Time period definitions

In the present study, the period ranging from 60 to 10 min before seizure onset was defined as the interictal period (−60 to −10 min). The last 10 min before seizure onset were assigned as the preictal period (−10 to 0 min). To mitigate potential confounding effects arising from transitional states, the interval between −40 and −10 min was excluded during the training phase. Henceforth, we refer to the entire one-hour interval preceding seizure onset as "seizure" for simplicity, in both text and figure legends.

2.3 | Experimental design

Two different patient-specific experiments were conducted using the same dataset. The first experiment did not consider causality, while the second experiment did.

2.3.1 | Non-causal analysis

Goal of the first (non-causal) experiment was to examine the separability of training data. Employing a support vector machine (SVM), we conducted a recursive search to identify the feature combination yielding the highest area under the receiver operating characteristic curve (ROC-AUC) scores for each patient. Due to the exponential increase in computational complexity with the number of input features, we used an Extra Trees classifier for dimensionality reduction, retaining only the top 10 features based on their Gini Index. The performance was evaluated with a leave-one-out cross-validation (LOOCV), treating seizures as separate events (Figure 1A). The experimental design for each patient can be summarized as follows:

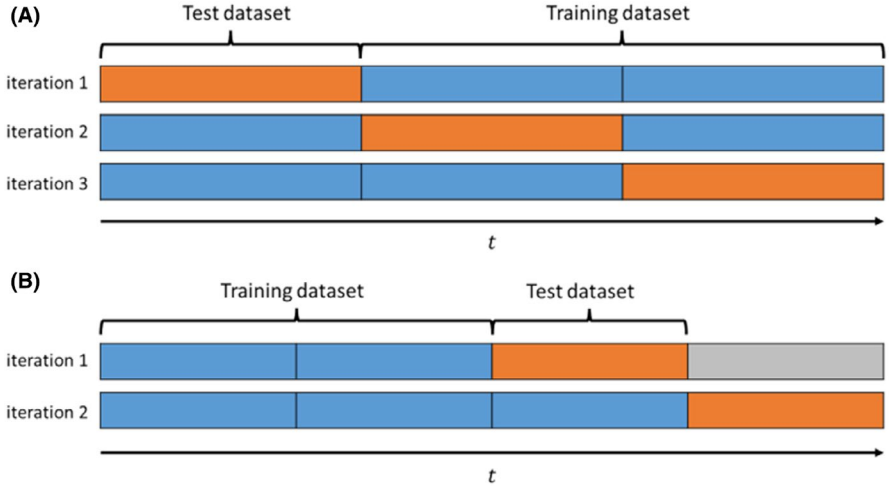

FIGURE 1 Illustration of the (A) leave-one-out cross-validation (LOOCV) and (B) pseudo-prospective testing. The horizontal arrow represents the time axis, denoted as *t* and each block represents the 1-h preceding each seizure. LOOCV dismisses chronological order of events.

1. Extract top 10 features using an Extra Trees classifier
2. Create $n = 968$ unique combinations of features (min. 3–max. 10)
3. Test n combinations with LOOCV of k splits (number of seizures) using SVM
4. Obtain the highest average ROC-AUC

2.3.2 | Causal analysis

This experiment employed a pseudo-prospective testing approach, wherein each seizure was successively considered an independent test set, while past data was assigned to the training set (Figure 1B). To optimize models and enhance performance, this experiment was subdivided into three complexity levels. At the first level, pseudo-prospective testing was conducted with an SVM classifier, incorporating all features for training in each iteration. The second level introduced feature selection (FS) in each pseudo-prospective testing iteration, selecting only features surpassing the 90th percentile of the Extra Trees classifier's Gini Index from the training set. In the third level, FS was combined with hyperparameter optimization through a grid search in a nested LOOCV, repeated in each iteration. Thereby, any potential information leakage was excluded in all levels.

Due to the stochastic nature of the implemented FS strategy, the latter two complexity levels were conducted five times for each patient. The results were averaged and improvement over chance (IoC) classification score was accepted, if the majority of the five iterations achieved better than chance classification performance for each seizure. The experimental design for each patient can be summarized as follows:

1. Past k seizures (≥ 2) serve as training(−validation) set
2. Selecting features from k seizures (complexity levels 2 & 3)
3. Optimizing hyperparameters with LOOCV with k splits (complexity level 3)
4. The chronologically next seizure in time axis t serves as test set
5. For n seizures, the previous steps are repeated $n - 2$ times

2.4 | Tools and evaluation metrics

2.4.1 | Support vector machine

Similar to previous studies,[19,20] an SVM was used for classification. In Support vector machine (SVM), given a set of binary-labeled training data, examples are mapped into a higher-dimensional space. A function is created trying to maximize the gap between the negative and positive classes. The resulting maximum-margin hyperplane, or decision boundary, is responsible for the classification of new, unseen data. By mapping unseen samples into the same space, the algorithm assigns a class, depending on their relative spatial position with respect to the hyperplane.[23] In this work, the SVM classifier utilized a radial basis function (RBF) kernel.

2.5 | Extra tree classifier

The Extra Tree classifier was employed for dimensionality reduction. These classifiers are ensembles of multiple decision trees. Trees are trained on the original input sample and randomly select cut-points for features, rather than the optimal. The use of random cut-point leads to highly de-correlated decision trees. For feature selection purposes, the Gini Index is computed, as the normalized total reduction criterion during feature split and assigned to each feature as importance or weight.[24] The selected features provide a lower-dimensional representation of

the original dataset. The Extra Tree classifier in this study consisted of 1000 estimators (decision trees).

2.6 | Multidimensional scaling

Multidimensional scaling (MDS) is a technique used for mapping high-dimensional observations onto a lower-dimensional space. Its major advantage consists of its ability to preserve the original similarities or dissimilarities in data in the form of shorter and longer distances in geometric space, respectively.[25] In the present study, MDS was used to visualize complex HRV data in more familiar 2D space and to formulate hypotheses based on these observations.

2.7 | Evaluation metrics

The area under the curve (AUC) of the receiver operating characteristic (ROC) is a measure of a model's ability to distinguish between two classes. The ROC curve expresses the relationship between the true positive rate (sensitivity) and the false positive rate (1—specificity) at various classification thresholds. This threshold-free metric was chosen to simplify model comparison, as it summarizes multiple threshold-dependent metrics into one.

More specifically, when classifying a finite number of samples, the AUC values obtained by a naive predictor should follow a normal distribution with $\mu = 0.5$. A model's performance is defined as improvement over chance (IoC), if the chance of a naive predictor achieving an equal or better score than the tested model is less than 5% ($p < 0.05$).[26] Therefore, a model's performance is defined as IoC if the AUC value is significantly greater than that of a naive predictor.

All analyses were conducted with Python (v. 3.9.7), with the use of custom algorithms and open source libraries, including Neurokit 2 (v. 2.1) for HRV extraction and with Scikit-learn (v. 1.1.2) for machine-learning models.[27,28]

3 | RESULTS

3.1 | Non-causal experiment

The non-causal experiment terminated after exhausting all unique feature combinations, fitting and testing $n \times k$ models for each patient. The average ROC-AUC (SD) across patients was 0.823 (±0.12). IoC classification scores were achieved in 208 out of 252 (82.5%) seizures (Table 1). The best-performing feature combinations varied significantly among patients, with no specific feature or category being preferred. On average, the selected combinations comprised 3.9 features, which cannot be safely interpreted physiologically, or assigned to a specific state. The obtained scores should approximate the upper limit of data separability using predefined training labels.

3.2 | Pseudo-prospective experiment

The average ROC-AUC score across all three complexity levels of the pseudo-prospective evaluation was approximately 0.56. Despite optimization efforts exploiting FS and hyperparameter optimization, there was no significant uplift in performance (Table 1). IoC classifications ranged from 45.4% to 49.4% of the tested seizures across all levels. The respective patient-wise performance varied significantly between the three tested pseudo-prospective testing methods, none of which demonstrated clear superiority (Figure 2).

There is a wide dispersion in scores, with some patients scoring below chance level, while others scored nearly perfectly. Patients with a small number of seizures in the test set tended to contribute to both extremes of the average patient-specific AUC dispersion. Score values obtained from patients with a seizure number above median (3) demonstrate more consistency. Two of these patients exhibit above-average classification score, consistently achieving ≥75% IoC classification across the different pseudo-prospective testing methodologies (red bars Figure 2).

TABLE 1 Scores for the non-causal and the causal experiments.

Experiment design	Method	Average ROC-AUC	SD	IoC classification (*n*, %)		Avg. IoC % per patient
Non-causal (LOOCV)	FS with recursive search	0.823	0.12	208	82.5%	86.0%
Causal (Pseudo-prospective testing)	All features	0.569	0.17	86	49.4%	44.4%
	FS	0.566	0.17	83	47.7%	48.7%
	FS and hyperparameter optimization (grid)	0.559	0.18	79	45.4%	44.5%

Abbreviations: avg., Average; FS, Feature selection; IoC, Improvement over chance; LOOCV, Leave-one-out cross-validation; SD, Standard deviation.

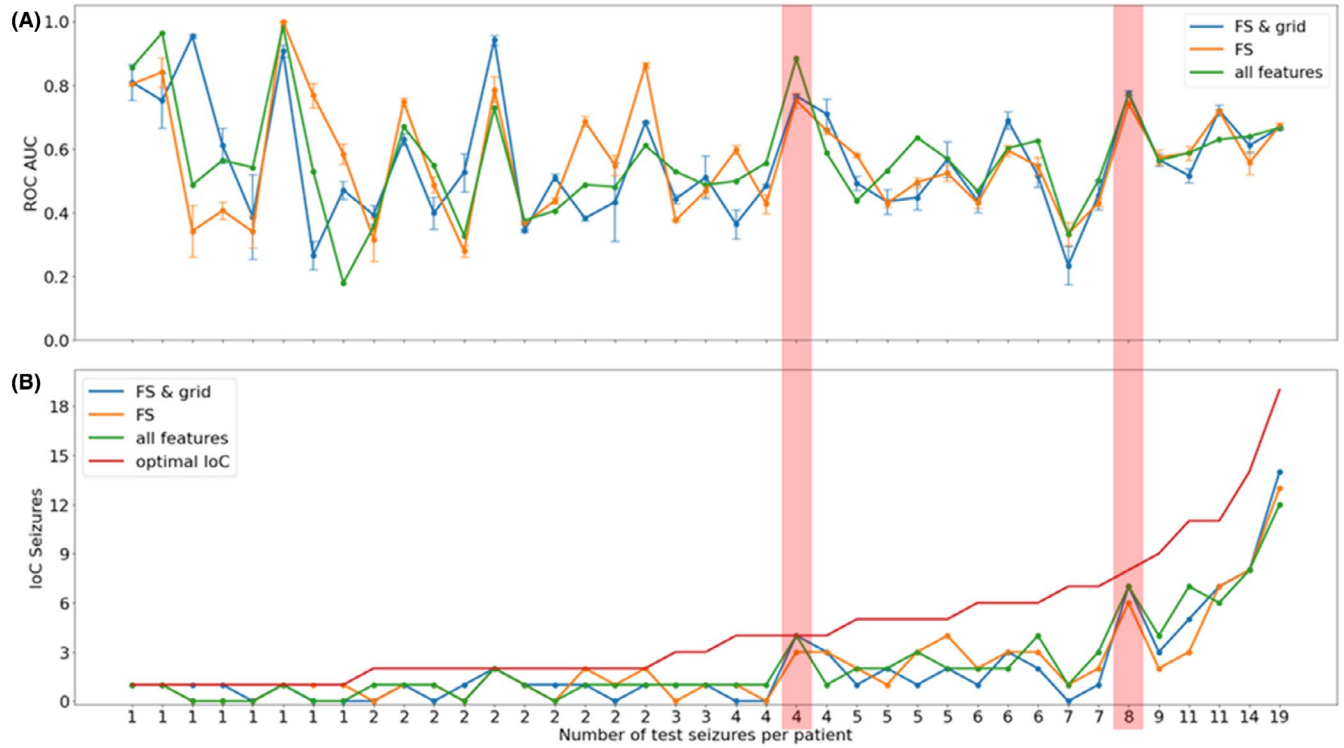

FIGURE 2 Pseudo-prospective trial: Scores for individual patients. (A) The average ROC-AUC score obtained for each patient sorted by the number of test seizures in ascending order. In the case of FS and FS &grid methods, the standard deviation (SD) bars for the five iterations illustrate the impact of heuristic feature selection (FS) behavior on the results. Red bars highlight two true score outliers. (B) The IoC classifications achieved for each patient. (FS: Feature selection; FS and grid: Feature selection and hyperparameter optimization with grid search; IoC: Improvement over chance)

We correlated our cumulative and seizure-wise results with clinical metadata and finally used MDS for data visualization. Prior to MDS, we used Extra Trees classifier to select the 10 best features, thus highlighting any differences in HRV between the interictal and preictal states for each patient. Each data point represents an observation window in the 1-h interval preceding seizure onset, with similarity in data depicted as shorter distance in the geometric space.

We have identified three closely intertwined factors that contribute to the models' performance loss: the signal's non-stationarity, the variability in preictal duration, and dynamics. For each factor, we have selected example cases from our cohort.

3.3 | Non-stationarity

Physiologic changes in HRV, for example induced by sleep–wake cycles, may lead to an increased data heterogeneity in the states preceding seizures during training. These changes can be perceived as the primary variation in data by the classifier, especially when the preictal dynamics are comparatively minor, leading to a weak prediction performance.

One such example is presented in Figure 3A, where three of four seizures occurred in wake state and share some relative similarities (closer in feature space). In contrast, the orange seizure (2) occurred during sleep and occupies a distinct position in feature space. The preictal periods have more affinity to the parent seizure interval than between them.

The clinical metadata of the two patients with the consistent high scores (Figure 2, red bars) reveal that all of their seizures occurred under similar circumstances, during sleep stage 2. Their data exhibit greater similarity, and preictal states tended to converge and cluster in a distinct regime in feature space as illustrated in Figure 3B.

3.4 | Variable preictal dynamics

By examining patients with seizures stratified by their sleep–wake cycles to compensate for physiologic HRV changes, our data provide evidence of variability in preictal dynamics of seizures originating, both between seizure onset zones (SOZ) and within a single SOZ.

Figure 4B illustrates a case with bitemporal epilepsy, where six out of seven seizures originated from the left

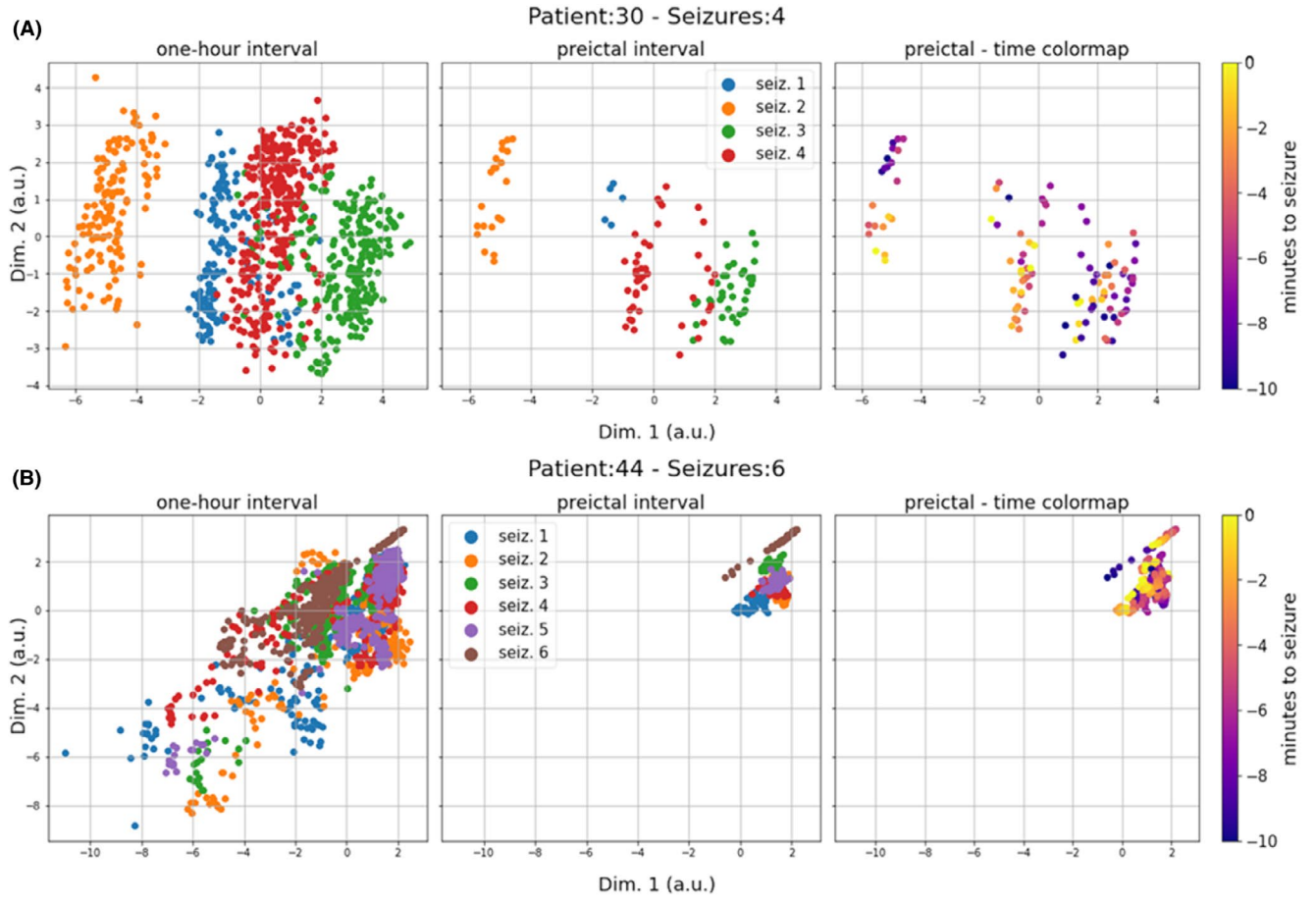

FIGURE 3 Non-stationarity in data. MDS mapping of consecutive seizures registered for patient 30 (A) and patient 44 (B). Each color represents a one-hour-long interval before each seizure. On the left MDS mapping of the 1-h preceding seizure onset and on the middle and right, only the preictal observation windows. On the right, the evolution of data over time. (MDS: Multidimensional scaling; Dim: Dimension; a.u.: Arbitrary units; seiz.: Seizure).

temporal lobe and only seizure 7 from the right temporal lobe. The right temporal preictal dynamics (circle) are distinct from those originating from the left (dashed oval), indicating a lateralization difference or inter-SOZ variability. Moreover, seizure 3 is taking a distinct path with time, after initially sharing similarities with all left temporal preictal dynamics. Clinical data reveal an unforeseen preictal tachycardia, suggestive of intra-SOZ variability.

In certain cases, it is possible that seizures might not be preceded by clearly identifiable or sufficiently long preictal dynamics at all (Figure 3A).

3.5 | Variable preictal duration

The conventionally assumed preictal time definition, based on a fixed duration, did not align with the time point at which features showed a distinct change prior to seizure. The actual duration of preictal dynamics varied significantly, observing differences both between patients and seizures.

For three consecutive seizures of patient 18 (Figure 4A), the preictal data points closer to seizure onset occupy a distinct space (dashed oval). However, far from onset preictal data points belonging to seizure 1 and 3 (−10 min to approximately −5 min prior to seizure onset) share more similarities with their interictal data (solid oval).

In the case of the bitemporal epilepsy, seizures show both shorter and longer preictal dynamics. The true preictal period of the seventh seizure extends beyond 10 min (Figure 4B, small oval), while data points from seizure 5 converge towards left temporal dynamics at −5 min before seizure onset.

4 | DISCUSSION

Most seizure prediction efforts to date build on the idea of a transitional state between interictal intervals and ictal events. In theory, this transition (i.e., the preictal period) is characterized by changes in HRV that should be clearly distinguishable from those in interictal intervals.

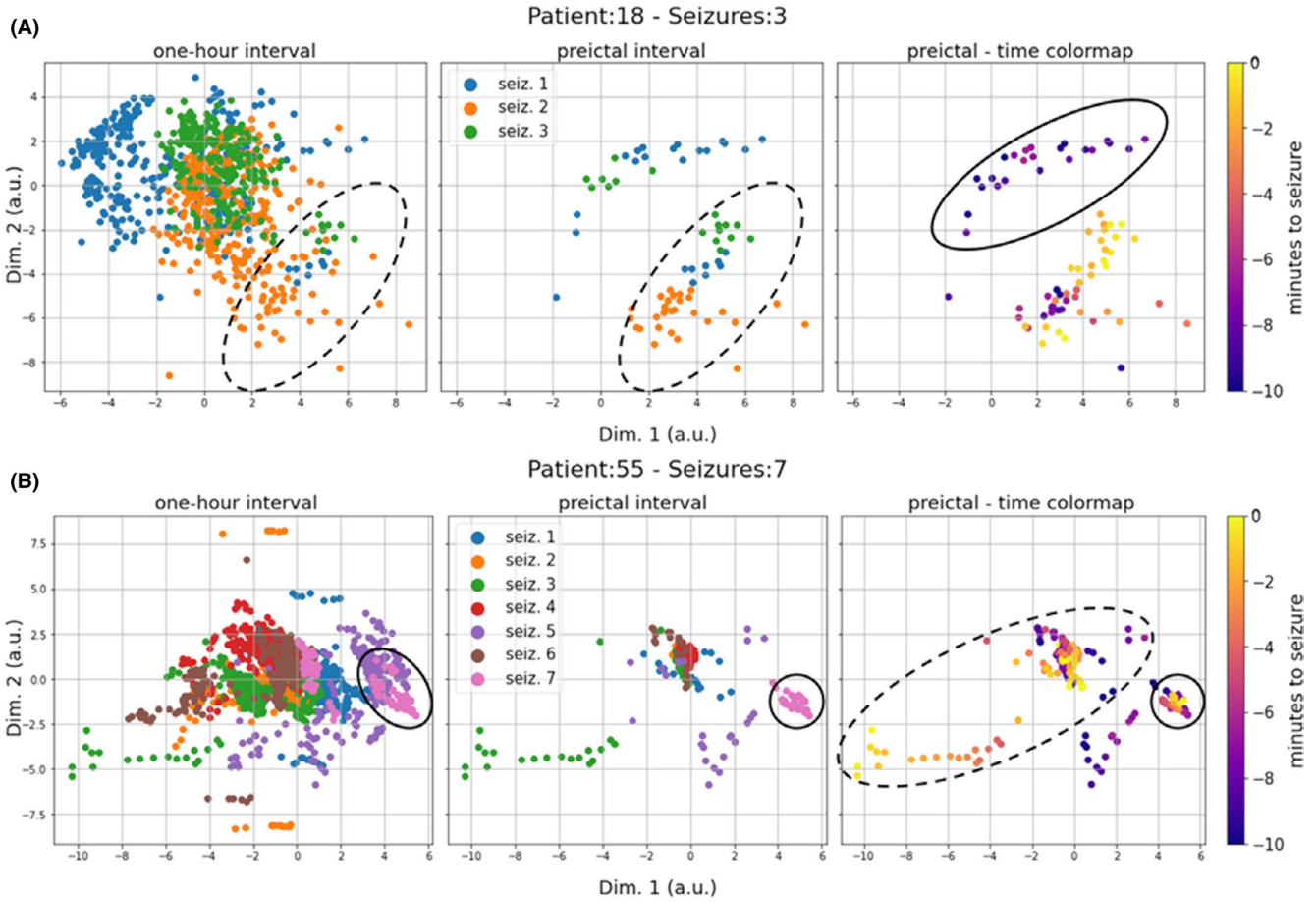

FIGURE 4 Variability in preictal duration and dynamics. MDS mapping of seizures stratified by sleep–wake cycles for patient 18 (A) and patient 55 (B). Each color represents a 1-h-long interval before each seizure. On the left MDS mapping of the 1-h preceding seizure onset and on the middle and right, only the preictal observation windows. On the right, the evolution of data over time. (MDS: Multidimensional scaling; Dim: Dimension; a.u.: Arbitrary units; seiz.: Seizure).

Breaking the chronological order of seizures and the strict training and test data separation with a recursive FS process, we have demonstrated that there are noticeable differences in HRV between interictal and preictal intervals. Under these conditions, the SVM achieved promising results. In a pseudo-prospective trial that aimed to mimic real conditions from a temporal perspective, the classification error increased significantly. However, the two highlighted exceptions and the IoC percentage of approximately 50%, demonstrated that class discrimination cannot be purely attributed to random feature fluctuations detected by the in-sample optimization of the non-causal analyses. Despite any causal, off-sample optimization, an unoptimized SVM classifier scored comparably to its optimized counterparts. This suggests that factors contributing to the suboptimal performance may indeed be rooted in the data. Finally, by correlating the visualized HRV data, the classification results, and clinical metadata, a number of closely intertwined factors impeding real-world application were identified. These include the signal's non-stationarity and the variability in preictal duration and dynamics.

In this work, we used a fixed preictal duration of 10 min, similar to previous HRV seizure prediction studies.[18–20] Our results indicate that the actual preictal duration is highly variable between seizures, an observation consistent with a recently published paper exploiting unsupervised learning methods.[29] The majority of preictal changes could be identified up to 40 min before seizures and showed high intra-patient variability in duration. Nonetheless, the preictal duration identified by most of their clustering solutions ranged from 2 to 9 min. Similarly, Behbahani et al. report that a prediction horizon of 4:30 min yielded the lowest FP/h in their cohort.[30]

Furthermore, as HR and HRV vary following longer or shorter cycles and trends, such as sleep–wake cycles and aging trends, this creates temporally varying statistical characteristics in features, termed non-stationarity.[31] Similar to an observation made by Billeci et al. in a smaller cohort,[20] the two patients highlighted in Figure 2 suggest that homogeneity in the conditions under which seizures occurred favors predictability,

potentially by reducing the non-stationarity of the signal (Figure 3).

After stratifying seizures based on seizure timings to compensate for non-stationarity effects, our observations suggest that a patient's seizures can be preceded by different preictal dynamics. The l patient with bilateral epilepsy presented in Figure 4B, provides evidence of lateralization in preictal dynamics. While this might be expected,[20] unilateral patients exhibiting inherently different dynamics originating from a single SOZ are of greater interest. However, a clear distinction between genuine intra-SOZ variability and non-stationarity effects might not be possible.

In addition to the previous findings of variable dynamics, the possibility of seizures occurring without autonomic precursors must be considered.[13] Since not all seizures exhibit ANS dysregulation during the ictal event, it raises doubts about the presence of ANS precursors. A previous seizure detection study using HRV showed that a detection rate of 66% or more is possible in only half of their patients.[32] Similarly, a review on ictal tachycardia states that only about 70% are associated with ictal tachycardia.[33] In the previously mentioned study exploiting unsupervised ML models, preictal dynamics were only detected in 41% of seizures.[29]

Recent studies suggest that the occurrence of seizures might be a probabilistic rather than a deterministic phenomenon, with seizure likelihood governed by cyclic modulation.[34–37] These cycles have been shown to comodulate with circadian and multidien cycles of epileptic brain activity[35] and HR,[38] and have been used to forecast seizure risk, even days in advance.[39,40] This novel approach frequently defines periods of low and high seizure risk. The latter is often referred to as the proictal state instead of preictal, to emphasize the uncertainty of seizure occurrence and the shift in underlying hypotheses.[36,41]

All these phenomena create two major limitations for current data-driven prediction models. First, regardless of evaluation method, the variability in preictal duration creates an unreliable ground truth for approaches employing fixed inter- and preictal periods for training and testing. Moreover, the probabilistic nature of seizure occurrence blurs the definition of ground truth further, rendering the underlying deterministic model behind seizure occurrence inadequate to describe it with a binary label. Second, the effects attributed to non-stationarity and variable preictal dynamics become more pronounced when evaluating performance pseudo-prospectively, as future data can potentially differ significantly from past events. In such a scenario, insufficiently represented interictal states or preictal dynamics in the training data can only lead to misclassification. Essentially, our classifiers may always be overfitted for the training data and fail to generalize appropriately to future data.

A series of similar observations have been reported in studies investigating iEEG data for seizure prediction. Non-stationarity in iEEG data is also regarded as a major obstacle to data-driven approaches.[36] While many studies on the field assume a preictal time period of 1 h or more[42,43] the variability in actual duration between seizures has also been discussed.[44,45] Experiments with iEEG suggest that a single SOZ has the capacity to generate different seizure subtypes, each being represented by a distinct ictal network evolution.[46] Seizures of a certain subtype tended to occur closer in time, and changes follow circadian or slower time scales. There is strong evidence that a given seizure subtype may be preceded by different preictal dynamics.[47] Consequently, each subtype may also affect different networks exhibiting neurocardiac control, which might explain the different preictal HRV dynamics observed with a single SOZ. As such, progress made in one field might be transferrable to other.

The similarities in the limiting factors of iEEG and HRV data for seizure prediction reveal aspects of a broader neurocardiac information flow. This is further supported by a study showing that the predictive potential of the ECG is comparable to that of a 21-channel scalp EEG.[48]

A major limitation of our study is the short duration of analyzed ECG prior to seizure onset. While this allowed for the inclusion of more patients and seizures, thereby increasing the sample size and heterogeneity of the cohort, it also comes with trade-offs. By examining equal amounts of data for each upcoming seizure, the comparison of classification performance between patients and seizures within our cohort is less biased. However, since the 60-min interval analyzed is relatively short and our primary scope was to unveil the physiologic factors that impede its real-world application, we refrained from setting a prediction horizon and reporting clinically relevant prediction metrics. The use of a fixed 10-min preictal period was chosen to facilitate comparison with previous literature.[18–20] Such rigid definition may produce misleading results given the variability in true preictal duration and its probabilistic nature. Finally, all factors that impede the causal prediction of seizures are in most cases closely intertwined. Therefore, inferring true causality or asses their universality is challenging.

This study provides supporting evidence that ECG preictal classification might be possible for a group of patients with specific characteristics. Their seizures should preferably share a single SOZ, evoke significant preictal HRV modulation, and occur under conditions favoring data stationarity, such as sleep-related seizures. However, it is essential to note that there is no guarantee that these conditions will consistently result in successful preictal classification. This underscores the complexity of ECG-based seizure prediction on an individual basis. When

considering cross-patient prediction approaches, the inherent heterogeneity in data is expected to amplify and further undermine the predictive capacity.

The highlighted challenges and their impact on classification performance are mostly relevant when conducting causal studies. Non-causal studies often neglect such effects by design, for example in-sample optimization or inappropriate validation strategies, which can lead to overly optimistic results with limited practical utility. This major methodological concern has been underscored recently using EEG data[49] and is also evident in the differences observed in our own results. It is critical to acknowledge the current methodological weaknesses and insufficiencies of the data-driven prediction approaches to achieve significant progress. Therefore, we strongly advocate the adoption of a methodological and evaluation framework for seizure warning systems, similar to the worldwide iEEG seizure prediction competition,[43,50,51] or to the My Seizure Gauge Trial[52] for seizure forecasting. Both frameworks provide a more realistic estimation of performance by pseudo-prospectively or prospectively examining long-term data as demonstrated in several studies incorporating wearable data, including HR data.[40,52–55]

In conclusion, current seizure prediction approaches exploiting HRV short-term dynamics are not yet ready for clinical application. Given the nature of data, more probabilistic approaches should be considered in the future. The contribution of deterministic seizure prediction, complementary to seizure forecasting that exploits mid-term dynamics, remains to be determined.

ACKNOWLEDGMENTS

The work was supported by the Innovation Projects MedTech ALERT of Else Kröner Fresenius Center for Digital Health of the TU Dresden and the University Hospital Carl Gustav Carus.

CONFLICT OF INTEREST STATEMENT

The authors have no conflict of interest to declare. We confirm that we have read the Journal's position on issues involved in ethical publication and affirm that this report is consistent with those guidelines.

DATA AVAILABILITY STATEMENT

Research data are not shared.

ETHICS STATEMENT

The study was approved by the Ethics Committee of the Technische Universität Dresden (BO-EK-398082021 and BO-EK-116022021).

ORCID

Sotirios Kalousios https://orcid.org/0009-0007-7072-9576
Jens Müller https://orcid.org/0000-0001-9875-3534
Hongliu Yang https://orcid.org/0000-0002-6021-0187
Georg Leonhardt https://orcid.org/0009-0005-5892-0525

REFERENCES

1. Fiest KM, Sauro KM, Wiebe S, Patten SB, Kwon CS, Dykeman J, et al. Prevalence and incidence of epilepsy. Neurology. 2017;88(3):296–303.
2. Fisher RS, van Emde Boas W, Blume W, Elger C, Genton P, Lee P, et al. Epileptic seizures and epilepsy: definitions proposed by the international league against epilepsy (ILAE) and the International Bureau for Epilepsy (IBE). Epilepsia. 2005;46(4):470–2.
3. Baker GA. The psychosocial burden of epilepsy. Epilepsia. 2002;43(s6):26–30.
4. Laxer KD, Trinka E, Hirsch LJ, Cendes F, Langfitt J, Delanty N, et al. The consequences of refractory epilepsy and its treatment. Epilepsy Behav. 2014;1(37):59–70.
5. Löscher W, Schmidt D. Modern antiepileptic drug development has failed to deliver: ways out of the current dilemma. Epilepsia. 2011;52(4):657–78.
6. Brodie MJ, Barry SJE, Bamagous GA, Norrie JD, Kwan P. Patterns of treatment response in newly diagnosed epilepsy. Neurology. 2012;78(20):1548–54.
7. Mormann F, Andrzejak RG, Elger CE, Lehnertz K. Seizure prediction: the long and winding road. Brain. 2007;130(2):314–33.
8. Schulze-Bonhage A, Sales F, Wagner K, Teotonio R, Carius A, Schelle A, et al. Views of patients with epilepsy on seizure prediction devices. Epilepsy Behav. 2010;18(4):388–96.
9. Bruno E, Simblett S, Lang A, Biondi A, Odoi C, Schulze-Bonhage A, et al. Wearable technology in epilepsy: the views of patients, caregivers, and healthcare professionals. Epilepsy Behav. 2018;1(85):141–9.
10. Kuhlmann L, Lehnertz K, Richardson MP, Schelter B, Zaveri HP. Seizure prediction—ready for a new era. Nat Rev Neurol Oktober. 2018;14(10):618–30.
11. Morrell M. Brain stimulation for epilepsy: can scheduled or responsive neurostimulation stop seizures? Curr Opin Neurol. 2006;19(2):164–8.
12. Bruno E, Biondi A, Richardson MP. Pre-ictal heart rate changes: a systematic review and meta-analysis. Seizure. 2018;55:48–56.
13. Delamont RS, Walker MC. Pre-ictal autonomic changes. Epilepsy Res. 2011;97(3):267–72.
14. Shaffer F, McCraty R, Zerr CL. A healthy heart is not a metronome: an integrative review of the heart's anatomy and heart rate variability. Front Psychol. 2014;5:1040.
15. Bouzid Z, Al-Zaiti SS, Bond R, Sejdić E. Remote and wearable ECG devices with diagnostic abilities in adults: a state-of-the-science scoping review. Heart Rhythm Juli. 2022;19(7):1192–201.
16. Novak V, Reeves AL, Novak P, Low PA, Sharbrough FW. Time-frequency mapping of R–R interval during complex partial seizures of temporal lobe origin. J Auton Nerv Syst. 1999;77(2):195–202.
17. Kerem DH, Geva AB. Forecasting epilepsy from the heart rate signal. Med Biol Eng Comput. 2005;43(2):230–9.
18. Fujiwara K, Miyajima M, Yamakawa T, Abe E, Suzuki Y, Sawada Y, et al. Epileptic seizure prediction based on

multivariate statistical process control of heart rate variability features. IEEE Trans Biomed Eng. 2016;63(6):1321–32.
19. Pavei J, Heinzen RG, Novakova B, Walz R, Serra AJ, Reuber M, et al. Early seizure detection based on cardiac autonomic regulation dynamics. Front Physiol. 2017;8:765.
20. Billeci L, Marino D, Insana L, Vatti G, Varanini M. Patient-specific seizure prediction based on heart rate variability and recurrence quantification analysis. PLoS One. 2018;13(9):e0204339.
21. Brammer JC. Biopeaks: a graphical user interface for feature extraction from heart- and breathing biosignals. J Open Source Softw. 2020;5(54):2621.
22. Orphanidou C, Bonnici T, Charlton P, Clifton D, Vallance D, Tarassenko L. Signal-quality indices for the electrocardiogram and Photoplethysmogram: derivation and applications to wireless monitoring. IEEE J Biomed Health Inform. 2015;19(3):832–8.
23. Cortes C, Vapnik V. Support-vector networks. Mach Learn. 1995;20(3):273–97.
24. Geurts P, Ernst D, Wehenkel L. Extremely randomized trees. Mach Learn. 2006;63(1):3–42.
25. Borg I, Groenen PJF. Modern multidimensional scaling: theory and applications. 2nd ed. New York [Heidelberg]: Springer; 2005. p. 614. (Springer series in statistics).
26. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology. 1982;143(1):29–36.
27. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O. Scikit-learn: machine learning in python. J Mach Learn Res. 2011;12(85):2825–30.
28. Makowski D, Pham T, Lau ZJ, Brammer JC, Lespinasse F, Pham H, et al. NeuroKit2: a python toolbox for neurophysiological signal processing. Behav Res Methods. 2021;53(4):1689–96.
29. Leal A, Pinto MF, Lopes F, Bianchi AM, Henriques J, Ruano MG, et al. Heart rate variability analysis for the identification of the preictal interval in patients with drug-resistant epilepsy. Sci Rep. 2021;11(1):5987.
30. Behbahani S, Dabanloo NJ, Nasrabadi AM, Dourado A. Prediction of epileptic seizures based on heart rate variability. Technol Health Care off J Eur Soc Eng Med. 2016;24(6):795–810.
31. Berntson GG, Bigger JT Jr, Eckberg DL, Grossman P, Kaufmann PG, Malik M, et al. Heart rate variability: origins, methods, and interpretive caveats. Psychophysiology. 1997;34(6):623–48.
32. Jeppesen J, Fuglsang-Frederiksen A, Johansen P, Christensen J, Wüstenhagen S, Tankisi H, et al. Seizure detection based on heart rate variability using a wearable electrocardiography device. Epilepsia. 2019;60(10):2105–13.
33. Eggleston KS, Olin BD, Fisher RS. Ictal tachycardia: the head–heart connection. Seizure. 2014;23(7):496–505.
34. Kalitzin S, Koppert M, Petkov G, Velis D, da Silva FL. Computational model prospective on the observation of proictal states in epileptic neuronal systems. Epilepsy Behav. 2011;22:S102–S109.
35. 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(1):88.
36. Müller J, Yang H, Eberlein M, Leonhardt G, Uckermann O, Kuhlmann L, et al. Coherent false seizure prediction in epilepsy, coincidence or providence? Clin Neurophysiol. 2022;133:157–64.
37. Karoly PJ, Rao VR, Gregg NM, Worrell GA, Bernard C, Cook MJ, et al. Cycles in epilepsy. Nat Rev Neurol Mai. 2021;17(5):267–84.
38. Karoly PJ, Stirling RE, Freestone DR, Nurse ES, Maturana MI, Halliday AJ, et al. Multiday cycles of heart rate are associated with seizure likelihood: an observational cohort study. EBioMedicine. 2021;72:103619.
39. Proix T, Truccolo W, Leguia MG, Tcheng TK, King-Stephens D, Rao VR. Forecasting seizure risk in adults with focal epilepsy: a development and validation study. Lancet Neurol. 2021;20(2):127–35.
40. Xiong W, Stirling RE, Payne DE, Nurse ES, Kameneva T, Cook MJ, et al. Forecasting seizure likelihood from cycles of self-reported events and heart rate: a prospective pilot study. EBioMedicine. 2023;93:104656.
41. Stirling RE, Cook MJ, Grayden DB, Karoly PJ. Seizure forecasting and cyclic control of seizures. Epilepsia. 2021;62(S1):S2–S14.
42. Brinkmann BH, Wagenaar J, Abbot D, Adkins P, Bosshard SC, Chen M, et al. Crowdsourcing reproducible seizure forecasting in human and canine epilepsy. Brain. 2016;139(6):1713–22.
43. Kuhlmann L, Karoly P, Freestone DR, Brinkmann BH, Temko A, Barachant A. Epilepsyecosystem.Org: crowd-sourcing reproducible seizure prediction with long-term human intracranial EEG. Brain. 2018;141(9):2619–30.
44. Bandarabadi M, Rasekhi J, Teixeira CA, Karami MR, Dourado A. On the proper selection of preictal period for seizure prediction. Epilepsy Behav. 2015;46:158–66.
45. Mormann F, Kreuz T, Rieke C, Andrzejak RG, Kraskov A, David P. On the predictability of epileptic seizures. Clin Neurophysiol. 2005;116(3):569–87.
46. Schroeder GM, Diehl B, Chowdhury FA, Duncan JS, de Tisi J, Trevelyan AJ. Seizure pathways change on circadian and slower timescales in individual patients with focal epilepsy. Proc Natl Acad Sci. 2020;117(20):11048–58.
47. Freestone DR, Karoly PJ, Cook MJ. A forward-looking review of seizure prediction. Curr Opin Neurol. 2017;30(2):167–73.
48. Meisel C, Bailey KA. Identifying signal-dependent information about the preictal state: a comparison across ECoG, EEG and EKG using deep learning. EBioMedicine. 2019;45:422–31.
49. West J, Bozorgi ZD, Herron J, Chizeck HJ, Chambers JD, Li L. Machine learning seizure prediction: one problematic but accepted practice. J Neural Eng Januar. 2023;20(1):016008.
50. Snyder DE, Echauz J, Grimes DB, Litt B. The statistics of a practical seizure warning system. J Neural Eng Dezember. 2008;5(4):392–401.
51. Winterhalder M, Maiwald T, Voss HU, Aschenbrenner-Scheibe R, Timmer J, Schulze-Bonhage A. The seizure prediction characteristic: a general framework to assess and compare seizure prediction methods. Epilepsy Behav. 2003;4(3):318–25.
52. Brinkmann B, Nurse E, Viana P, Nasseri M, Kuhlmann L, Karoly P, et al. Seizure forecasting and detection with wearable devices and subcutaneous EEG – outcomes from the my seizure gauge trial (PL4.001). Neurology. 2023;100(17_supplement_2):4322.
53. Meisel C, El Atrache R, Jackson M, Schubach S, Ufongene C, Loddenkemper T. Machine learning from wristband sensor data for wearable, noninvasive seizure forecasting. Epilepsia. 2020;61(12):2653–66.
54. Nasseri M, Pal Attia T, Joseph B, Gregg NM, Nurse ES, Viana PF, et al. Ambulatory seizure forecasting with a wrist-worn device using long-short term memory deep learning. Sci Rep. 2021;11(1):21935.

55. Stirling RE, Grayden DB, D'Souza W, Cook MJ, Nurse E, Freestone DR, et al. Forecasting seizure likelihood with wearable technology. Front Neurol. 2021;12:704060.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Kalousios S, Müller J, Yang H, Eberlein M, Uckermann O, Schackert G, et al. ECG-based epileptic seizure prediction: Challenges of current data-driven models. Epilepsia Open. 2025;10:143–154. https://doi.org/10.1002/epi4.13073