IOP Publishing J. Neural Eng. 23 (2026) 046009 https://doi.org/10.1088/1741-2552/ae7693

Journal of Neural Engineering

Check for updates

OPEN ACCESS

RECEIVED
10 November 2025

REVISED
9 April 2026

ACCEPTED FOR PUBLICATION
2 June 2026

PUBLISHED
9 July 2026

PAPER

Multiscale predictive modeling robustly improves the accuracy of pseudo-prospective seizure forecasting in drug-resistant epilepsy

Gagan Acharya[1], **Erin Conrad**[2,3,4], **Kathryn A Davis**[2,3,4] **and Erfan Nozari**[1,5,6,7,*]

[1] Department of Electrical and Computer Engineering, University of California, Riverside, CA, United States of America
[2] Department of Neurology, Hospital of the University of Pennsylvania, Philadelphia, PA, United States of America
[3] Center for Neuroengineering and Therapeutics, University of Pennsylvania, Philadelphia, PA, United States of America
[4] Department of Bioengineering, University of Pennsylvania, Philadelphia, PA, United States of America
[5] Department of Mechanical Engineering, University of California, Riverside, CA, United States of America
[6] Department of Bioengineering, University of California, Riverside, CA, United States of America
[7] Neuroscience Graduate Program, University of California, Riverside, CA, United States of America
[*] Author to whom any correspondence should be addressed.

E-mail: erfan.nozari@ucr.edu

Keywords: seizure prediction, drug resistant epilepsy, dynamical modeling, iEEG, classification

Supplementary material for this article is available online

Abstract

Objective. Extensive research over the past two decades has focused on identifying a preictal period in scalp and intracranial encephalography (iEEG). This effort has led to numerous seizure prediction and forecasting algorithms, with moderate success on datasets consisting of curated and pre-segmented EEG. When evaluated pseudo-prospectively on continuous recordings, existing algorithms often exhibit low sensitivity, high time in warning, or both. In this study, we investigate whether predictive modeling of temporal dynamics of iEEG features and seizure risk can improve pseudo-prospective (PP) forecasting performance. *Approach.* Using iEEG data from $n = 5$ patients undergoing presurgical evaluation at the Hospital of the University of Pennsylvania and six state-of-the-art baseline models, we shift the focus from designing new features and classifiers to modeling the temporal evolution of iEEG features (classifier inputs) and seizure risk (classifier outputs). We develop autoregressive (AR) models to predict iEEG features and seizure risk over timescales of several minutes and incorporate these predictions into existing forecasting pipelines. *Main results.* We first demonstrate that a wide range of iEEG features are predictable over time, with over 99% and 35% of features achieving $R^2 > 0$ for 10 s and 10-minute-ahead predictions (mean R^2 of 0.85 and 0.2), respectively. We observe a strong correlation between feature predictability and classification-based feature importance. Accordingly, we show that incorporating an AR model that predicts iEEG features approximately 12 ± 4 min into the future improves PP performance, with a mean increase of 28% in the area under the sensitivity versus time-in-warning curve (PP-AUC). The addition of a second AR model at the level of seizure risk yields further gains, resulting in a total mean improvement of 51% in PP-AUC. *Significance.* These results provide evidence for the long-term predictability of seizure-relevant iEEG features and demonstrate the value of time-series predictive modeling for improving seizure forecasting from continuous intracranial EEG.

1. Introduction

Despite continuous advancements in antiseizure medications, over one-fifth of epilepsy patients remain refractory to pharmacological interventions (Picot *et al* 2008, Berg 2009, Fattorusso *et al* 2021). Motivated by the potential to improve the lives of patients with drug-resistant epilepsy and powered by evidence that epileptic seizures may begin well in advance of clinical onset (Litt *et al* 2001), seizure prediction and forecasting have been the subject of intensive research efforts over the past two and a half decades (Leijten *et al* 2018, Paul 2018, Sharmila and Geethanjali 2019, Kim *et al* 2020, Shoeibi *et al* 2020, Andrzejak *et al* 2023, Baud *et al* 2023, Cheng and Goldenholz 2025), all aiming to achieve the same goal: reliably identifying the preictal period with increasing temporal and statistical accuracy. The field has attracted considerable interest, as evidenced by various international initiatives and competitions (Brinkmann and Cukierski 2014, Kuhlmann *et al* 2016, 2018, Laboratory 2025), collectively propelling substantial innovation in algorithmic approaches to seizure prediction. These efforts have resulted in a diverse range of methodologies, from statistical models to deep learning frameworks. Yet, despite this progress, accurate seizure forecasting remains an elusive goal. Existing algorithms universally suffer from low sensitivity, high false-positive rates, or both, particularly when tested pseudo-prospectively on continuous electroencephalogram (EEG) data. This gap highlights a pressing need for novel approaches that go beyond the commonly pursued goals of improved feature extraction and advanced classification techniques.

Existing seizure detection, prediction, and forecasting algorithms commonly operate by splitting the time series into (possibly overlapping) segments, extracting certain features (biomarkers) from each, and training a classifier on those features (Kassiri *et al* 2017, Leijten *et al* 2018, Paul 2018, Dümpelmann 2019, Sharmila and Geethanjali 2019, Kim *et al* 2020, Shoeibi *et al* 2020, Costa *et al* 2024). The features may be purely signal-based (Liang *et al* 2011, Kassiri *et al* 2017, Skarpaas *et al* 2019, Khambhati *et al* 2024, Yang *et al* 2024, Halimeh *et al* 2025) or based on dynamical systems theory (Kiymik *et al* 2004, Chisci *et al* 2010, Aarabi and He 2014, Li *et al* 2017, Chang *et al* 2018, Solaija *et al* 2018, Rafik and Larbi 2019, Ashourvan *et al* 2020); they may be derived from time domain (Gotman 1990, Esteller *et al* 2001, Logesparan *et al* 2012, Halimeh *et al* 2025), frequency domain (Saab and Gotman 2005, Logesparan *et al* 2012, Bandarabadi *et al* 2014, Hopfengärtner *et al* 2014, Zhang and Parhi 2015), or both (Bartolomei *et al* 2008, Liang *et al* 2011, Hills 2014, Khambhati *et al* 2024, Yang *et al* 2024); and the classification may be done via a simple threshold crossing (Berényi *et al* 2012, Sun and Morrell 2014) or various state-of-the-art machine learning algorithms (Shoeb *et al* 2004, Tzallas *et al* 2009, Chen *et al* 2013, Donos *et al* 2015, Zheng *et al* 2015, Hügle *et al* 2018, Heller *et al* 2018, Kiral-Kornek *et al* 2018, Manzouri *et al* 2018, Gleichgerrcht *et al* 2022, Feng *et al* 2025). In many existing approaches, however, this formulation effectively treats the features extracted from different segments as *independent and identically distributed (i.i.d.)* observations from some underlying distributions. While certain methods incorporate temporal structure implicitly or through post hoc processing, the independence assumption remains prevalent in a large class of statistical and machine learning pipelines. This stands in contrast to mounting evidence of the multiple-timescale dynamics of ictogenesis itself (Guirgis *et al* 2015, El Houssaini *et al* 2020), as well as the various slow and ultra-slow dynamics of the circadian and multidien rhythms (Karoly *et al* 2018, Proix *et al* 2019, Stirling *et al* 2021, Baud *et al* 2023, Khambhati *et al* 2024), sleep-wake states (Rocamora *et al* 2013, Cordeiro *et al* 2015), and seizure clustering (Cook *et al* 2016, Ferastraoaru *et al* 2016) which all have well-documented modulatory effects on ictogenesis. As such, many existing approaches do not fully exploit these temporal dependencies, limiting their ability to perform true 'forecasting', where the risk of a seizure occurring s minutes later is determined not only by the brain's current state, but also by its predicted future evolution.

A number of prior works have attempted to incorporate temporal dynamics into seizure prediction, albeit in limited forms. Dynamical smoothing approaches based on state-space modeling (Zhang and Parhi 2016) treat the preictal probability from a baseline classifier as a noisy observation of an underlying latent risk, which is then estimated via a Kalman filter. While this enforces temporal consistency, the dynamics are typically simplistic and operate only at the level of the risk estimate, without modeling the evolving structure of intracranial EEG (iEEG) features. In contrast, adaptive predictive filtering methods (Garcés Correa *et al* 2019) directly exploit the predictability of iEEG signals, using deviations from expected dynamics to detect seizures. While beneficial, these approaches do not leverage supervised learning and rely on threshold-based decisions, which limits their ability to capture complex relationships between features and seizure risk. Consequently, existing methods either smooth classifier outputs or bypass learned representations altogether.

In this paper, we seek to improve the performance of existing machine learning-based seizure forecasting algorithms by augmenting them with two layers of predictive modeling: one at the level of iEEG features used as input for classification, and one at the level of seizure risk produced as output by the classification models. Our approach leverages the long-range temporal correlations in classification features to expand contextual information and refine its estimates of seizure risk by filtering historical trends in its past risk estimates. Instead of proposing a new classifier, feature set, or model architecture, we introduce a general framework that operates independently of the specific choice of algorithm or features, adding additional layers to existing seizure prediction methods in a way that requires no retraining or fine-tuning of the baseline model. By integrating this framework with various baseline models, we demonstrate its capacity to consistently improve performance across a wide range of models, from support vector machines (SVMs) and random forests to convolutional neural networks (CNNs) and transformers.

2. Methods

In this section, we describe the methodological framework used to implement and evaluate the proposed approach.

2.1. Multi-level predictive modeling for improved seizure prediction

Significant prior work has pursued the design of seizure prediction algorithms from scalp and iEEG (Leijten *et al* 2018, Paul 2018, Sharmila and Geethanjali 2019, Kim *et al* 2020, Shoeibi *et al* 2020, Andrzejak *et al* 2023, Baud *et al* 2023, Cheng and Goldenholz 2025). Instead of seeking to design another such algorithm, in this work, we take a complementary approach by choosing a seizure prediction algorithm off the shelf and *augmenting it* with predictive temporal statistical relationships that are not explicitly modeled in many existing pipelines. A typical seizure prediction/forecasting algorithm can be abstracted into a classifier f that maps a vector $\mathbf{x}[t]$ of features extracted from a moving window of (i) EEG up to time t (say, $[t-d,t]$) to the probability $q_B[t]$ that the given window is preictal (figure 1):

$$q_B[t] = f(\mathbf{x}[t])\,. \tag{1}$$

In what follows, we will refer to the function f as the *baseline model* and to time-varying probability $q_B[t]$ as the *seizure risk* at time t generated by this model. $q_B[t]$ may subsequently be thresholded to obtain a binary interictal/preictal decision if needed.

By nature, the above procedure is reactive rather than predictive. A positive (i.e. preictal) detection, even if made correctly, says something about a time window in the past (i.e. $[t-d,t]$) rather than a potentially impending seizure in the future. Therefore, we augment the above approach with two predictive, autoregressive (AR) dynamical models that are placed before and after the classifier, respectively (figure 2). The former takes the general form

$$\hat{\mathbf{x}}_f[t+s] = h(\mathbf{x}[t], \mathbf{x}[t-1], \ldots, \mathbf{x}[t-l+1])\,, \tag{2}$$

where h is a generally nonlinear dynamical model that takes as input the past l lags of classification features $\mathbf{x}$ from $t-l+1$ to t, and predicts the future evolution of iEEG features s steps into the future. We tune the structure and hyper-parameters of h (particularly, s and l) and train equation (2) as a regression model to minimize the error between predicted $\hat{\mathbf{x}}_f[t+s]$ and true $\mathbf{x}[t+s]$ features over a patient-specific training set of data, separately for each patient and each feature. Once learned, we feed the predictions of equation (2) at test time to the same classifier f, giving the predicted risk

$$q_{B+F}[t+s] = f(\hat{\mathbf{x}}_f[t+s]) \tag{3}$$

which is then used for pseudo-prospective (PP) seizure forecasting on a separate, within-patient test data. The subscript $_{B+F}$ reflects the addition of feature dynamics in equation (2) to the baseline classifier f, and distinguishes q_{B+F} from the *baseline risk* in equation (1).

A second AR dynamical model is, in turn, built to capture temporal dynamics at the output of the classifier. This model generalizes the commonsense approach of looking at a window $(q_B[t-k+1], \ldots, q_B[t])$ of seizure risks to make an interictal/preictal decision, and similar to equation (2) takes the general form (figure 2)

$$q_{B+R}[t] = \phi(q_B[t], q_B[t-1], \ldots, q_B[t-k+1])\,. \tag{4}$$

The function ϕ and the lag hyperparameter are trained separately for each patient. Given that each $q_B[t-\tau]$ term is dependent on the underlying feature $\mathbf{x}[t-\tau]$, this approach can also be considered a way to indirectly increase the iEEG context available for seizure forecasting and to refine the seizure risk $q_B[t]$ at time t based on its historical trends. Finally, we allow for the combination of feature and risk dynamics, giving rise to the combined estimate

$$\begin{aligned} q_{B+F+R}[t+s] = \phi(&q_{B+F}[t+s], q_{B+F}[t+s-1], \ldots, \\ &q_{B+F}[t+s-k+1])\,. \end{aligned}$$

In general, the choice of dynamical model (feature, risk, or combined) that leads to the best PP accuracy varies from patient to patient and needs to be optimized per patient. This patient-specific optimization requires data that capture sufficient variability in seizure dynamics within each individual, including multiple annotated seizure events and extended interictal periods. Such data are necessary both to reliably train the baseline classifiers and dynamical models, and to enable unbiased PP evaluation of their forecasting performance. We therefore base our analysis on long-duration iEEG recordings, described next.

2.2. Data

All analyses and experiments presented in this paper were conducted using publicly available iEEG data from the International Epilepsy Electrophysiology Portal (iEEG.org). The data for all patients consisted of stereoelectroencephalography (sEEG) recordings. We present results on $n = 5$ patients, each with varying seizure characteristics, electrode channels, and seizure durations, as summarized in table 1. Patient

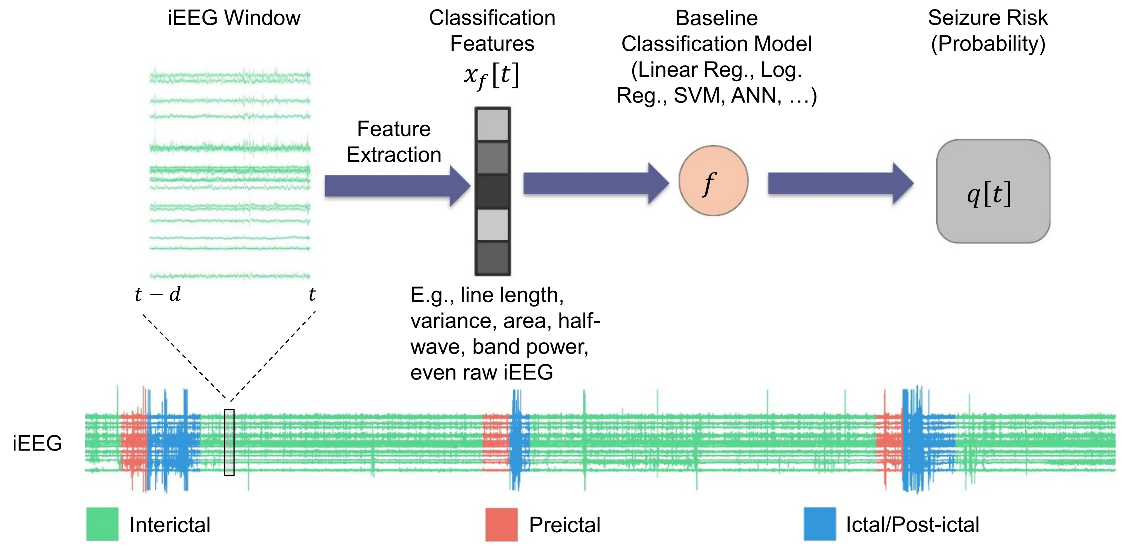


Figure 1. Standard seizure prediction pipeline. Existing seizure prediction/forecasting algorithms involve processing raw iEEG data through a series of steps to classify brain states as either preictal (seizure-impending) or interictal (baseline). This pipeline often starts with preprocessing to remove artifacts, followed by extracting relevant features ($\mathbf{x}[t]$), a classification model that uses these features to predict the preictal probability of the given window ($q[t]$), and an optional final thresholding to obtain a binary decision. While the same scheme may be used for any electrophysiological data, in this work, we use pre-surgical iEEG from $n = 5$ patients obtained from the iEEG.org database (see methods). In all of what follows, we compute features over windows of size $d = 2$ min, shifted each time by 1 step = 10 s.

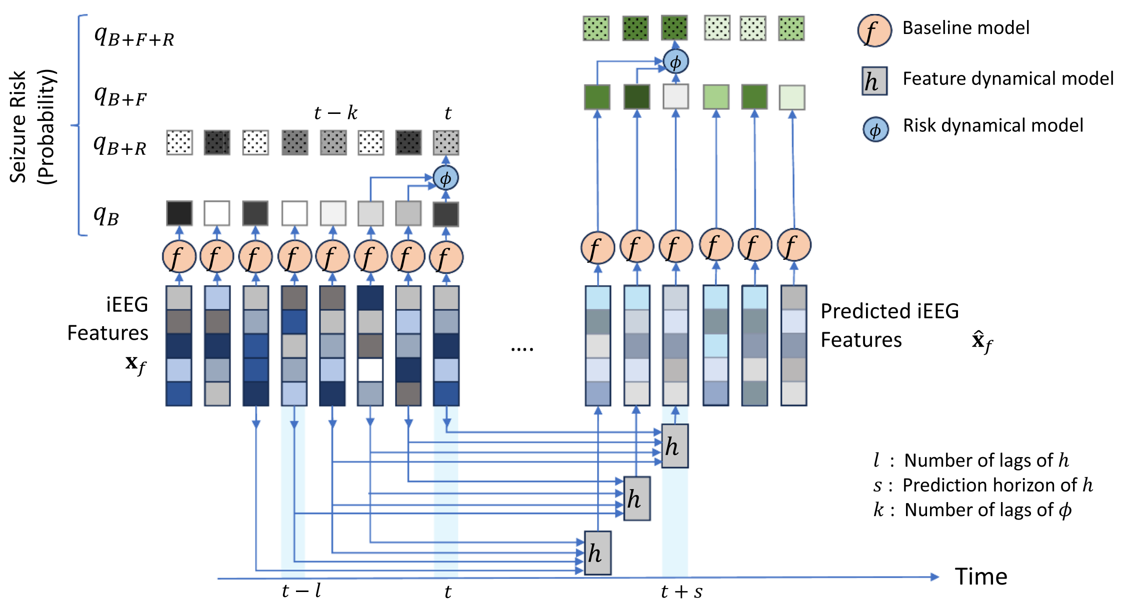


Figure 2. Proposed multi-level predictive modeling to improve seizure forecasting. Unlike the bulk of prior research, we start with an off-the-shelf seizure forecasting model f (cf figure 1) and augment it with one or both of two autoregressive predictive models: h, that operates on iEEG features (inputs to f), and ϕ that operates on seizure risk (outputs from f). All models and modeling choices, including the structure, parameters, and hyperparameters of h and ϕ are tuned separately for each patient. This results in 4 estimates of seizure risk q_M, where $M = B$ (baseline), $M = B + F$ (baseline + feature dynamics), $M = B + R$ (baseline + risk dynamics), or $M = B + F + R$ (baseline + both dynamics). Each q_M is used for pseudo-prospective seizure forecasting on a held-out continuous test duration of each patient's own iEEG.

selection was based on the availability of a sufficient number of lead seizures ($\geqslant$10), which is critical for enabling a reliable train-test split in the seizure prediction framework. While additional patients were available in the database, only a small subset satisfied this criterion, and we further restricted our analyses to five patients in order to manage the extensive computational experiments required for this study. We down-sampled all recordings to 200 Hz for uniformity and lowering the computation complexity of

Table 1. Data characteristics of the patients used in this study. 'Other cortex' refers to seizure onset zones outside the mesial temporal or temporal neocortical structures.

Patient ID	Seizure Onset Zone	# Channels	Data duration (days)	# Lead Seizures
HUP083	Other Cortex	92	13.7	11
HUP097	Multifocal	102	13.1	12
HUP101	Multifocal	102	12.3	10
HUP170	Other Cortex	228	6.7	13
HUP204	Mesial Temporal Lobe	117	6.1	10

Table 2. List of baseline seizure forecasting models used in this work.

Baseline model	Classifier (f)	Features ($\mathbf{x}$)	References
A	K-nearest neighbor	Power features, statistical features (line-length, signal average)	Based on Cook *et al* (2013)
B	Logistic regression	Power features	Based on Karoly *et al* (2017)
C	Support vector machine	Power features, power correlation, eigenvalues	'Birchwood' model in Brinkmann *et al* (2016)
D	Random forest	Power features, time correlation, eigenvalues, statistical features (channel kurtosis, standard deviation, line-length)	Team B in Kuhlmann *et al* (2018)
E	Convolutional neural network	Power features	Based on Truong *et al* (2018)
F	Transformer	Power feature	Model in Yan *et al* (2022)

feature extraction. For each patient, seizure onset and offset times were manually annotated by a trained epileptologist and accompanied the raw data on iEEG.org.

Using the inter-seizure intervals, we annotated a seizure as a lead seizure if it occurred at least 8 h away from the previous seizure offset. Interictal periods were defined as durations occurring at least 3 h away from any seizure event in both temporal directions, whereas the interval [65, 5] min before each lead seizure was labeled as preictal. For each patient, we constructed the training set using the preictal periods from the first three lead seizures, together with three neighboring interictal periods of 1 h each. To further ensure that the baseline classifier had sufficient training data, we also randomly selected one or two additional lead seizures (with matched interictal segments) for inclusion in the training set. Limiting this number to at most two strikes a balance between providing enough data for classifier training and retaining the majority of seizures for unbiased PP evaluation. The corresponding segments were excluded from testing, and all remaining iEEG data after the third lead seizure were reserved for evaluation.

2.3. Baseline classification models

As noted earlier, throughout this work, we start from a given (off-the-shelf) 'baseline' seizure forecasting model (equation (1)) and study the benefits of augmenting it with predictive models at the level of its input (iEEG features, $\mathbf{x}$) and/or output (seizure risk, q). Therefore, our approach is in general agnostic to the choice of baseline model f and its iEEG features $\mathbf{x}_f$, and we have evaluated it using a diverse set of baseline models and features. Specifically, we used six different baseline models, as summarized in table 2. We selected these models based on the following criteria: those employed in influential papers on seizure forecasting (Models A and B), those shown to be effective in crowd-sourced Kaggle competitions (Models C and D), and those representative of modern machine learning approaches (Models E and F). For the latter four, publicly available code facilitated implementation, whereas for the first two, we replicated their classifiers using our custom code based on our best understanding of the corresponding papers and their available supplementary material.

For each baseline model, we computed its respective features as originally proposed on a 2 min moving window. The window was shifted each time by 50 ms and 10 s on the training and test data, respectively. After feature extraction, we train each baseline model as a supervised binary classifier and extract its respective seizure risk $\hat{q}_B \in [0, 1]$ from before the final thresholding step. The resulting continuous seizure risk trajectory $\hat{q}_B[t]$ was computed on both the training and test samples. The former was used to train the risk-predictive models in equation (4), while the latter were used to compute each baseline model's PP performance (see evaluation protocol).

Table 3. List of autoregressive dynamical models used for s-step ahead prediction of iEEG features in equation (2). KNN: k-nearest neighbor; MLP: multi-layer perceptron.

Model	Equation
Linear	$\hat{x}_i[t+s] = \sum_{n=0}^{l-1} \alpha_n x_i[t-n]$
Quadratic	$\hat{x}_i[t+s] = \sum_{m,n=0}^{l-1} \beta_{m,n} x_i[t-m] x_i[t-n]$
KNN	$\hat{x}_i[t+s] = \frac{1}{K}\sum_{n=0}^{K-1} x_i^{\text{train}}[t_n+s]$ where $t_0,\ldots,t_{K-1}$ are time indices of the K-nearest neighbors among training data to $\left(x_i[t-l+1],\ldots,x_i[t]\right)$. We used K = 5 for all patients.
MLP	$\hat{x}_i[t+s] = \sigma\left(\mathbf{W}_2\sigma\left(\mathbf{W}_1\mathbf{x}_i^l[t]\right)\right)$ where $\sigma = \tanh$ is the activation function, $\mathbf{W}_1 \in \mathbb{R}^{H\times l}, \mathbf{W}_2 \in \mathbb{R}^{1\times H}$ are weight matrices, and $\mathbf{x}_i^l[t] = \left(x_i[t-l+1],\ldots,x_i[t]\right)$. We used $H = 50$ for all patients.
Koopman autoencoder	$\hat{x}_i[t+s] = E(\mathbf{x}_i^l[t])^\top\theta + b$ and $\mathbf{x}_i^l[t+1] = D\left(KE(\mathbf{x}_i^l[t])\right)$, where the encoder is $E(\mathbf{x}) = \mathbf{W}_e^2\sigma(\mathbf{W}_e^1\mathbf{x})$ with $\mathbf{W}_e^1 \in \mathbb{R}^{p\times l}, \mathbf{W}_e^2 \in \mathbb{R}^{p\times p}$, and the decoder is $D(\mathbf{z}) = \mathbf{W}_d^2\sigma(\mathbf{W}_d^1\mathbf{z})$ with $\mathbf{W}_d^1 \in \mathbb{R}^{p\times p}, \mathbf{W}_d^2 \in \mathbb{R}^{l\times p}$. Here, $\sigma = \text{ReLU}$, p is the latent dimension (set to 64), and $K \in \mathbb{R}^{p\times p}$ is the learnable Koopman operator that evolves the latent state forward by one step. $\theta \in \mathbb{R}^{p\times 1}$ and $b \in \mathbb{R}$. Based on Takeishi *et al* (2017).

2.4. Modeling feature dynamics

For each patient, we learned the feature-predictive model h in equation (2) as a nonlinear regression problem. We considered five different forms for h, as detailed in table 3, and tuned the parameters $\boldsymbol{\theta}$ of each model by minimizing its sum of squared error

$$J(\boldsymbol{\theta}) = \sum_t \left\| \mathbf{x}[t+s] - h\left(\mathbf{x}[t],\ldots,\mathbf{x}[t-l+1];\boldsymbol{\theta}\right)\right\|^2$$

over the training dataset. In all cases, the function h was learned element by element, such that the future values of each iEEG feature were predicted using its own past history. The hyper-parameters of each model were fixed as listed in table 3, except for the main hyper-parameters l and s, which were optimized by iterating over $l \in \{5, 10, 15\}$ min and $s \in \{1/6, 1/2, 1, 2, 5, 10, 15, 20, 30\}$ min separately for each patient.

2.5. Modeling risk dynamics

The second-level AR model ϕ implements more holistic decision making by looking at historical trends in seizure risk instead of only instantaneous risk. Further, in contrast to simpler techniques such as majority voting or ensemble methods, which aggregate *decisions* from multiple classifiers or time points, the AR model ϕ learns sequential trends in data which can lead to more nuanced decision-making. For each patient and baseline model f (equation (1)), we train ϕ as a standard supervised classification problem using the same training data used for learning the baseline model f and feature-predictive model h. The values of $\left(q_B[t-k+1],\ldots,q_B[t]\right)$ are provided as input, and the preictal/interictal labels are provided as target output. In all cases, we have used a random forest classifier for ϕ with a lag of $k = 90$ (15 min), which we have empirically observed to perform better than alternative methods such as logistic regression and k-nearest neighbors classifier (KNN). Once the classifier is trained, for each test sample we obtain $q_{B+R}[t]$ by computing the ratio of the number of positive (preictal)-classifying trees divided by the number of all decision trees in the random forest.

2.6. Evaluation protocol

In comparison to the more common approach of evaluating a seizure forecasting model based on pre-segmented, often balanced datasets, PP analysis (PPA) measures the accuracy of a seizure forecasting model based on its ability to continuously monitor an iEEG stream over multiple days and correctly predict clinically marked seizures while minimizing false alarms during extended interictal periods. Following (Snyder *et al* 2008, Karoly *et al* 2017), we implement PPA as follows. Given a test period $[t_0, t_1]$, we first evaluate the seizure risk time series $[q_M[t_0], q_M[t_0+1],\ldots,q_M[t_1]]$ where $M = B$, $B+F$, $B+R$, or $B+F+R$. For $M = B$ and $B+R$, this time series is directly shifted by s steps to compensate for the lack of an s-step ahead predictive component. We then sweep over the value of threshold τ between 0 and 1, and for each fixed value of τ we calculate the values of sensitivity (true positive rate) and time in warning as follows. At any time $t \in [t_0, t_1]$, a warning period is triggered when $q_M[t]$

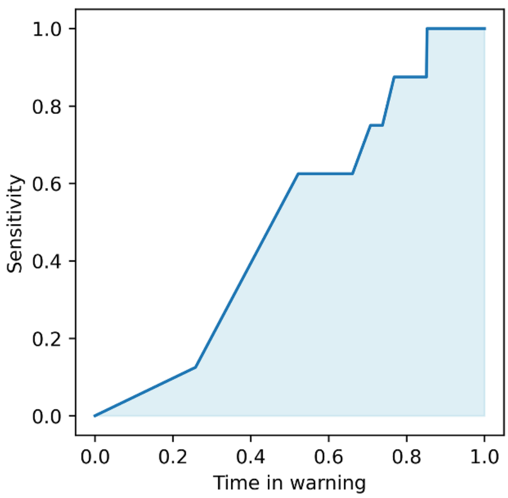


Figure 3. Illustrative example of performance metric used for comparison between methods, namely, the pseudo-prospective area under the curve (PP-AUC). This metric closely resembles the standard area under the receiver operating characteristic curve, with the main difference being the substitution of false positive rate with time in warning.

exceeds τ, and the warning remains active for a fixed duration of 30 min following the threshold crossing. This blocking is important in PPA for the stability of the warning raised by the algorithm, as an otherwise 'flickering' warning would have little utility and/or lead to major confusion in decision making. The sensitivity is then computed as the fraction of seizures that occur during an active warning period, and time in warning as the percentage of total test time spent in warning. These measures, both a function of τ, capture the burdens of false negatives and false positives, respectively.

By varying the threshold τ we then obtain the PP sensitivity-time in warning curve, as exemplified in figure 3. The bottom left and top right corners of this curve are achieved trivially by all models, corresponding to $\tau = 1$ and $\tau = 0$, respectively. We measure the accuracy of each seizure forecaster based on the area under this curve, denoted by PP area under the curve or PP-AUC. PP-AUC quantifies the trade-off between sensitivity and time in warning, allowing for an analysis of how well the model balances early detection and false alarms. PP-AUC $\simeq 1$ denotes a near-perfect model that correctly predicts all seizures with no false positives, PP-AUC $\simeq 0$ denotes the opposite, and PP-AUC $\simeq 0.3 \sim 0.4$ denotes an approximately at-chance model (chance is often lower than 0.5 because of the temporal blocking in PPA).

Unlike standard AUC for which the chance level is 0.5, even with imbalanced data, chance-level PP-AUC is often less than 0.5 because of the temporal (here, 30min) blocking in PPA. Further, we cannot simply shuffle the test samples to obtain a chance level, as seizures typically cluster in time and shuffling would destroy their clustering, which is beneficial for any (chance included) seizure forecaster. Therefore, to obtain chance-level PP-AUC, we randomize the baseline model f by training it over a shuffled training set in which the preictal/interictal labels are randomly shuffled. This randomized model is then tested in PPA as usual, ensuring a fair chance level for PP-AUC that truly reflects a randomized decision maker without altering the structure of PPA.

To allow for statistical comparisons between models, we repeated the classifier training multiple times, each time using the preictal periods from the first three lead seizures together with three neighboring interictal segments, and additionally including one or two randomly selected lead seizures (with matched interictal segments) to augment the training set. The corresponding segments were excluded from evaluation, and the remaining iEEG data after the third lead seizure were used for testing. For each random draw of training segments, seizure likelihood estimates from the trained classifier were used to calculate the corresponding PP-AUCs.

3. Results

We evaluated the proposed multi-level predictive framework on iEEG data from $n = 5$ patients with epilepsy. In this section, we begin by examining the extent to which commonly used iEEG features exhibit predictable temporal dynamics, which underlie the feasibility of our overall approach. We then investigate the relationship between the predictability of features and their relevance for seizure classification. Finally, we assess the impact of incorporating feature and risk dynamics on PP seizure forecasting performance, and compare the framework with existing dynamics-inspired methods.

3.1. iEEG features follow predictable dynamics up to 30 min into the future

We first investigated the temporal predictability of various iEEG-based features that are commonly used for seizure forecasting. This is important because at the core of the predictive scheme described above (figure 2) is the AR model in equation (2) that predicts the evolution of iEEG features into the future.[8] The viability of our scheme is therefore contingent on the presence of predictable dynamics in iEEG features $\mathbf{x}[t]$.

Using an AR form for feature dynamics as in equation (2), we examined how accurately each

[8] The AR model ϕ, while important, is secondary and plays a filtering (smoothing) role.

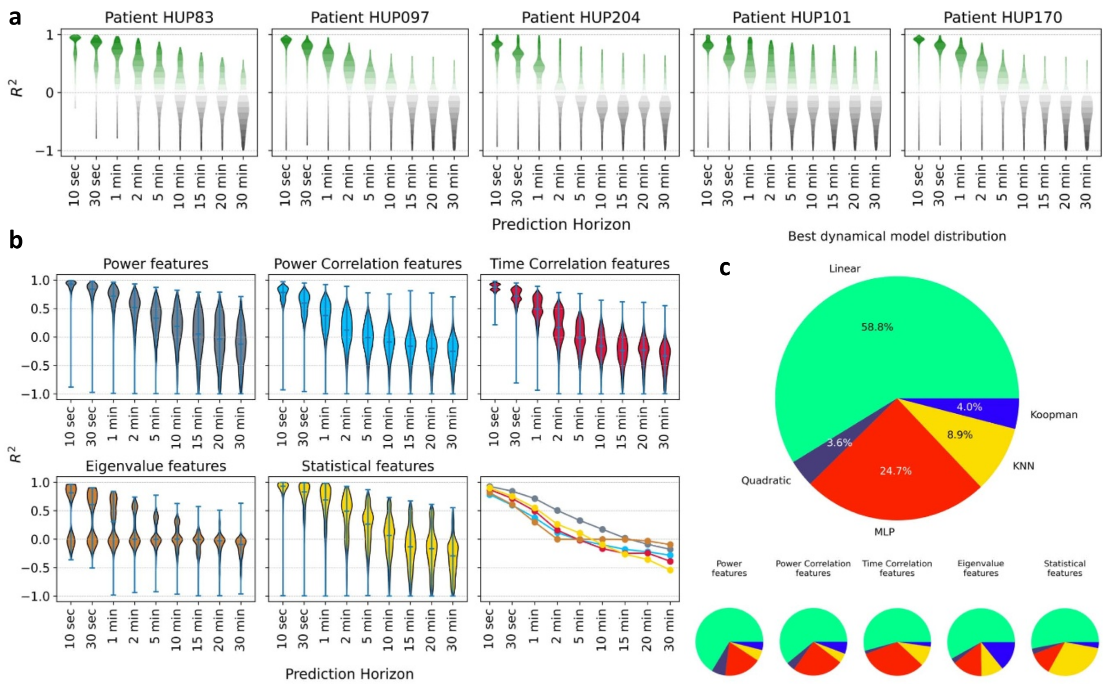


Figure 4. Long-horizon predictability of seizure-relevant iEEG features using autoregressive dynamical models. (a) Each panel, corresponding to one patient indicated at the top, shows the distribution of cross-validated R^2 of a linear s-step-ahead AR model as in equation (2) as a function of prediction horizon s. The vertical range in all panels is limited to $[-1, 1]$ for better visualization. Each violin plot is combined across all features, which are themselves a union of various sets of features used by different seizure classification models in the literature (see table 2 for a categorical breakdown and Methods for details). Green (resp., gray) indicates positive (negative) R^2 and better (worse) than chance performance. Most features remain predictable above chance up to 2–10 min, depending on the patient, and across all patients, some features remain predictable above chance up to 30 min. (b) Predictability of iEEG features by category. Data is the same as panel (a) but now combined across patients and separated by feature category (average number of features across patients: power features $\simeq 2800$, power correlation features $\simeq 8000$, time correlation features $\simeq 8000$, statistical features $\simeq 1600$, eigenvalue features $\simeq 250$). For a complete list of features in each category, see methods and supplementary material. The bottom right sub-panel shows the medians of all R^2 distributions overlaid and color-coded by feature category. Power features clearly have the longest predictability compared to other categories. (c) Distribution of AR models that explain most variance (have maximum R^2) across all features and patients. For each patient-feature, we fit five different AR models (1 linear, 4 nonlinear) and selected the one with the highest cross-validated R^2. MLP: multi-layer perceptron, KNN: k-nearest neighbor. see methods for details on each model.

iEEG feature can be predicted into the future, with a focus on assessing whether features remained predictable over extended periods. In general, the AR function h can (and often should, see figure 6) be nonlinear, but for this analysis we limited h to be linear for simplicity and computational feasibility. We then parametrically varied the horizon length s for every patient and every feature, and measured cross-validated (out-of-sample) regression accuracy (R^2) as a function of s.

The results, as shown in figure 4(a), display the distribution of regression R^2 values as a function of prediction horizon s, combined across a wide range of seizure-relevant iEEG features separately for each patient (see methods). As expected, the predictive accuracy generally decreases with prediction horizon s. Interestingly, however, most features maintained above-chance predictability ($R^2 > 0$) even over long horizons up to 10 min. Beyond the 10-minute mark, more than half of the features often exhibit a negative (lower-than-chance) predictability. Yet, in each patient, there are some features that remain predictable even up to 30 min. These results support the hypothesis that iEEG features hold intrinsic temporal structures that persist over long horizons, providing a basis for long-term seizure forecasting.

Importantly, the long-horizon predictability of iEEG features varies significantly across feature categories (time-domain, frequency-domain, etc). As shown in figure 4(b), power features have the farthest predictability, while the predictability of eigenvalue features (eigenvalues of channel cross-correlation matrices) drops to chance level most quickly (see supplementary material for details). Yet, not all frequency-domain features are necessarily better predictable than time-domain features; power correlation features, e.g. decay to chance-level more quickly than time correlation features. Though beyond the scope of this work, these differences in predictability can be very valuable for improved feature engineering and classifier design for seizure forecasting.

Further, using potentially nonlinear AR models can further improve the long-horizon predictability of iEEG features, but only marginally. For each patient and iEEG feature, we compared five different functional forms for equation (2)–the linear and

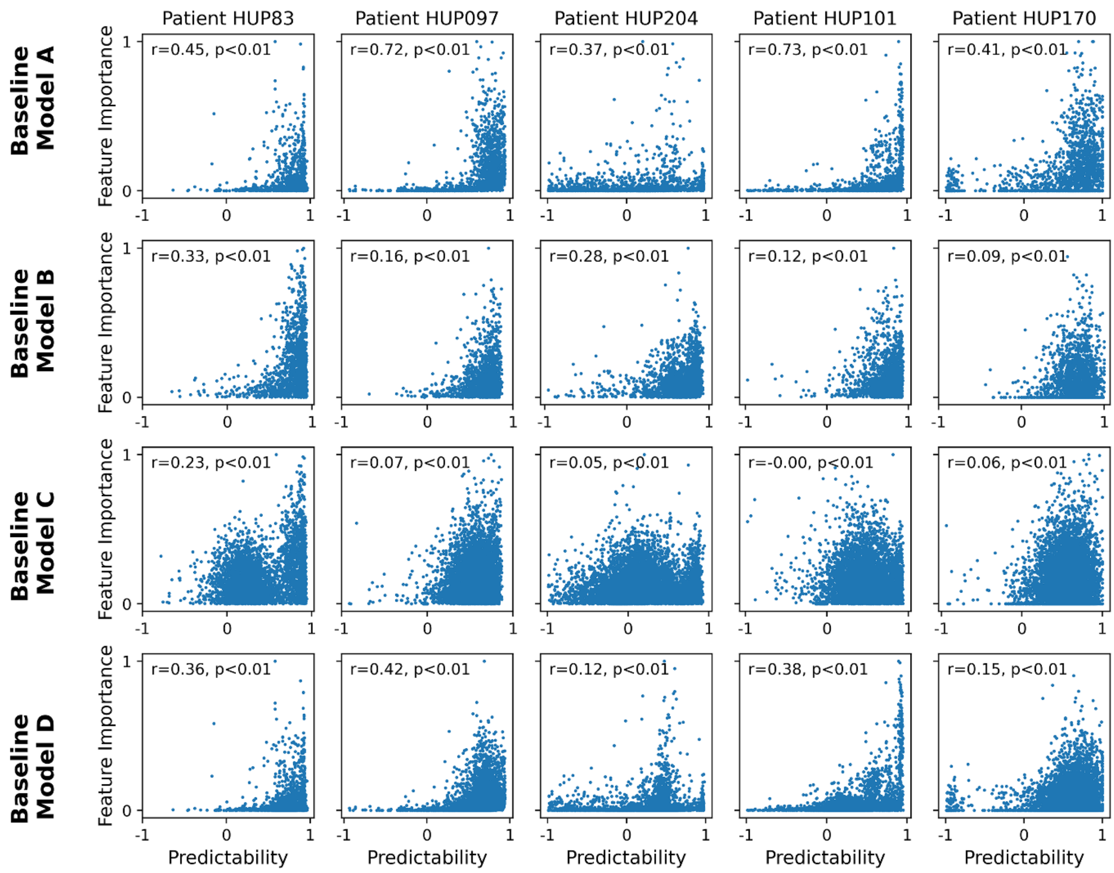


Figure 5. Presence of significant positive correlation between feature predictability and feature importance across various state-of-the-art seizure classification models. Each panel shows the relationship between feature predictability and feature importance for one patient and one classification model. Each dot represents one iEEG feature, and predictability is computed as the cross-validated R^2 of a linear AR model with a prediction horizon (s) of 2 min. The horizontal range of each panel is limited to $[-1, 1]$ for better visualization. Feature importance is computed based on the respective baseline model; see supplementary material for details. The Spearman correlation coefficient and associated p-value are shown in each panel.

4 nonlinear forms. We then found the most accurate model for each patient-feature and computed the win percentage of model across all patient-features. As shown in figure 4(c), in about three-quarters of cases, the linear model achieves the highest R^2, followed by the multi-layer perceptron artificial neural network that approximately covers the remaining quarter of patient-features. Notably, this result shows that the dynamic linearity we had found earlier at the level of raw iEEG potentials (Nozari *et al* 2020) *approximately* holds also for a wide range of iEEG features. Though weak, however, the nonlinearity of iEEG features is not negligible and will in fact prove to be critical for dynamics-augmented seizure forecasting (see figure 6).

3.2. iEEG features with higher predictability tend to also have higher importance in seizure classification

We next examined the importance of each iEEG feature for seizure classification across several state-of-the-art seizure forecasting models, and compared the importance of each feature with its long-horizon predictability (figure 5). While the exact definition of feature importance varies between classification models (see supplementary material), it generally quantifies the sensitivity of the classifier's output to variations in each input (feature) and is a metric of how relevant each feature has been for seizure classification based on training data. Note that, in general, the importance and predictability of features are independent quantities coming from two orthogonal approaches to the iEEG time series (cross-sectional seizure-focused classification vs. temporal seizure-independent regression). Therefore, the relationship between feature importance and feature predictability is critical for the success of integrating feature dynamics (equation (2)) into seizure forecasting. In particular, it is the *conjunction* of feature importance and predictability that will make dynamics-augmenting potentially beneficial, whereas either alone is unlikely to be sufficient.

As seen from figure 5, in nearly all baseline models and all patients, there exists a strong positive correlation between feature importance and feature predictability. As with most aspects of seizure

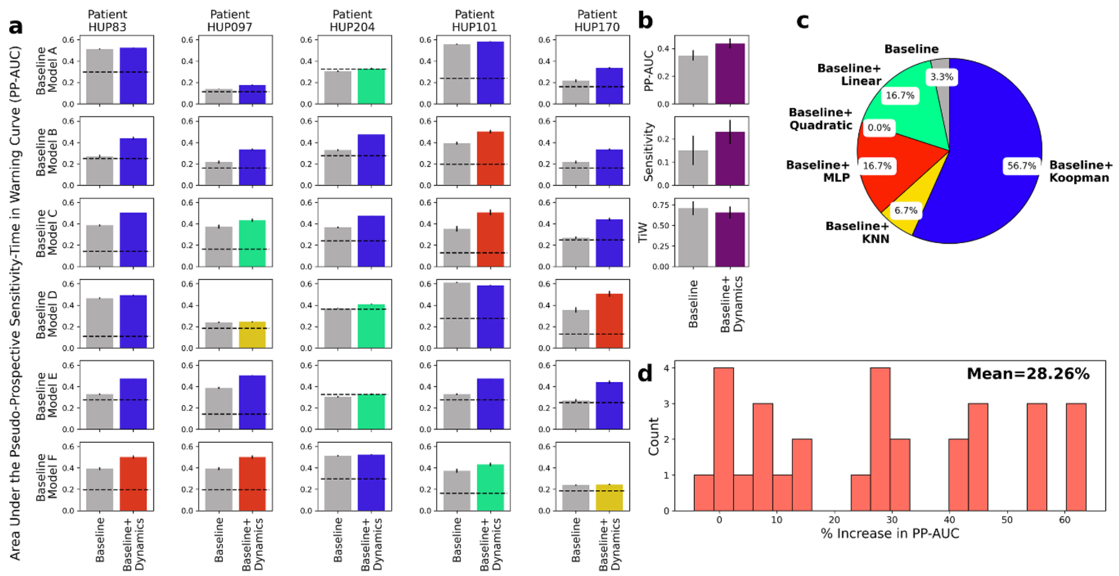


Figure 6. Pseudo-prospective seizure forecasting accuracy improves after incorporating predictive dynamical modeling of iEEG features. (a) Area under the pseudo-prospective sensitivity-time in warning curve (PP-AUC, cf figure 3 and Methods for details) for the baseline model (left, gray) and best dynamics-augmented alternative (right, colored) across five patients and six baseline models. The color of the right bar in each panel illustrates the best-performing model h and matches the color code in (c). Each bar shows mean $\pm$ 2 s.e.m. over randomized train-test splits. Black dotted horizontal lines in each panel correspond to chance-level PP-AUC. (b) Summarizing the results in panel (a) across patients and baseline models. The use of dynamics results in higher PP-AUC (top), higher sensitivity at 30% time in warning (TiW) (middle), and lower TiW at 60% sensitivity (bottom). (c) Distribution of AR models h that achieved the highest PP-AUC across all patients and baseline models. Note the major differences with the similar distribution in figure 4. (d) Distribution (histogram) of the relative improvement in PP-AUC across all patient-baseline combinations in (a). Relative improvement is measured as ($\text{PP-AUC}_{\text{B+D}}$—$\text{PP-AUC}_{\text{B}}$) / PP-AUC_{B}.

forecasting, there exists strong heterogeneity among patients as well as baseline models. In general, the positive correlation between feature importance and predictability is strongest for baseline model B and weakest for the baseline model C. Further, except for model C, this relationship is often generally monotonic and approximately exponential, with a rapid and strong increase in feature importance as predictability approaches 1. These results thus strongly support the potential benefit of dynamics-augmentation for seizure forecasting (figure 2), as we directly examine next.

3.3. Incorporating feature dynamics improves accuracy of PP seizure forecasting

Building on the established predictability of seizure-relevant iEEG features over extended time intervals, we used the estimated seizure risk q_{B+F} in equation (3) for continuous PP seizure forecasting using iEEG data from five patients from the Hospital of the University of Pennsylvania (HUP) on the iEEG.org portal. To demonstrate that our approach is robust and can adapt to various baseline models, we implemented six different state-of-the-art seizure forecasting models as the baseline model (figure 1 and equation (1)) for each patient (see methods). These models reflect a wide range of effective strategies in seizure forecasting, including models recognized for their strong performance in seizure research (Baseline Models A and B), models that have excelled in Kaggle-hosted crowdsourced competitions (Models C and D), and approaches using state-of-the-art machine learning (Models E and F). In each case, we measure seizure forecasting accuracy based on the area under the PP sensitivity-time in warning curve (PP-AUC, figure 3, see methods for details).

The results, summarized in figure 6, demonstrate a strong improvement in seizure forecasting accuracy across nearly all patients and baseline models. For each patient-baseline combination, we evaluated the same five candidates for feature dynamical modeling (h in equation (2)) as in figure 4(c) and selected the one with the highest PP-AUC. Figure 6(a) shows a breakdown of average PP-AUC for baseline and best-performing dynamics-augmented alternative across all patients and baseline models. The success rate of each dynamical model and the distribution of relative improvement in PP-AUC are summarized in panels b and c, respectively. Except for one case where the baseline model alone is slightly better than all dynamical alternatives, in all other patient-baseline combinations augmenting the baseline model with AR feature dynamics results in a higher PP-AUC. Across all patient-baseline combinations, we obtain a mean increase in PP-AUC of 28%, signifying the substantial but untapped potential of simple AR modeling for improved seizure forecasting.

Besides improving the accuracy of seizure forecasting, predictive dynamical models can provide

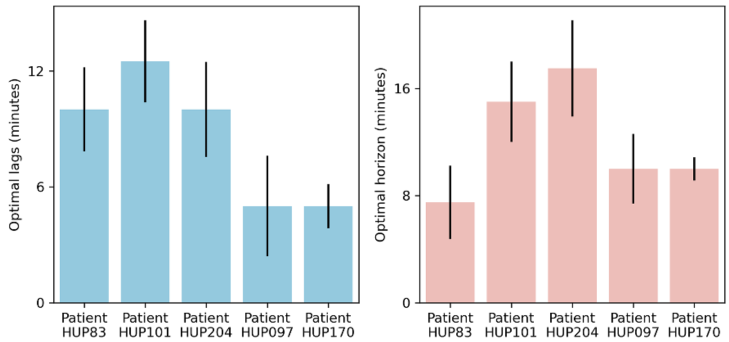


Figure 7. Optimal temporal timescale of iEEG feature dynamics for seizure forecasting. The left and right panels show the optimal number of autoregressive lags (history length) and the optimal prediction horizon (step size), respectively. Both values are obtained using the linear h model based on maximizing the PP-AUC score and reported separately per patient as median (across all features for that patient) $\pm$ 1 s.e.m. Note that both the optimal number of lags and the optimal prediction horizon vary considerably across patients, reflecting the substantial heterogeneity among epilepsy patients and the need for patient-specific modeling.

insights into the nature of seizure dynamics. In particular, the optimal length of 'past' and 'future' that maximizes seizure forecasting accuracy are important markers of the timescale at which seizures arise in each patient (figure 7). Note that these values are based on optimizing seizure forecasting accuracy (PP-AUC) and are thus different from R^2 values shown in figure 4 that reflect pure feature AR accuracy. Notably, we can see from figure 7 that both the optimal number of AR lags (past) and prediction horizons (future) are internally stable for each patient, externally variable across patients, and overall on the order of $\sim$10 min for all patients. These values, while consistent with the earliest reports of the preictal period (Litt *et al* 2001), provide detailed measures of temporal structures in iEEG that are sufficiently robust and reproducible in each patient that can be used for automated seizure forecasting.

3.4. Modeling risk dynamics further enhances seizure predictability

Next, we examined the effects of adding a second layer of predictive dynamics at the level of seizure risk, equation (4), on PP seizure forecasting. Similar to above, for each patient-baseline combination, we used a training portion of iEEG data to fit the baseline model, five feature-predictive models h, and one risk-predictive model ϕ, resulting in 12 possible combinations (baseline alone or baseline with either of the 5 forms of h, each with or without ϕ). We then evaluated the PP-AUC of each of these 12 possibilities in patient-specific test data, repeated over all train-test splits, and found the model with the highest mean PP-AUC for each patient-baseline combination.

The results, shown in figure 8, demonstrate the additional benefit of including risk dynamics over not only the baseline but also the combination of baseline and feature dynamics. Similar to figure 6, inclusion of dynamic predictive models improves PP-AUC in all except one patient-baseline combination (figure 8(a)). However, the addition of risk dynamics has further improved PP-AUC in half of the cases (figure 8(c)), bringing the total mean improvement in PP-AUC over baseline to 51% (figure 8(d)).

Figure 8(e) shows the overall distribution of PP-AUC across all patient-baseline combinations using q_B, q_{B+F}, q_{B+R}, and q_{B+F+R}. The addition of risk dynamics, both in the absence and presence of feature dynamics, plays a more regulatory role, reducing outliers and increasing the robustness and reliability of seizure forecasting via a dynamic filtering of the risk dynamics. This effect can also be seen from the detailed breakdown of seizure forecasting with and without the risk dynamics alone (supplementary figure 1) and results in only a marginal gain in the absence of feature prediction. It is the latter that is the main driver of improved PP-AUC, and the combination of the two dynamic layers that results in the best seizure forecasting overall.

3.5. Comparison with existing dynamics-inspired methods

To further contextualize these performance gains, we compare the proposed framework with existing works in the literature that incorporated temporal structure in seizure forecasting at different levels of the pipeline. The first work we compared against considered a dynamical smoothing approach based on state-space modeling (Zhang and Parhi 2016), in which the preictal probability produced by a baseline classifier is treated as a noisy observation of an underlying latent seizure risk and recursively filtered using a Kalman filtering framework. This method introduces temporal consistency through stochastic dynamics imposed on the risk estimate. The second comparison is with the adaptive predictive filtering approach of Garcés Correa *et al* (2019), which directly models the predictability of iEEG signals using channel-wise adaptive filters. Unlike classifier-based approaches, this method produces a seizure risk estimate directly from

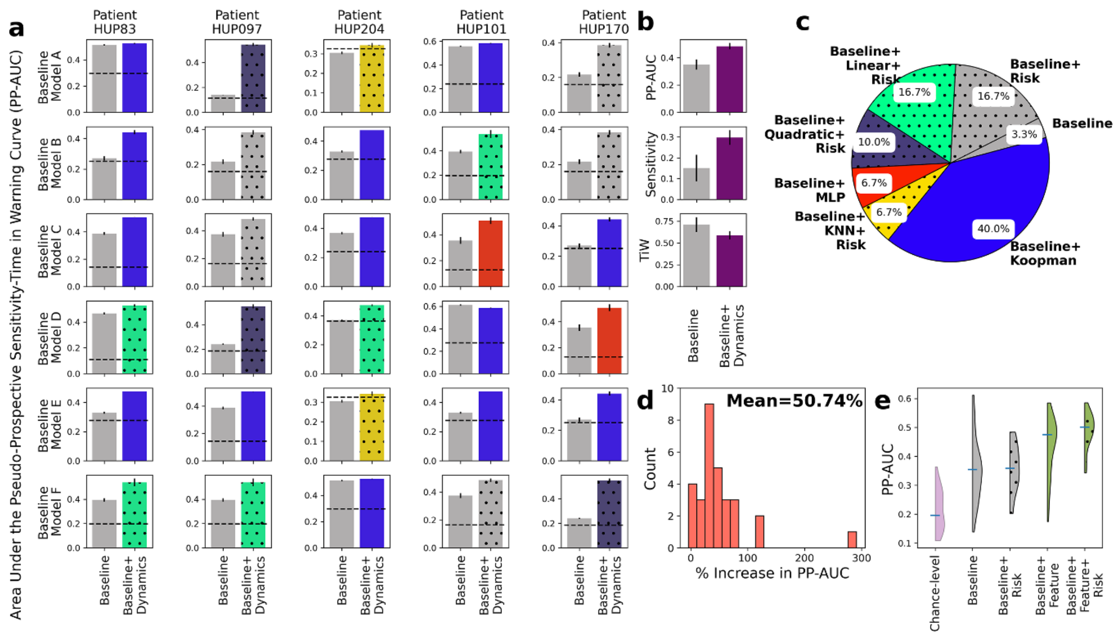


Figure 8. Further improvements in seizure forecasting accuracy with incorporation of feature *and risk* dynamics. (a) Side-by-side breakdown of improvements in PP-AUC resulting from incorporation of both feature (equation (2)) and risk (equation (4)) dynamics. Details are similar to figure 6(a). In each sub-panel, the color of the bar on the right still shows the form of feature dynamics h that resulted in the highest PP-AUC, and the dotted hatches show the presence of risk dynamics in the best-performing model. (b) Summary comparisons of PP-AUC, sensitivity at 30% TiW, and TiW at 60% sensitivity between baseline and dynamics-augmented forecasters across all patients and baseline models. Details parallel those in figure 6(b). (c) Distribution of models that achieved the highest PP-AUC across all patients and baseline models. Details are similar to figure 6(c) with the addition of dotted hatches showing the presence of risk dynamics in the best-performing model as in (a). (d) Distribution of relative improvement in PP-AUC across all patient-baseline combinations similar to figure 6(d). (e) Distributions of raw PP-AUCs for all four possible combinations of baseline models with feature and/or risk dynamics. Each distribution combines across all 5 patients and 6 baseline models.

signal dynamics without relying on a pre-trained classifier. All methods were evaluated under the same PP framework. For the Kalman filtering approach, the inputs were the preictal probabilities generated by the baseline classifiers, while the adaptive filtering method was applied directly to the raw iEEG signals as described in the supplementary material.

Figure 9 shows the distribution of PP-AUC across all patient–baseline combinations for the baseline model, the two alternative approaches to dynamics-augmentation described above, and our proposed framework incorporating both feature and risk dynamics. The Kalman filtering approach provides modest and inconsistent improvements over the baseline, reflecting the benefits of enforcing temporal consistency at the level of seizure likelihood. The adaptive filtering method, on the other hand, exhibits substantially lower predictive performance than baseline methods. Finally, our proposed framework consistently achieves higher PP-AUC across nearly all patient–baseline combinations, with statistically significant improvements over the baseline as well as both alternative approaches (Wilcoxon rank-sum test, $p < 0.001$). These differences highlight the distinct role that dynamics play in each approach. The existing approaches either (i) impose temporal smoothness on classifier outputs or (ii) extract dynamical structure directly from raw signals without leveraging discriminative feature representations. In contrast, our proposed framework explicitly models the evolution of both feature representations and seizure risk through predictive AR structures. This enables forward modeling of preictal dynamics, rather than retrospective smoothing or signal-level transformations, leading to improved forecasting performance.

3.6. Computational complexity and parameterization

While the proposed framework yields consistent improvements in forecasting performance, practical deployment requires assessing the additional computational and parametric costs incurred by the temporal modeling components. Table 4 summarizes the parameter count and per-sample computational complexity of the approaches considered in this work, as well as that of the two alternative approaches discussed in section 3.5. Assuming a linear AR model, the proposed framework introduces additional complexity that scales linearly with the number of features and temporal lags. In our implementation, typical values of l range from 30 to 60 (approximately 5–10 min of history), while $k \approx 90$ (approximately 15 min), resulting in modest computational overhead over the baseline model. Accordingly, in our experiments, we have found the additional cost

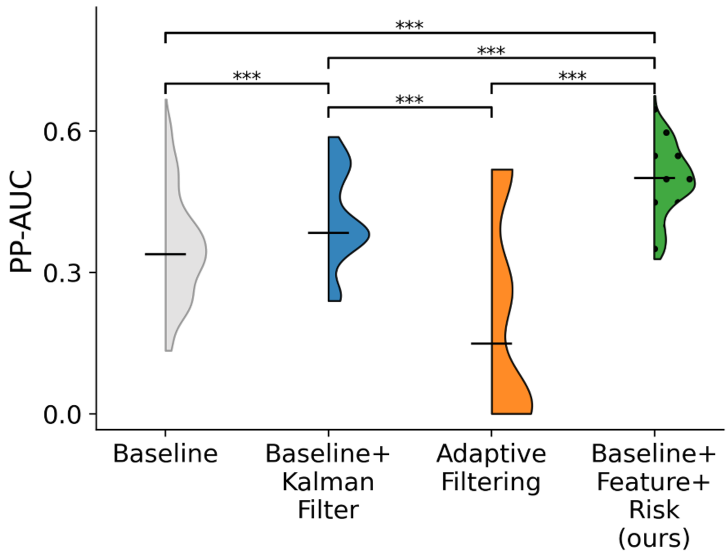


Figure 9. Comparison with existing dynamics-inspired approaches to seizure forecasting. Distributions of PP-AUC across all patient–baseline combinations for the baseline classifier, the dynamical smoothing via Kalman filtering approach of Zhang and Parhi (2016) (Baseline + Kalman Filter), the adaptive predictive filtering approach of Garcés Correa *et al* (2019) (Adaptive Filtering), and our proposed framework (Baseline + Feature + Risk). Horizontal bars indicate means. The proposed method consistently outperforms others, with statistically significant differences (Wilcoxon rank-sum test, $***p < 0.001$).

Table 4. Parameter count and per-sample computational complexity of the temporal modeling approaches proposed in this work. The complexity of the two alternative approaches discussed in section 3.5 is shown for comparison. All complexity bounds are for the additional computations required to incorporate temporal dynamics using a linear AR model and exclude the independent computational complexity of the baseline classifiers. d: feature dimension, l: lag order for feature dynamics, k: lag order for risk dynamics, C: number of channels, and L: FIR filter order used for adaptive filtering.

Method	Parameters ($\propto$ time per step)
Feature dyn. modeling ('+F')	$O(dl)$
Risk dyn. modeling ('+R')	$O(k)$
Hybrid dyn. modeling ('+F+R')	$O(dl+k)$
Kalman Filtering	$O(1)$
Adaptive Filtering	$O(CL)$

of temporal modeling to often be negligible relative to baseline feature extraction and classifier evaluation, which dominate the computational load. Thus, the improved forecasting performance that is achieved by dynamics augmentation is generally accompanied by minimal computational overhead and can be feasibly implemented on the same hardware capable of implementing the respective baseline model.

4. Discussion

Seizure prediction and forecasting research over the past two decades has largely reduced these problems to statistical classification and/or regression, and relied primarily on advancements in machine learning and data acquisition to improve outcomes. In this study, we provided pioneering evidence for the utility of an orthogonal dimension of improvement, i.e. explicitly modeling temporal dynamics over classification features and seizure risk. Using publicly available pre-surgical iEEG data from five patients at the HUP and six state-of-the-art seizure forecasting models as baseline, we showed that adding simple AR predictive models at the level of iEEG features can increase PP forecasting accuracy by an average of 28%. This improvement increased to 51% when we added a second AR model at the level of seizure risk, while the latter alone had marginal benefit. Additional analyses revealed that these benefits primarily stem from the convergence of two factors: long-term predictability of many iEEG features up to tens of minutes, and strong positive correlation between the predictability of iEEG features and their importance for seizure classification. In addition to revealing a new dimension of temporal dynamics in the context of ictogenesis, these results open the door to various future studies into the benefits of predictive modeling in specific applications of seizure prediction and forecasting.

In general, the benefits of predictive modeling in a context such as seizure forecasting depend on a trade-off between accuracy and latency. On the one hand, predictive modeling (equation (2)) uses temporal statistical correlations to predict iEEG features several minutes before they occur, and this creates an advantage in terms of latency. However, on the other hand, any predictive model will inevitably introduce errors due to the stochastic nature of the underlying time series, suboptimal modeling, etc. This creates a trade-off (which becomes stronger as the prediction horizon s increase) between using 'future' features which are only approximate, or using present features which are perfectly accurate. As such, it is neither trivial nor guaranteed that adding predictive modeling should improve seizure forecasting, and which effect dominates is, in general, patient-, baseline classifier-, and horizon-dependent. Given this context, our findings are significant as they show that for almost all patient-baseline combinations and at least some prediction horizons, it is latency that dominates accuracy.

Our study further addresses an important gap in seizure forecasting research, namely, moving beyond curated and pre-segmented preictal-interictal data and benchmarking algorithms based on PPA on continuous data. As clearly seen from figure 8(e), the median pseudo-prospective AUC across all the models and patients we tested is about 0.35 at baseline (and about 0.5 even after the addition of predictive modeling). This is alarmingly lower than commonly reported AUCs on the basis of curated data (see, e.g. Epilepsy Ecosystem (2025)). The latter values are likely to be overly optimistic and not the most accurate reflection of how each algorithm would actually perform in a clinical setting. One major reason behind this gap is the *additional difficulty of preventing false positives* when an algorithm is monitoring continuous iEEG with long interictal durations, which can contain numerous 'seizure-looking' motifs. This is a serious challenge, however, and in need of more focused attention in future research. Whether a seizure-forecasting algorithm is used to deliver advisory warnings to the patient or closed-loop neurostimulation to the brain, the additional challenges of PPA are crucial and inevitable in both settings.

In the context of PPA, notable as well are the differences between acutely- and chronically-recorded iEEG. In this study, we used pre-surgical iEEG data recorded acutely at the HUP epilepsy monitoring unit (EMU). By nature, however, EMU conditions are highly non-stationary, and recorded seizures are likely more heterogeneous in their mechanisms and dynamics than each patient's normal seizures outside of the hospital. The number of recorded seizures is also very limited in acute data, with only a few for most patients and hardly over a dozen for any patient. Both factors (abnormally high heterogeneity and very few recorded seizures) are major bottlenecks in using acute data, particularly for data-driven algorithms like those investigated here. We therefore expect seizure forecasting algorithms to reach (potentially significantly) higher PP-AUCs if used with chronically-recorded data. On the other hand, however, chronically recorded iEEG has significantly lower spatiotemporal coverage than acute iEEG, and future work is needed to rigorously test and benchmark the algorithms studied in this work on chronically recorded data. Also a promising avenue for future research is testing the benefits of predictive modeling on seizure forecasting using scalp EEG and even other physiological time series (heart rate, seizure occurrence, etc).

A long-standing question in epilepsy research is what exactly constitutes the 'preictal' period (Litt and Echauz 2002). In this work, we initially used the common one-hour pre-seizure window (precisely, 65–5min before seizure onset) for labeling data for the purpose of training each classifier. However, our predictive modeling results yielded optimal prediction horizons (i.e. value s^* that resulted in maximum PP-AUC) that varied widely across patients, ranging approximately between 7 and 20 min (figure 7). These values are interestingly consistent with the earliest reports of the preictal period (Litt *et al* 2001), and are promising from a clinical standpoint as even a five-minute warning could give patients enough time to reach a safe space or administer medication to reduce the risk of seizure occurrence. At the same time, these results highlight the importance of accurate patient-specific characterization of the preictal period, and suggest s^* as a potential data-driven biomarker that can be used for this purpose.

An interesting finding of our study was a major disagreement between the choice of predictive models of iEEG features that directly maximized regression R^2 (figure 4(c)) and the choice that maximized downstream PP forecasting accuracy (figure 6(c)). As far as the former is concerned, a linear AR model best described iEEG feature dynamics in most cases. However, complex, nonlinear models (particularly those based on the Koopman operator theory) led to greater accuracy when used in the context of seizure forecasting. This is interesting from at least two perspectives. First is the importance of end-to-end (rather than piece-by-piece) optimization when multiple analyses/models are combined within a larger pipeline. Second is the additional complexity of seizure forecasting compared to pure iEEG feature prediction. This additional complexity may be the reason why a linear model can be optimal for pure feature prediction, but a more expressive nonlinear model may give predictions that are more useful for seizure forecasting. Future work is needed to dive more deeply into and mechanistically explain the nature of the dynamics of iEEG features and seizure risk.

This study has a number of limitations. The exploratory nature of our work required us to implement and test a combinatorially large number of structural and parametric choices for each patient, each of which required running computationally expensive algorithms on continuous iEEG from tens to hundreds of channels over days to weeks of recording. Therefore, our current implementation does not impose limitations on processing power or memory usage, an assumption that will need to be addressed for future translations to implantable devices. Furthermore, in this work, we only used data from 5 patients who had a (relatively) large numbers of lead seizures (table 1). In this selection we did not further restrict the patient's type of epilepsy in order to maximize the number of lead seizures per patient, a choice that also showcases the promise of our approach across diverse epilepsies. Future work will aim to extend our findings to larger cohorts through more focused, hypothesis-driven investigations. For the same reason of computational complexity, we uniformly down-sampled all iEEG data to 200 Hz, which retains frequency content in the range 0–100 Hz. While this excludes higher-frequency oscillations that might include additional information about ictal transitions, it was inherited by the frequency range of band-power features used by our baseline models and includes, in particular, (up to) the common range of 'low-voltage fast' activity that is often monitored for seizure onset detection. Thus, while we do not anticipate significant benefits to arise from faster sampling, the model-agnostic nature of our framework is applicable to all features and baseline models regardless of their choice of sampling or frequency range.

Another limitation of this work is that our dynamical models, similar to the baseline classifiers they augment, are data-driven and do not provide a mechanistic understanding of the relationship between each individual's pathophysiology and their predictive dynamics. This, in particular, makes it harder to generalize models across individuals and limits the success of seizure forecasting to the quality and quantity of available data. Finally, the seizure annotations that we have used (i.e. 'the ground truth') have been performed manually by a board-certified epileptologist, limiting our analyses to seizures that have been documented in clinical records. Human errors and inter-rater variability have been documented and studied extensively, but manual annotations still remain the gold standard due largely to the complexities of automated seizure detection (Tveit *et al* 2023, Xu *et al* 2025).

In summary, our findings underscore the importance of slow dynamics in seizure forecasting, provide an innovative approach for using slow dynamics at the levels of iEEG features and seizure risk, and illustrate a largely untapped potential for improved seizure forecasting. These insights pave the way for future work focused on enhancing model robustness and generalizability, meta-learning across large cohorts of patients, and prospective clinical testing in both acute and chronic contexts.

5. Conclusion

In this study, we focused on long-range temporal predictability as a key and previously underutilized resource in seizure forecasting and showed that modeling the dynamics of both iEEG features and seizure risk can substantially improve PP forecasting accuracy. We achieved this via a general, model-agnostic framework that allows for integrating multi-scale dynamical system modeling into any baseline forecasting model without retraining the latter. Our results thus lay the groundwork for the integration of dynamical system identification and seizure forecasting, and open a new avenue for improving closed-loop seizure monitoring and control in epilepsy.

Acknowledgment

The research performed in this study was supported in part by the National Science Foundation Award Number 2239654 (to E.N.) and the National Institute of Health Award Numbers R01-NS-148491 (to E.N.) and R01-NS-116504 (to K.A.D.).

Data availability statement

The data that support the findings of this study are openly available at the following URL/DOI: www.ieeg.org/.

Implementation details

The iEEG data used in this study is publicly available from the iEEG.org Portal at www.ieeg.org/.

All analyses were performed in Python version 3.9 using standard scientific computing libraries, including NumPy (v1.26.4), SciPy (v1.11.1), scikit-learn (v1.3.0), and PyTorch (v2.3.1). Data preprocessing, feature extraction, and model evaluation were implemented using custom Python scripts.

Experiments were conducted on a workstation running Ubuntu 22.04 LTS and equipped with an AMD Ryzen Threadripper PRO 3975WX CPU (32 cores, 64 threads) and an NVIDIA Quadro RTX 4000 GPU (8 GB VRAM). GPU acceleration was used for model training where applicable, although all methods can be executed on CPU-only systems. Fixed random seeds were used where applicable to support reproducibility.

Author contributions

Gagan Acharya 0009-0000-7285-0950
Formal analysis (lead), Investigation (equal), Visualization (lead), Writing – original draft (equal), Writing – review & editing (equal)

Erin Conrad
Data curation (lead), Resources (supporting), Supervision (supporting), Writing – original draft (equal)

Kathryn A Davis
Data curation (supporting), Methodology (equal), Supervision (equal), Writing – original draft (equal)

Erfan Nozari 0000-0002-1981-8959
Conceptualization (lead), Methodology (equal), Supervision (lead), Validation (equal), Writing – original draft (equal), Writing – review & editing (equal)

References

Aarabi A and He B 2014 Seizure prediction in hippocampal and neocortical epilepsy using a model-based approach *Clin. Neurophysiol.* **125** 930–40

Andrzejak R G, Zaveri H P, Schulze-Bonhage A, Leguia M G, Stacey W C, Richardson M P, Kuhlmann L and Lehnertz K 2023 Seizure forecasting: where do we stand? *Epilepsia* **64** S62–S71

Ashourvan A *et al* 2020 Model-based design for seizure control by stimulation *J. Neural Eng.* **17** 026009

Bandarabadi M, Rasekhi J, Teixeira C A and Dourado A 2014 Sub-band mean phase coherence for automated epileptic seizure detection *The Int. Conf. on Health Informatics* (Springer) pp 319–22

Bartolomei F, Chauvel P and Wendling F 2008 Epileptogenicity of brain structures in human temporal lobe epilepsy: a quantified study from intracerebral EEG *Brain* **131** 1818–30

Baud M O, Proix T, Gregg N M, Brinkmann B H, Nurse E S, Cook M J and Karoly P J 2023 Seizure forecasting: bifurcations in the long and winding road *Epilepsia* **64** S78–S98

Berényi A, Belluscio M, Mao D and Buzsáki G 2012 Closed-loop control of epilepsy by transcranial electrical stimulation *Science* **337** 735–7

Berg A T 2009 Identification of pharmacoresistant epilepsy *Neurologic Clin.* **27** 1003–13

Brinkmann B H *et al* 2016 Crowdsourcing reproducible seizure forecasting in human and canine epilepsy *Brain* **139** 1713–22

Brinkmann B H and Cukierski W 2014 American epilepsy society seizure prediction challenge (Kaggle) (available at: https://kaggle.com/competitions/seizure-prediction)

Chang W-C *et al* 2018 Loss of neuronal network resilience precedes seizures and determines the ictogenic nature of interictal synaptic perturbations *Nat. Neurosci.* **21** 1742–52

Chen W *et al* 2013 A fully integrated 8-channel closed-loop neural-prosthetic CMOS SOC for real-time epileptic seizure control *IEEE J. Solid-State Circuits* **49** 232–47

Cheng J C and Goldenholz D M 2025 Seizure prediction and forecasting: a scoping review *Curr. Opin. Neurol.* **38** 135–9

Chisci L, Mavino A, Perferi G, Sciandrone M, Anile C, Colicchio G and Fuggetta F 2010 Real-time epileptic seizure prediction using ar models and support vector machines *IEEE Trans. Biomed. Eng.* **57** 1124–32

Cook M J *et al* 2013 Prediction of seizure likelihood with a long-term, implanted seizure advisory system in patients with drug-resistant epilepsy: a first-in-man study *Lancet Neurol.* **12** 563–71

Cook M J, Karoly P J, Freestone D R, Himes D, Leyde K, Berkovic S, O'Brien T, Grayden D B and Boston R 2016 Human focal seizures are characterized by populations of fixed duration and interval *Epilepsia* **57** 359–68

Cordeiro I M, Von Ellenrieder N, Zazubovits N, Dubeau F, Gotman J and Frauscher B 2015 Sleep influences the intracerebral EEG pattern of focal cortical dysplasia *Epilepsy Res.* **113** 132–9

Costa G, Teixeira C and Pinto M F 2024 Comparison between epileptic seizure prediction and forecasting based on machine learning *Sci. Rep.* **14** 5653

Donos C, Dümpelmann M and Schulze-Bonhage A 2015 Early seizure detection algorithm based on intracranial EEG and random forest classification *Int. J. Neural Syst.* **25** 1550023

Dümpelmann M 2019 Early seizure detection for closed loop direct neurostimulation devices in epilepsy *J. Neural Eng.* **16** 041001

El Houssaini K, Bernard C and Jirsa V K 2020 The epileptor model: a systematic mathematical analysis linked to the dynamics of seizures, refractory status epilepticus and depolarization block *Eneuro* **7** 0485

EPFL Embedded Systems Laboratory 2025 Seizure detection challenge (2025) (Epilepsy Benchmarks website) (available at: https://epilepsybenchmarks.com/challenge/)

Epilepsy Ecosystem 2025 Leaderboard (available at: www.epilepsyecosystem.org/leaderboard) (Accessed 18 July 2025)

Esteller R, Echauz J, Tcheng T, Litt B and Pless B 2001 Line length: an efficient feature for seizure onset detection *2001 Conf. Proc. 23rd Annual Int. Conf. of the IEEE Engineering in Medicine and Biology Society* vol 2 (IEEE) pp 1707–10

Fattorusso A, Matricardi S, Mencaroni E, Dell'Isola G B, Di Cara G, Striano P and Verrotti A 2021 The pharmacoresistant epilepsy: an overview on existent and new emerging therapies *Front. Neurol.* **12** 674483

Feng T, Ni J, Gleichgerrcht E and Jin W 2025 Seizureformer: a transformer model for IEA-based seizure risk forecasting (arXiv:2504.16098)

Ferastraoaru V, Schulze-Bonhage A, Lipton R B, Dümpelmann M, Legatt A D, Blumberg J and Haut S R 2016 Termination of seizure clusters is related to the duration of focal seizures *Epilepsia* **57** 889–95

Garcés Correa A, Orosco L L, Diez P and Leber E L 2019 Adaptive filtering for epileptic event detection in the EEG *J. Med. Biol. Eng.* **39** 912–8

Gleichgerrcht E, Dumitru M, Hartmann D A, Munsell B C, Kuzniecky R, Bonilha L and Sameni R 2022 Seizure forecasting using machine learning models trained by seizure diaries *Physiol. Meas.* **43** 124003

Gotman J 1990 Automatic seizure detection: improvements and evaluation *Electroencephalogr. Clin. Neurophysiol.* **76** 317–24

Guirgis M, Chinvarun Y, Campo M D, Carlen P L and Bardakjian B L 2015 Defining regions of interest using cross-frequency coupling in extratemporal lobe epilepsy patients *J. Neural Eng.* **12** 026011

Halimeh M, Jackson M, Loddenkemper T and Meisel C 2025 Training size predictably improves machine learning-based epileptic seizure forecasting from wearables *Neurosci. Inf.* **5** 100184

Heller S, Hügle M, Nematollahi I, Manzouri F, Dümpelmann M, Schulze-Bonhage A, Boedecker J and Woias P 2018 Hardware implementation of a performance and energy-optimized convolutional neural network for seizure detection *2018 40th Annual Int. Conf. of the IEEE Engineering in Medicine and Biology Society (EMBC)* (IEEE) pp 2268–71

Hills M 2014 Seizure detection using FFT, temporal and spectral correlation coefficients, eigenvalues and random forest *Technical Report* 1–10 (Github)

Hopfengärtner R, Kasper B S, Graf W, Gollwitzer S, Kreiselmeyer G, Stefan H and Hamer H 2014 Automatic seizure detection in long-term scalp EEG using an adaptive thresholding technique: a validation study for clinical routine *Clin. Neurophysiol.* **125** 1346–52

Hügle M, Heller S, Watter M, Blum M, Manzouri F, Dumpelmann M, Schulze-Bonhage A, Woias P and Boedecker J 2018 Early seizure detection with an energy-efficient convolutional neural network on an implantable microcontroller *2018 Int. Joint Conf. on Neural Networks (IJCNN)* (IEEE) pp 1–7

Karoly P J, Goldenholz D M, Freestone D R, Moss R E, Grayden D B, Theodore W H and Cook M J 2018 Circadian and circaseptan rhythms in human epilepsy: a retrospective cohort study *Lancet Neurol.* **17** 977–85

Karoly P J, Ung H, Grayden D B, Kuhlmann L, Leyde K, Cook M J and Freestone D R 2017 The circadian profile of epilepsy improves seizure forecasting *Brain* **140** 2169–82

Kassiri H, Tonekaboni S, Tariqus Salam M, Soltani N, Abdelhalim K, Velazquez J L P and Genov R 2017 Closed-loop neurostimulators: a survey and a seizure-predicting design example for intractable epilepsy treatment *IEEE Trans. Biomed. Circuits Syst.* **11** 1026–40

Khambhati A N, Chang E F, Baud M O and Rao V R 2024 Hippocampal network activity forecasts epileptic seizures *Nat. Med.* **30** 2787–90

Kim T, Nguyen P, Pham N, Bui N, Truong H, Ha S and Vu T 2020 Epileptic seizure detection and experimental treatment: a review *Front. Neurol.* **11** 701

Kiral-Kornek I *et al* 2018 Epileptic seizure prediction using big data and deep learning: toward a mobile system *EBioMedicine* **27** 103–11

Kiymik M K, Subasi A and Rıza Ozcalık H 2004 Neural networks with periodogram and autoregressive spectral analysis methods in detection of epileptic seizure *J. Med. Syst.* **28** 511–22

Kuhlmann L *et al* 2018 Epilepsyecosystem. org: crowd-sourcing reproducible seizure prediction with long-term human intracranial EEG *Brain* **141** 2619–30

Kuhlmann L, LizLopez, O'Connell M, rudyno5, Wang S and Cukierski W 2016 Melbourne university aes/mathworks/nih seizure prediction (Kaggle) (available at: https://kaggle.com/competitions/melbourne-university-seizure-prediction)

Leijten F S S *et al* 2018 Multimodal seizure detection: a review *Epilepsia* **59** 42–47

Li A, Inati S, Zaghloul K and Sarma S 2017 Fragility in epileptic networks: the epileptogenic zone *2017 American Control Conf. (ACC)* (IEEE) pp 2817–22

Liang S-F, Liao Y-C, Shaw F-Z, Chang D-W, Young C-P and Chiueh H 2011 Closed-loop seizure control on epileptic rat models *J. Neural Eng.* **8** 045001

Litt B *et al* 2001 Epileptic seizures may begin hours in advance of clinical onset: a report of five patients *Neuron* **30** 51–64

Litt B and Echauz J 2002 Prediction of epileptic seizures *Lancet Neurol.* **1** 22–30

Logesparan L, Casson A J and Rodriguez-Villegas E 2012 Optimal features for online seizure detection *Med. Biol. Eng. Comput.* **50** 659–69

Manzouri F, Heller S, Dümpelmann M, Woias P and Schulze-Bonhage A 2018 A comparison of machine learning classifiers for energy-efficient implementation of seizure detection *Front. Syst. Neurosci.* **12** 43

Nozari E, Stiso J, Caciagli L, Cornblath E J, He X, Bertolero M A, Mahadevan A S, Pappas G J, and Bassett D S 2020 Is the brain macroscopically linear? a system identification of resting state dynamics *bioRxiv* (available at: www.biorxiv.org/content/early/2020/12/22/2020.12.21.423856)

Paul Y 2018 Various epileptic seizure detection techniques using biomedical signals: a review *Brain Inf.* **5** 6

Picot M-C, Baldy-Moulinier M, Daurès J-P, Dujols P and Crespel A 2008 The prevalence of epilepsy and pharmacoresistant epilepsy in adults: a population-based study in a western european country *Epilepsia* **49** 1230–8

Proix T, Truccolo W, Leguia M G, King-Stephens D, Rao V R, and Baud M O 2019 Forecasting seizure risk over days *MedRxiv* 19008086

Rafik D and Larbi B 2019 Autoregressive modeling based empirical mode decomposition (EMD) for epileptic seizures detection using eeg signals *Trait. Signal* **36** 273–9

Rocamora R, Andrzejak R G, Jiménez-Conde J and Elger C E 2013 Sleep modulation of epileptic activity in mesial and neocortical temporal lobe epilepsy: a study with depth and subdural electrodes *Epil. Behav.* **28** 185–90

Saab M E and Gotman J 2005 A system to detect the onset of epileptic seizures in scalp EEG *Clin. Neurophysiol.* **116** 427–42

Sharmila A and Geethanjali P 2019 A review on the pattern detection methods for epilepsy seizure detection from EEG signals *Biomed. Eng./Biomed. Tech.* **64** 507–17

Shoeb A, Edwards H, Connolly J, Bourgeois B, Ted Treves S and Guttag J 2004 Patient-specific seizure onset detection *Epil. Behav.* **5** 483–98

Shoeibi A *et al* 2020 Epileptic seizure detection using deep learning techniques: a review (arXiv:2007.01276)

Skarpaas T L, Jarosiewicz B and Morrell M J 2019 Brain-responsive neurostimulation for epilepsy (RNS® system) *Epil. Res.* **153** 68–70

Snyder D E, Echauz J, Grimes D B and Litt B 2008 The statistics of a practical seizure warning system *J. Neural Eng.* **5** 392

Solaija M S J, Saleem S, Khurshid K, Hassan S A and Kamboh A M 2018 Dynamic mode decomposition based epileptic seizure detection from scalp EEG *IEEE Access* **6** 38683–92

Stirling R E, Cook M J, Grayden D B and Karoly P J 2021 Seizure forecasting and cyclic control of seizures *Epilepsia* **62** S2–S14

Sun F T and Morrell M J 2014 The RNS system: responsive cortical stimulation for the treatment of refractory partial epilepsy *Expert Rev. Med. Devices* **11** 563–72

Takeishi N, Kawahara Y and Yairi T 2017 Learning Koopman invariant subspaces for dynamic mode decomposition *Advances in Neural Information Processing Systems* vol 30

Truong N D, Nguyen A D, Kuhlmann L, Bonyadi M R, Yang J, Ippolito S and Kavehei O 2018 Convolutional neural networks for seizure prediction using intracranial and scalp electroencephalogram *Neural Netw.* **105** 104–11

Tveit J *et al* 2023 Automated interpretation of clinical electroencephalograms using artificial intelligence *JAMA Neurology* **80** 805–12

Tzallas A T, Tsipouras M G and Fotiadis D I 2009 Epileptic seizure detection in EEGs using time–frequency analysis *IEEE Trans. Inf. Technol. Biomed.* **13** 703–10

Wagenaar J B *et al* 2013 (https://doi.org/10.1109/NER.2013.6696201)

Xu Z *et al* 2025 Annotating neurophysiologic data at scale with optimized human input *J. Neural Eng.* **22** 046003

Yan J, Li J, Xu H, Yu Y and Xu T 2022 Seizure prediction based on transformer using scalp electroencephalogram *Appl. Sci.* **12** 4158

Yang H, Mueller J, Eberlein M, Kalousios S, Leonhardt G, Duun-Henriksen J, Kjaer T and Tetzlaff R 2024 Seizure forecasting with ultra long-term EEG signals *Clin. Neurophysiol.* **167** 211–20

Zhang Z and Parhi K K 2015 Low-complexity seizure prediction from iEEG/sRRG using spectral power and ratios of spectral power *IEEE Trans. Biomed. Circuits Syst.* **10** 693–706

Zhang Z and Parhi K K 2016 Seizure prediction using long-term fragmented intracranial canine and human EEG recordings *2016 50th Asilomar Conf. on Signals, Systems and Computers* (IEEE) pp 361–5

Zheng Y-X, Zhu J-M, Qi Y, Zheng X-X and Zhang J-M 2015 An automatic patient-specific seizure onset detection method using intracranial electroencephalography *Neuromodulation* **18** 79–84

Supplementary Material

Multiscale predictive modeling robustly improves the accuracy of pseudo-prospective seizure forecasting in drug-resistant epilepsy

Gagan Acharya[1], Erin Conrad[2,3,4], Kathryn A. Davis[2,3,4], and Erfan Nozari[1,5,6,7,*]

[1] *Department of Electrical and Computer Engineering, University of California, Riverside, CA, USA*
[2] *Department of Neurology, Hospital of the University of Pennsylvania, Philadelphia, Pennsylvania, USA*
[3] *Center for Neuroengineering and Therapeutics, University of Pennsylvania, Philadelphia, Pennsylvania, USA*
[4] *Department of Bioengineering, University of Pennsylvania, Philadelphia, Pennsylvania, USA*
[5] *Department of Mechanical Engineering, University of California, Riverside, CA, USA*
[6] *Department of Bioengineering, University of California, Riverside, CA, USA*
[7] *Neuroscience Graduate Program, University of California, Riverside, CA, USA*
[*] *Corresponding author (email: erfan.nozari@ucr.edu)*

1 Additional Experimental Details

1.1 Comparison Methods

To evaluate the proposed dynamical modeling framework, we compare its performance against two representative baseline approaches that incorporate temporal structure in distinct ways.

Adaptive Predictive Filtering We implement an adaptive predictive filtering approach to generate a continuous estimate of seizure probability from intracranial EEG. For each channel, an adaptive finite impulse response (FIR) filter is trained online to predict the current signal from its delayed past using a least mean squares (LMS) update with step size μ. The filter uses a delay of D seconds and a filter order of L, which together determine the temporal context used for prediction. The outputs of these channel-wise predictors are averaged, squared, and smoothed over a window of duration T_s to form a global activity measure that reflects the instantaneous predictability of the signal. As the system approaches a seizure, deviations from the learned dynamics increase, resulting in a rise in this measure prior to electrographic onset. An adaptive threshold, defined as k times the moving median of the activity measure over a window of duration T_m, captures baseline variability. The ratio between the activity measure and this threshold is then mapped through a smooth nonlinear transformation to obtain a continuous estimate of seizure probability. This probability signal is subsequently aggregated over non-overlapping intervals of duration T_p to yield a lower rate representation suitable for seizure prediction.

Kalman Smoothing of Seizure Likelihood We also consider a dynamical state-space approach in which the preictal probability is modeled as a noisy observation of an underlying latent state with smooth temporal evolution. In contrast to heuristic methods such as moving averages or exponential smoothing, this formulation enforces temporal consistency through an explicit evolution model on the latent state and its rate of change.

Let $y_k \in [0, 1]$ denote the preictal probability at time k. We introduce a two-dimensional latent state $\mathbf{x}_k = [x_k, \dot{x}_k]^T$, where x_k represents the underlying signal and $\dot{x}_k$ its rate of change. The state evolves according to a linear dynamical system,

$$\mathbf{x}_k = \begin{bmatrix} 1 & \Delta t \\ 0 & 1 \end{bmatrix} \mathbf{x}_{k-1} + \mathbf{w}_k, \tag{1}$$

where $\mathbf{w}_k \sim \mathcal{N}(0, Q)$ denotes process noise capturing slow variations in the latent dynamics. The observation model is given by

$$y_k = \begin{bmatrix} 1 & 0 \end{bmatrix} \mathbf{x}_k + v_k, \tag{2}$$

where $v_k \sim \mathcal{N}(0, R)$ represents observation noise arising from variability in classifier outputs.

The latent state $\mathbf{x}_k$ is estimated recursively using a Kalman filtering framework, yielding a smoothed estimate $\hat{x}_k$ of the underlying signal. Consistent with prior work (Zhang and Parhi (2016)), the observation noise variance is chosen to be larger than the process noise variance ($R \gg Q$), promoting smooth temporal evolution and reducing rapid fluctuations in y_k. The filtered estimate $\hat{x}_k$ is used as the smoothed seizure likelihood. This method serves as a dynamic baseline for comparison with the proposed approach.

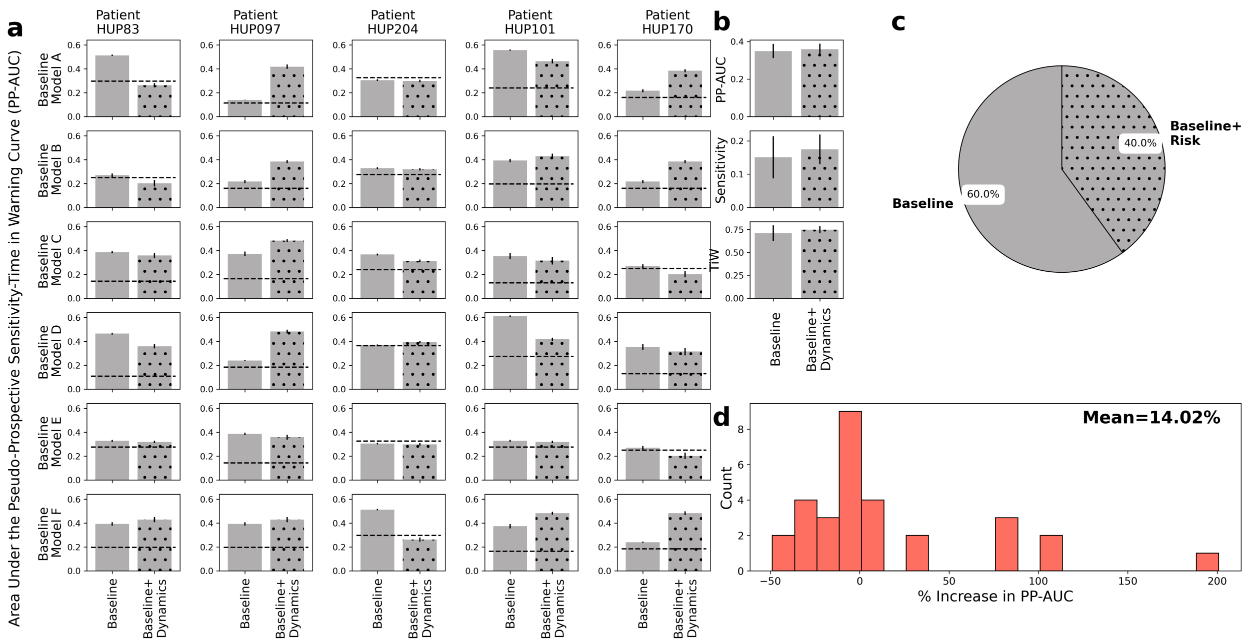
1.2 Feature categories

The features used for model training were grouped into five categories to better illustrate their differences in dynamics (Main Figure 4). *Power features* capture the spectral content of the iEEG by computing band-limited power for each channel, either in canonical frequency bands (delta, theta, alpha, beta, gamma) or in equally spaced frequency bins used in Baseline Models C. In total, for each patient we had $23C$ power features, where C denotes the number of iEEG channels (see Table 2). *Power correlation features* measure the interdependence of spectral activity across channels by computing correlation matrices between band power time series, e.g., correlating the power in band 1 of channel 12 with the power in band 3 of channel 4. The total number of power correlation features equals $C \times (C-1)/2$ for each patient. *Time correlation features* similarly compute pairwise correlations, but directly on the raw iEEG time series rather than on band powers, quantifying linear relationships between channels in the time domain. Similar to power correlation features, for each patient the total number of time correlation features is given by $C \times (C-1)/2$. *Eigenvalue features* summarize the global structure of these correlation matrices by using their eigenvalues ($2C$ features in total), which capture overall network synchrony and connectivity patterns. Finally, *Statistical features* are various simple time-domain signal statistics from each channel, which may vary slightly across baseline models. Typical examples include line length, mean, standard deviation, and the standard deviation of first differences. These features capture basic amplitude and variability characteristics of the signals. In total, we had $13C$ statistical features for each patient.

1.3 Computation of feature importance for baseline models

To relate feature predictability (from the feature dynamical model) to their relevance for classification, we computed feature importance scores for the baseline classifiers using model-appropriate approaches. For k-nearest neighbors (Model A), where importance is not explicitly defined, we used a permutation approach: feature values were shuffled across samples, and the resulting drop in accuracy was recorded. For logistic regression and linear SVM (Models B and C), importance was taken as the absolute value of the learned coefficients or weight vector, reflecting the strength of association between features and the decision boundary. For the random forest (Model D), importance was quantified by the mean decrease in Gini impurity, as implemented by the `scikit-learn` package in Python. For the deep learning models (Models E and F), we did not compute feature importance, since attribution methods (e.g., saliency maps) are computationally intensive and beyond the scope of this study.

2 Additional Results



Supplementary Figure 1: **Relative improvements in pseudo-prospective forecasting accuracy using risk dynamical modeling alone.** Panels parallel those in Main Figures 6 and 8. Note the lack of color coding (corresponding to different feature dynamical models which are absent here) and the use of dotted hatches for risk-augmented models.