
Computational Approaches to Seizure Forecasting

*Thesis submitted in partial fulfillment of the requirements for
the degree of “MASTER OF SCIENCE”*

By
NOAM SIEGEL



Submitted to the Department of Computer Sciences at
Ben-Gurion University of the Negev
The Faculty of Natural Sciences

JANUARY 2023

BEER SHEVA

Computational Approaches to Seizure Forecasting

*Thesis submitted in partial fulfillment of the requirements for
the degree of “MASTER OF SCIENCE”*

By
NOAM SIEGEL

This work was carried under the supervision of
PROF. OREN SHRIKI
AND
DR. DAVID TOLPIN

The Department of Computer Sciences

Approved by:

Student

Date

Supervisor

Date

Supervisor

Date

Chairperson of the Committee for
Graduate Studies

Date

JANUARY 2023

BEER SHEVA

Acknowledgements

I would like to express my deepest appreciation to Dr. Oren Shriki. Thank you for allowing me to travel my own path in the Garden of Science you've helped cultivate. Your heartfelt mentorship and support during my two years at the Computational Psychiatry Lab will never be forgotten.

This endeavor would not have been possible without Dr. David Tolpin. Thank you for advising, teaching and collaborating with me throughout my journey. I learned from you how to articulate basic scientific questions, and how to effectively apply myself to answering them.

Special thanks to my teachers and administrative staff at Ben-Gurion University of the Negev, and especially the Computer Science, Mathematics and Cognitive Science departments. You have provided me all of the educational and physical resources I could ask for to realize my academic dreams.

I could not have undertaken this journey without my lab mates. Thank you for giving me inspiration, fruitful arguments and being there for me in times of need: Aviv Dotan, Miriam Guendelman, Daniel Polyakov, Lahav Foox, Miriam Dissen, Asaf Harel, Shay Ben-Sasson, Adi Nissim, Yarden Ben-Horin, Sagi Furman, Ofer Avin, Ophir Almagor, Anat Shkedy-Rabani, Or Rabani.

To Miriam Guendelman, again, for your insightful remarks and editing suggestions. I hope to be able to pay you back soon.

To the Mayo Foundation for Medical Education and Research, for granting me permission to use your art (figure 3.1).

To my family - grandparents, parents, siblings, fiancé and others. Your inspiring achievements, loving support and affirmative relationships, at all times, are the main reasons why I'm here today. If it were not for you, this thesis would not exist.

To Shir

"The theory of probabilities is at bottom nothing but common sense reduced to calculus"

Essai Philosophique sur les Probabilités
Pierre-Simon Laplace

Abstract

Computationally modeling seizure timing in epilepsy from EEG would allow wearable devices to alert users before a seizure begins. Community focus on pattern recognition methods has achieved encouraging results in discriminating pre-seizure from normal brain function, as represented by EEG feature vectors. However, since seizure labeling is not available at scale, the success of these pattern recognition techniques will be constrained by out-of-budget labels as data grows.

We propose a Bayesian alternative to classifiers for seizure detection and prediction. It is label efficient and is shown to score 0.88 AUC-ROC with zero labels on a seizure detection task. A weakly supervised version which is biased towards circadian rhythms is shown to improve detection in canines. Our method composes two stages: First, the EEG distribution is modeled and is used to assign novel sequences higher seizure likelihoods. Then, a priori knowledge based on the time covariate is added for supervised improvements.

Preface

People who suffer from epilepsy undergo recurring *seizures* - a pervasive synchronous neuronal firing clinically manifested as loss of consciousness or motor control. These episodes are also termed *ictal* episodes. The uncertainty associated with seizure occurrences is deemed to be the leading cause of fear, stress and other comorbidities in patients with epilepsy. Therefore, a reliable method for estimating the likelihood of a near-term seizure is in the interest of epilepsy patients and their caregivers.

The field of seizure timing has grown into a branch of science, attracting almost every kind of expert - engineers, physicians and data scientists, to name a few. With this expansion, I feel lucky to have examined two types of approaches to seizure forecasting thoroughly. I hope that the report of my research will make it's way to an interested readership.

Noam Siegel BE'ER SHEVA, Israel, August 2022.

Contents

Abstract	i
Preface	ii
List of Figures	v
List of Tables	vi
1. Introduction	1
1.1. Background	1
1.1.1. Epileptic seizures	1
1.1.2. EEGs: electric potential recordings	1
1.1.3. Detecting and predicting seizures	2
1.1.4. Probabilistic modeling of seizure risk in epilepsy	3
1.1.5. Supervised machine learning and the data labeling problem	3
2. Pattern Recognition for Seizure Prediction	4
2.1. Methods	4
2.1.1. Preprocessing	4
2.1.2. Preictal and Interictal Binarization	5
2.1.3. Bivariate Feature Extraction Methods	6
2.1.4. Classifiers	9
2.2. Results	9
2.2.1. Evaluation Metrics	9
2.2.2. Classifier and Feature Comparison	10
2.2.3. Training Efficiency	10
2.2.4. Comparing Feature Sets	10
2.3. Conclusion & Discussion	12
3. Bayesian Estimation of Seizure Likelihood	15
3.1. Background	15
3.1.1. Problem description	15
3.1.2. Probabilistic representation of EEG and seizures in time	17
3.2. Model specification	17
3.2.1. EEG embedding	18
3.2.2. Bayesian estimation	18

3.3.	Empirical results	24
3.3.1.	Data	24
4.	Conclusion	26
Bibliography		28
Appendix A. Density estimation: mathematical derivations		33
A.1.	Embedding time series with Gaussian processes hyperparameters	33
A.2.	Gaussian Mixture Models	35
Appendix B. GP embedding: training details		36
Hebrew Abstract		37

List of Figures

1.1.	Electroencephalography (EEG) art	2
2.1.	EEG channel locations used in this chapter	5
2.2.	First 2 principal components of transformed data per patient and feature type	7
2.3.	Two types of bivariate features	8
2.4.	Three types of bivariate features	9
2.5.	Mean AUC-ROC scores for 5-fold cross validation.	11
2.6.	Standard deviation of AUC-ROC scores for 5-fold cross validation.	11
2.7.	Mean training times (in seconds) for fitting the classifiers to the datasets.	12
2.8.	Results of Linear SVM on different feature datasets	13
3.1.	Data collection and required output	16
3.2.	Sample space $\Omega_E \times \Omega_S$	17
3.3.	Probabilistically Modeling Seizure Occurrence, EEG signal and Annotations across Time	18
3.4.	EEG embedding in Gaussian process hyperparameter space	19
3.5.	Contour plots for multidimensional density estimation	20
3.6.	Circular Gaussian (von Mises) distribution and circadian profile	22
3.7.	Circadian Seizure Histogram and Circadian Profile	23
3.8.	Empirical results	25
D.1.	Hebrew Abstract	37

List of Tables

2.1.	Data used in this chapter.	5
2.2.	Classifiers used in chapter and their hyperparameter settings	10
B.1.	The GP parameters and training parameters used in our experiments.	36

1. Introduction

Seizures are ephemeral states of irregular brain function. They are associated with temporary loss of consciousness and physical motor control. It is accepted by the epilepsy community that predicting future seizures a few minutes or more before occurring is a worthy goal [1, 2].

This thesis lays forth an exploration of existing approaches with new, noninvasive EEG datasets, as well as the development of a probabilistic computational model for unsupervised, or weakly supervised, seizure detection and prediction.

1.1. Background

Here we review common literature related to epilepsy prediction from EEG data, as well as probabilistic modeling, machine learning and Bayesian inference.

1.1.1. Epileptic seizures

Epilepsy is broadly defined by recurring unprovoked seizures. Epilepsy is characterized by hyper-synchronous activity spreading in the brain. In some individuals, this pathology can vary with time; seizures can form in a nonepileptic brain, show evolving dynamics, and even be resolved, allowing some patients seizure-freedom [3].

A complete understanding of epilepsy requires accounting for the high variability in symptoms across patients, as depicted by the many classes of epilepsy reported in the literature. The taxonomy of epilepsy involves groups of seizure types, epilepsy types, behavioural symptoms and recognizable etiologies (see Figure I. in [4]).

1.1.2. EEGs: electric potential recordings

Electroencephalography (EEG) is a primary modality in the study of epilepsy (see figure 1.1). By capturing smoothed aggregations of local-field-potentials across cortical neural populations, the EEG is effective in capturing real-time neural firing dynamics. In particular, multiple channels measuring simultaneous activity across the scalp allow complex analyses of nonlinear spatiotemporal correlation from different brain regions.

EEGs may be intracranial (iEEG) or placed on the scalp's surface (sEEG). iEEG has a higher signal-to-noise ratio, but sEEG has the advantage of noninvasiveness. In any case, the resulting data has the form of a high-frequency multivariable time-series.

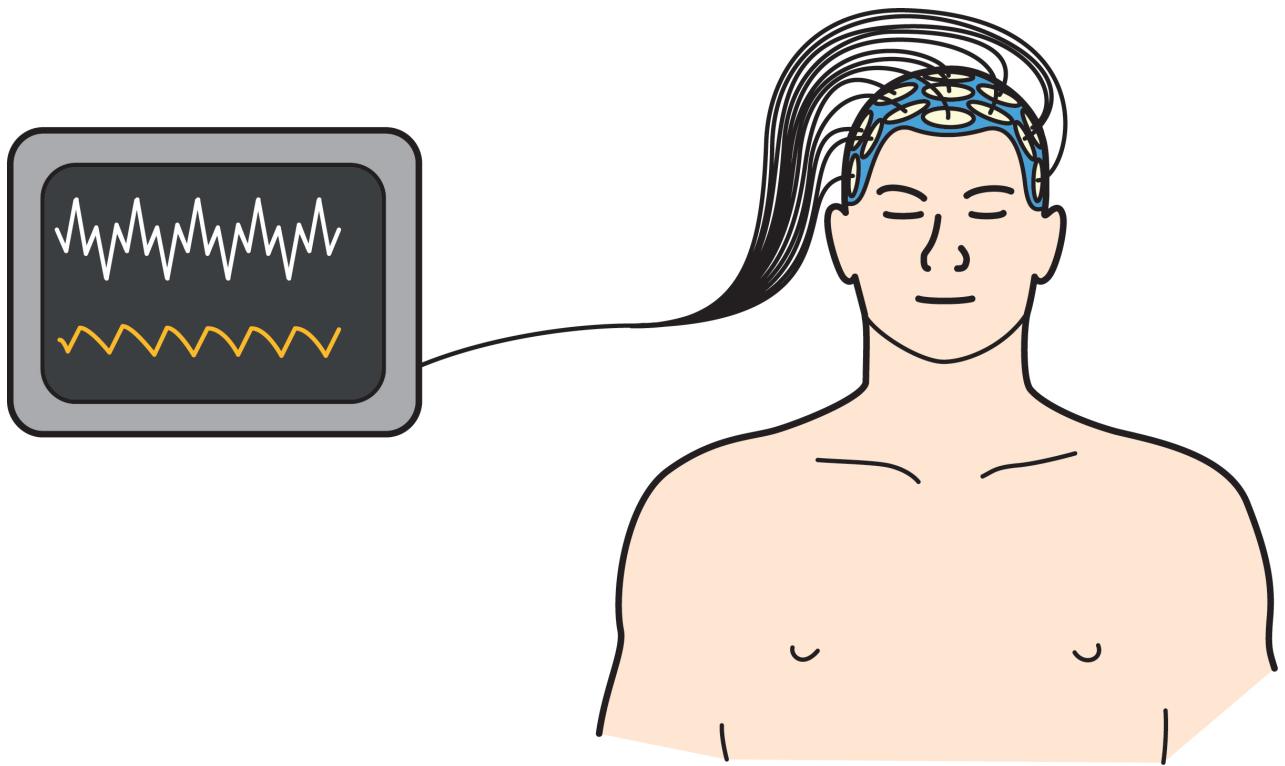


Figure 1.1. | Electroencephalography (EEG) art.

eeg from The Clear Communication People, licensed under CC BY-NC-ND 2.0.

EEG is a form of neuroimaging with high temporal accuracy. In the noninvasive setup, a wearable cap holds electrodes in contact with the scalp. The electric potentials induced by the brain are transmitted to a digital recording system for data analysis.

For seizure monitoring and epilepsy diagnosis, video-electroencephalograms (vEEG) are the standard systems used in clinical settings. The vEEG records simultaneously videography and EEG from a particular subject, so that clinicians can observe both visual and electrophysiological epilepsy-related phenomena. EEG is also used for patient evaluation during outpatients visits, and in the ICU these sessions typically do not include video monitoring. In such cases, documentation of seizures is a challenging task which can be alleviated with seizure detection algorithms [5].

1.1.3. Detecting and predicting seizures

Computational seizure detection and forecasting with EEG data continues to be an active field of research since the 1970's [6, 7]. Nonlinear methods were used in the 90's for analyzing cyclical, clustering and other time dependencies of seizure occurrences [8]. In the first decade of the millennium, models of EEG signals were developed to harness the dynamical processes involved in seizure generation for seizure forecasting [9]. Pattern recognition, including deep learning classifiers, have been shown to provide above chance classification in more than half of reported subjects [10, 11]. The last two decades have seen unprecedented efforts from physicians, computer scientists and engineers to develop detective and predictive systems for

epilepsy seizures [12–14].

1.1.4. Probabilistic modeling of seizure risk in epilepsy

The global clustering and cyclical properties of epilepsy seizures has led researchers to use probabilistic models for seizure risk estimation. Craley et al. [15] implement a hybrid deep-learning and probabilistic-graphical-model approach and show that it outperforms baseline methods at seizure detection. A logistic regression model, tweaked with Bayes' rule to incorporate a subject-specific prior, was shown to improve forecasting in humans [16]. Stochastic point process models have shown promising results for integrating prior information with real-time likelihood assessments to produce forecasts of seizures hours to days in advance, although prospective clinical studies are still required to validate the applicability of these algorithms [17].

1.1.5. Supervised machine learning and the data labeling problem

Prior work on automated seizure detection and prediction has focused on training machine learning models to classify EEG segments labeled as ictal, preictal or interictal. This requires expert labeling of EEG datasets for seizure events, which is both time consuming and requires extensive training. The dependence of supervised machine learning classifiers on expert-level domain knowledge makes these algorithms unsuitable for real-world forecasting devices, since expert categorical labeling solutions are still not available at scale [18]. Furthermore, even clinical-grade annotations, considered to be the "gold standard" in seizure documentation, suffer from judgement bias [19].

2. Pattern Recognition for Seizure Prediction

The field of pattern recognition includes designing pipelines to learn from data, represented as feature vectors, and recognize patterns which generalize across past and future data. A considerable amount of research has revealed that, indeed, machine learning classifiers are able to perform above-chance classification between preictal and interictal feature vectors in many patients.

Previously, the success of these methods for seizure prediction based on intracranial EEG has been demonstrated [20, 21] in human patients as well as animals. In this chapter, we show that similar results can be achieved by the same methods applied to scalp-surface EEG.

2.1. Methods

We follow closely the exploratory methods demonstrated in Mirowski et al. [20]. They show applications of several classifiers to seizure prediction. Bivariate feature extraction methods are used to create 2-d patterns for classification of preictal and interictal patterns (cf. figure 2.4c).

Data collection and specification

Typical publicly available epilepsy-seizure-prediction datasets consists of long-term raw EEG recordings from pre-surgical evaluation periods, supplemented with annotations provided by approved experts [22]. In these datasets, the occurrence of seizures is reported as a list of ictal (seizure) intervals

$$I := \{Ictal_i\} = \{\langle t_{start}^i | S_i | t_{end}^i \rangle\} \quad (2.1)$$

where t_{start} is the seizure onset, t_{end} the seizure offset, and S_i , when available, are the additional details reported such as epilepsy etiology, seizure type classification. In this project, we used data from the Epilepsiae Dataset [23]. Table 2.1 lists raw recording length, sample frequency, number of surface electrodes and size-on-disk per patient.

2.1.1. Preprocessing

Although it is common to apply spectral filters to EEG data to reduce line noise etc., we chose to focus on the effect of classifier and features choice on predictive performance, and thus kept preprocessing to a

Table 2.1. Data used in this chapter.

Patient Name	Length	Frequency	Electrodes	SOD
pat_3500	92.8 h	1024 Hz	32	21 GB
pat_3700	79.8 h	512 Hz	32	8.8 GB
pat_7200	94.6 h	1024 Hz	29	21 GB

minimum. First, EEG channel selection was performed to select a spatially far-reaching distributed subset of 19 channels, common to all patients (see figure 2.1).

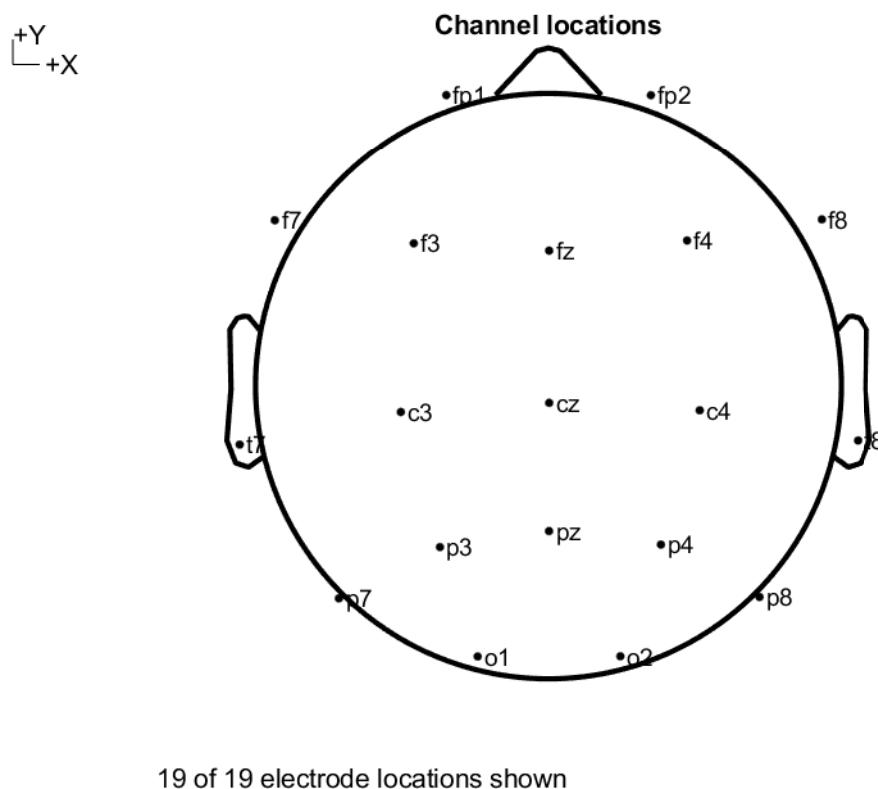


Figure 2.1. | EEG channel locations used in this chapter. The common set of channels ['Fp1', 'Fp2', 'F3', 'F4', 'C3', 'C4', 'P3', 'P4', 'O1', 'O2', 'F7', 'F8', 'Fz', 'Cz', 'Pz', 'T7', 'T8', 'P7', 'P8'] was selected for all patients. .

This was done to reduce data dimension as well as standardization amongst patients. The raw data was resampled to 256 Hz to reduce disk space and processing time.

2.1.2. Preictal and Interictal Binarization

A practical way to characterize time-varying signal dynamics is through a sliding window approach (see Fig. 1. in [24]). An EEG time series matrix X is partitioned temporally into non-overlapping windows

x_1, \dots, x_N . Each window x_i is given a label y_i , to form a dataset $D = \{(x_1, y_1), \dots, (x_N, y_N)\}$. Optionally, each window is transformed via a feature extraction function $f(x)$, which yields a feature-formed dataset $D_f = \{(f(x_1), y_1), \dots, (f(x_N), y_N)\}$.

Definition 2.1 (Preictal state). The state of preceding a seizure by a short time (minutes to hours).

Definition 2.2 (Interictal state). The state of being at least a few hours before and after a seizure.

Hence, each EEG window x_i is checked for distance to the provided seizure labels (see eq. 2.1) to create the appropriate label:

$$y_i := \begin{cases} 0 & \text{if } \text{preictal}(x_i) \\ 1 & \text{if } \text{interictal}(x_i) \end{cases} \quad (2.2)$$

Formally, $\text{ictal}(x_i)$ if and only if x_i temporally overlaps some ictal interval I_j . Similarly, $\text{preictal}(x_i)$ if and only if x_i overlaps with some τ_p time interval preceding an ictal interval I_j . Finally, $\text{interictal}(x_i)$ if and only if x_i does not closely precede or proceed any seizure interval, with thresholds τ_a before and τ_b after, respectively. In this chapter, τ_a and τ_b are each taken to be 4 hours, and τ_p is taken to be 1 hour, in line with previous works and competitions such as Mirowski et al. [20], UPenn and Mayo Clinic and Kaggle [21].

2.1.3. Bivariate Feature Extraction Methods

The deterministic classification method relies on hand-crafted, manually engineered, features. Specifically, following [20], we focus on bivariate measures of synchronicity between pairs of EEG channels (see figure 2.2 for examples).

For each patient, the recording is segmented into 5 minute windows. At loading time the EEG was shifted and scaled to zero mean and unit standard deviation, per 5-minute window. Each window is segmented into 60 frames, each 5 seconds long. Each 5 second frame is reduced into a vector of length $c \cdot (c - 1)/2$ (where c is number of channels), via one of the feature extraction methods described below. Each 5 minute window is regarded a single pattern with a single label (preictal vs. interictal).

Maximal Linear Cross Correlation (figure 2.3a)

In order to quantify the similarity of two signals $\{x_i\}$ and $\{y_i\}$ the maximum of a normalized cross-correlation function is taken as a measure for lag synchronization [25]:

$$C_{max} = \max_{\tau \in [-N, N]} \left| \frac{C_{xy}(\tau)}{\sqrt{C_{xx}(0) \cdot C_{yy}(0)}} \right| \quad (2.3)$$

Where

$$C_{xy} = \begin{cases} \frac{1}{N-\tau} \sum_{i=1}^{N-\tau} x_{i+\tau} y_i, & \text{for } \tau \geq 0 \\ C_{yx}(-\tau), & \text{for } \tau < 0 \end{cases} \quad (2.4)$$

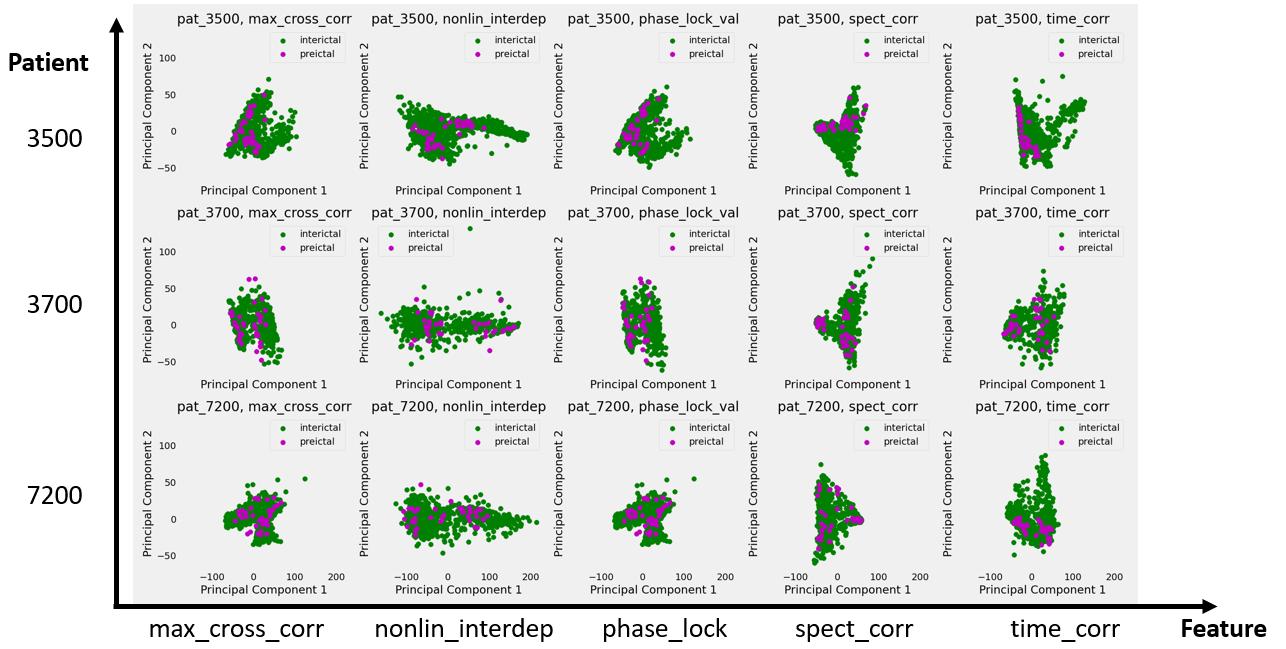


Figure 2.2. | First 2 principal components of transformed data per patient and feature type. The feature types are (left to right): maximal linear cross correlation, nonlinear interdependence, phase locking value, spectral domain correlation, time domain correlation. Each point represents a flattened 5-minute window (i.e., a pattern), and it's color denotes class label (interictal or preictal).

is the linear cross-correlation function. C_{max} is confined to the interval $[0, 1]$ with high values indicating that the two signals have a similar course in time (though possibly shifted by a time lag τ) while dissimilar signals will result in values close to zero.

Phase Locking Value (figure 2.3b)

Introduced in [26], the *Phase Locking Value* measures synchronicity between EEG channels in different locations on the scalp. First, for each channel i , the instantaneous phase $\sigma_i^a(t)$ of the analytical signal $x_i^a(t)$ is extracted. Then, for each pair (i, j) of channels, we compute the modulus of the time averaged phase difference mapped onto the unit circle:

$$PLV_{ij} = \left| \frac{1}{T} \sum_t e^{i(\phi_i^a(t) - \phi_j^a(t))} \right| \quad (2.5)$$

Correlation Coefficients and Eigenvalues in the Time and Frequency Domains (figure 2.3)

We compute the correlation coefficients between each pair of EEG channels, along with the eigenvalues of the correlation coefficients matrix, in both the time and frequency domains. This provides two more sets of

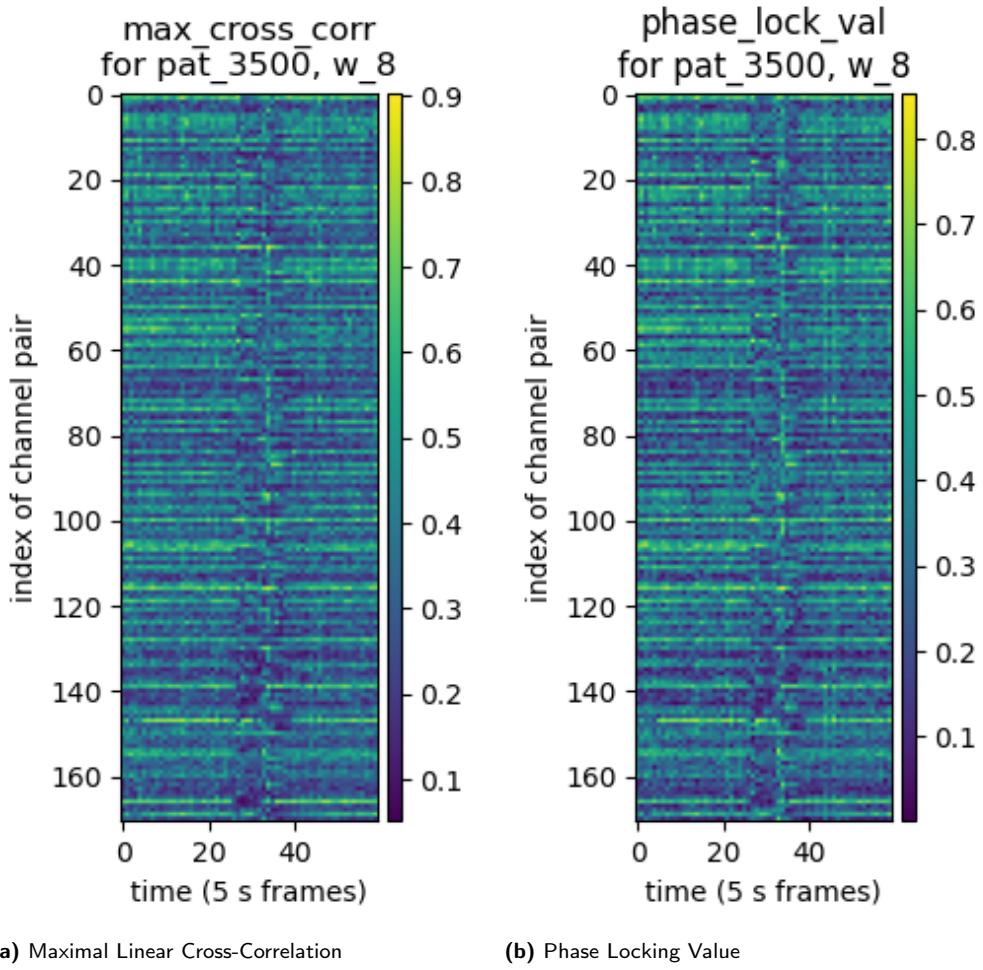


Figure 2.3. | Two types of bivariate features. Here are shown example patterns created by transforming 19-channel EEG recordings of length 5 minutes into maximal linear cross correlation (a) and phase locking values (b). Each row represents the feature values for a pair of channels, and each column represents those values for a particular 5 second frame. The frames are concatenated to form the window.

measures of synchronicity across EEG channels.

Nonlinear Interdependence (figure 2.4c)

The *non-linear interdependence* measure for generalized synchronization between two EEG single-channel signals $\{x_i\}$ and $\{y_i\}$ is presented in [6]. First, the two signals are represented as trajectories in a state space, via time-delay embedding. Then, an asymmetric statistic measuring the Euclidean distance, in the reconstructed state-space, between trajectories $\{\vec{x}_i\}$ and $\{\vec{y}_i\}$ is calculated for each time point. This is done for each pair of channels. See [10] for more details.

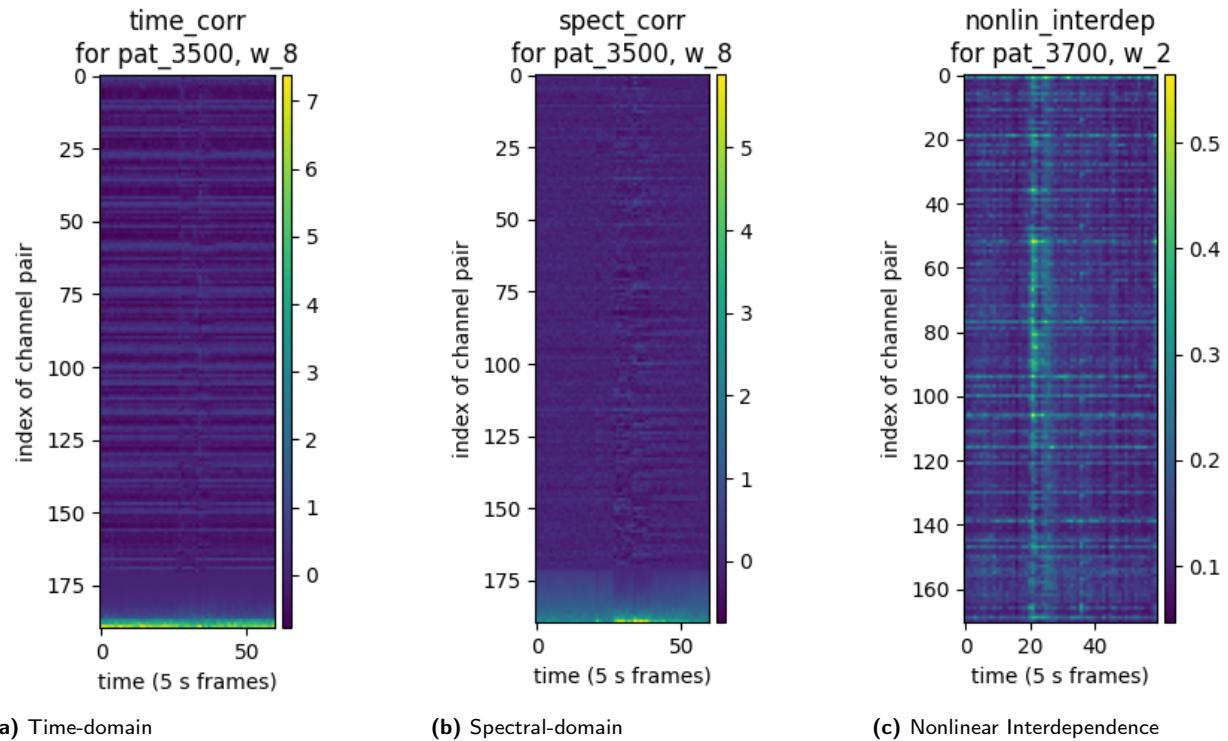


Figure 2.4. | Three types of bivariate features. Example patterns created by transforming 19-channel EEG recordings of length 5 minutes into inter-channel correlation coefficients with eigenvalues appended at the bottom (a and b), and nonlinear interdependence (c).

2.1.4. Classifiers

We trained and tested 9 of classifiers implemented in the *scikit-learn* [27] ML toolkit (v1.0.1), listed in table 2.2. We chose a variety of classifiers from different families (i.e., neural networks, ensembles and decision trees).

2.2. Results

A total of 117 classifier-datasets pairs were trained and evaluated with 5-fold cross validation, yielding a total of 585 model fits. In each fold, the classifier was trained on a random subset of 80% of the data, and tested on the remaining 20%.

2.2.1. Evaluation Metrics

We report the following metrics for all features-classifiers pairs:

1. ROC AUC
2. Precision

Table 2.2. | Classifiers used in chapter and their hyperparameter settings. The settings chosen were the default settings provided by sci-kit learn v.1.0.1. They are repeated here for reference.

Classifier Name	Hyperparameters
Nearest Neighbors	K=3
Linear SVM	kernel="linear", C=0.025
RBF SVM	gamma=2, C=1
Decision Tree	max_depth=5
Random Forest	max_depth=5, n_estimators=10
Neural Net	alpha=1, max_iter=1000
AdaBoost	n_estimators=50, learning_rate=1.0, algorithm='SAMME.R'
Naive Bayes	None
QDA	reg_param=0.0

3. Recall
4. fit time
5. score time

2.2.2. Classifier and Feature Comparison

Figures 2.5 and 2.6 show the mean and standard-deviation for each classifier on each feature set, for each patient. It is found that the Linear SVM and Neural Net (MLPClassifier) are the top two performers consistently, sometimes followed closely by the Nearest Neighbor Classifier.

2.2.3. Training Efficiency

Figure 2.7 shows the mean time it took to fit each classifier to each feature-set, for each patient, over 5-fold CV. It is shown that the neural network, ensemble method (AdaBoost) and support-vector-machines consistently take longer than the decision tree, naive-Bayes, quadratic discriminant analysis, and random forest classifiers.

2.2.4. Comparing Feature Sets

The Linear SVM, which performed best in the previous experiments, is selected to compared the predictive performance of the different feature types. Notably, the classifier performed above chance in all feature sets, with the spectral correlation coefficients features slightly outperforming the others with respect to the ROC AUC score. See figures 2.8a to 2.8c.

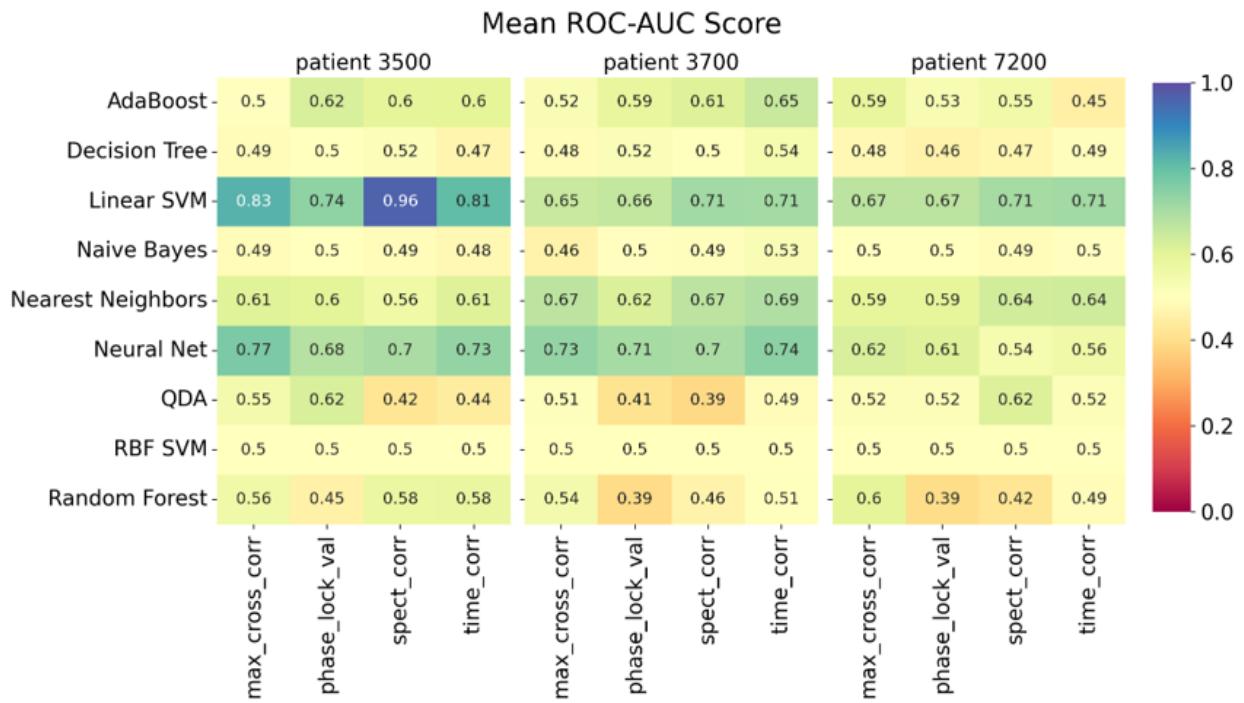


Figure 2.5. | Mean AUC-ROC scores for 5-fold cross validation..

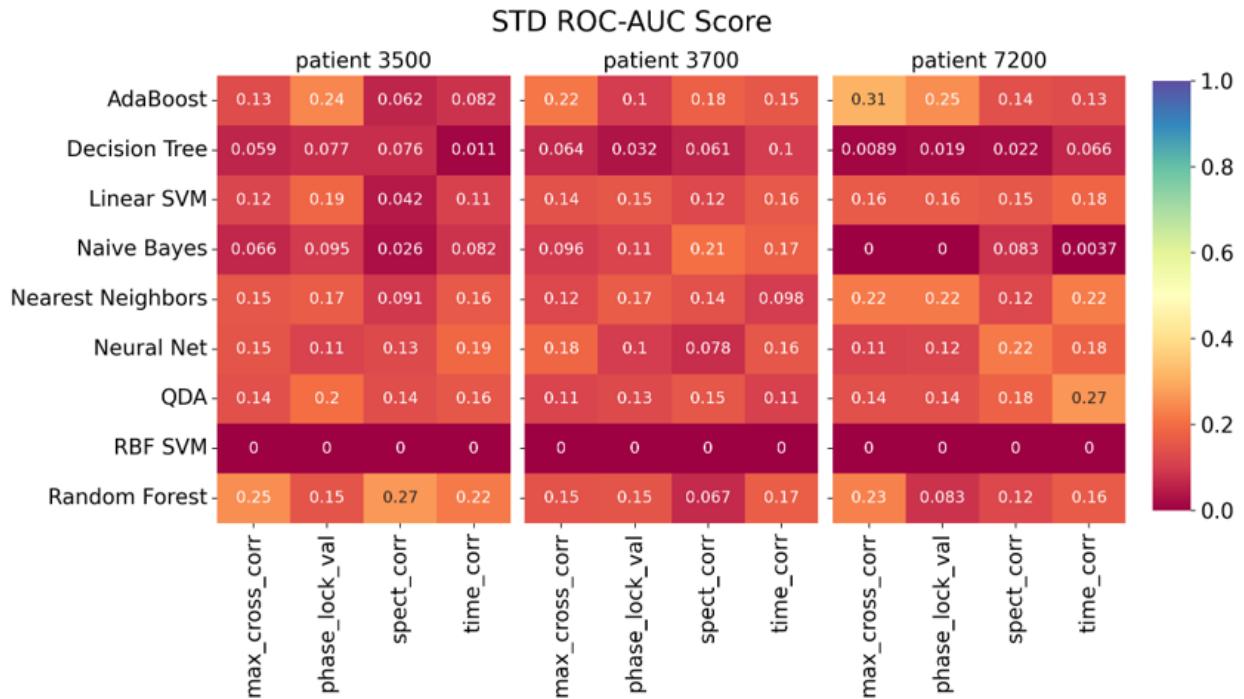


Figure 2.6. | Standard deviation of AUC-ROC scores for 5-fold cross validation..

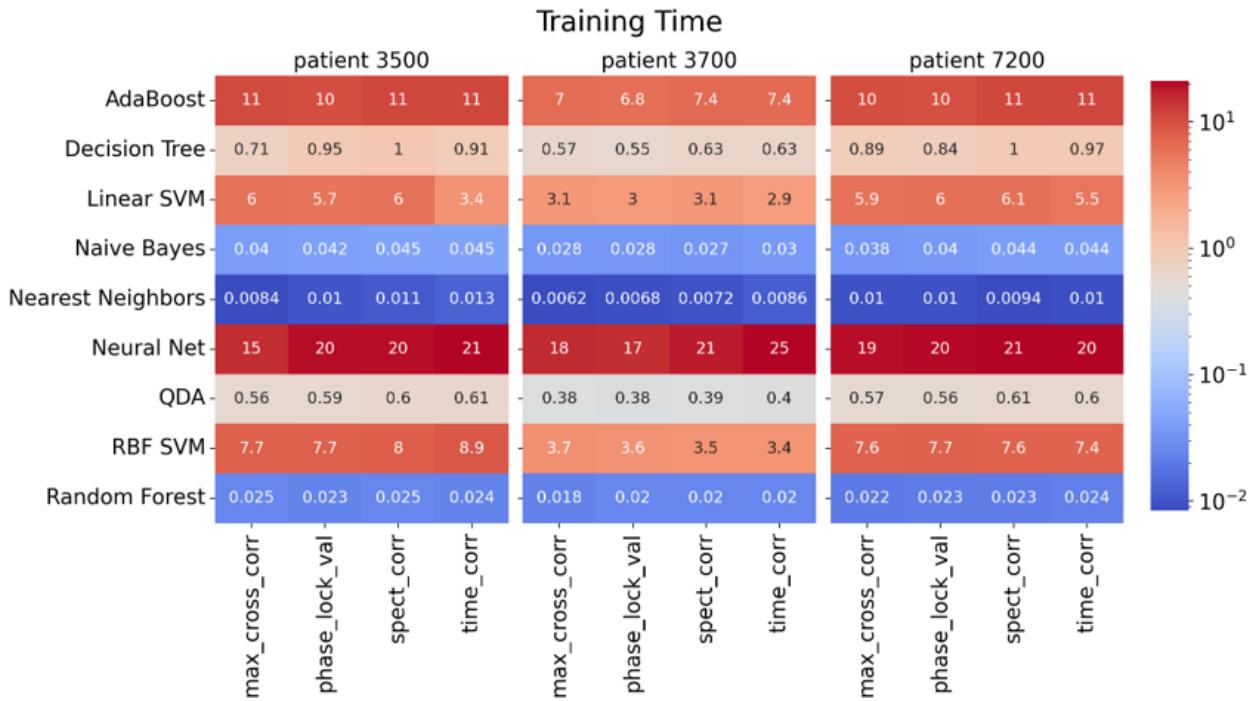


Figure 2.7. | Mean training times (in seconds) for fitting the classifiers to the datasets..

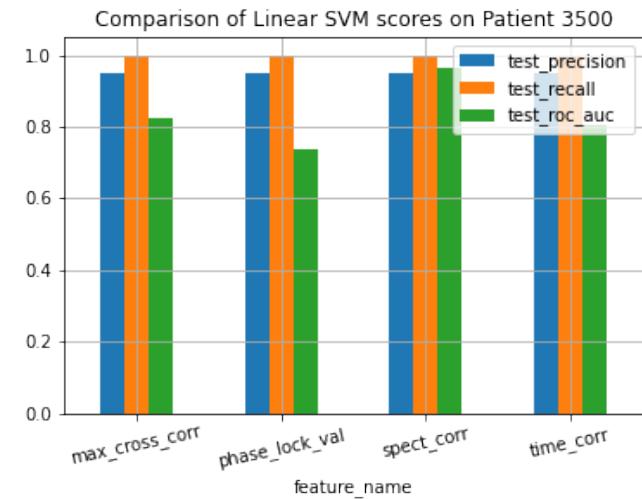
2.3. Conclusion & Discussion

In this chapter, we explore the pattern recognition approach in a similar fashion to Mirowski et al. [20], albeit on a newer dataset of noninvasive, 19 channel surface-EEG. We achieve comparable results across 3 subjects, with ROC-AUC scores commonly between 0.7-0.8 for a variety of bivariate feature sets related to inter-channel synchronicity with support vector machines.

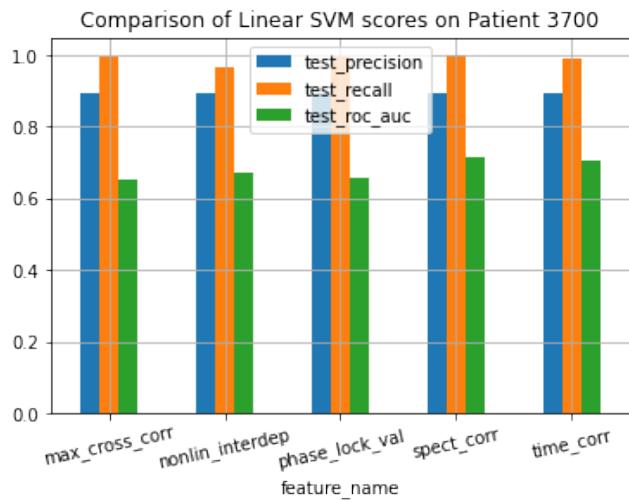
Mirowski et al. [20] evaluated their pipeline on 6-channel, intracranial data from 21 epilepsy patients. In this chapter, we use 19-channel data from surface EEG, from 3 patients. They reported 71% sensitivity with 0 false alarms, on 15 out of the 21 patients they assessed. We trained 117 classifiers in different combinations of classifier-dataset. On the best classifier for each patient, only Patient 3700 achieves over 70% sensitivity at the 0 false alarms threshold. For Patients 3500 and 7200, the sensitivities at 0 false alarms are 42% and 40%, respectively.

The main contribution of this chapter is to show that the classification of synchronicity-based features for seizure prediction is possible not just with intracranial data as in [20], but also with surface-EEG from Ihle et al. [23]. Another difference between our data and the data analyzed in the previous work is the number of EEG-channels. Although the increase from 6 (15 pairs) to 19 (171 pairs) EEG channels gives us more sources of information, it also leads to a larger sample space such that probably much more data is needed.

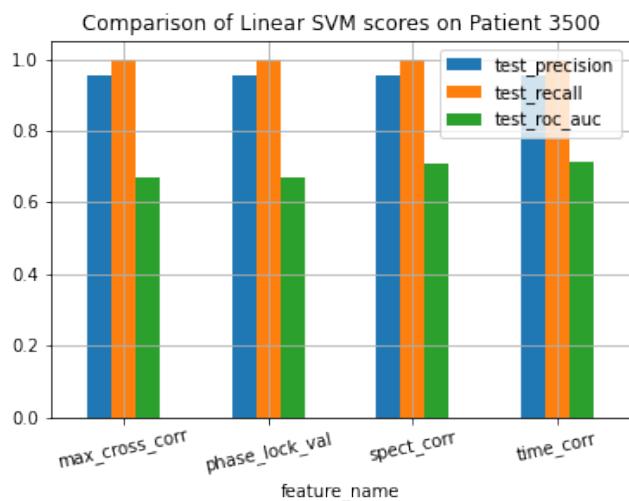
This chapter showcases the predictive power of different classifiers and synchronicity-based features applied to noninvasive EEG data. We trained various classifiers. While some classifiers failed to generalize at



(a) patient 3500



(b) patient 3700



(c) patient 7200

Figure 2.8. | Results of Linear SVM on different feature datasets.

all, most classifiers were successful in classifying better than chance, with AUC-ROCs in the range 0.67-0.81 commonly achieved. When comparing the synchronicity features, we did not find any of the feature sets to be remarkably better than the others in any of the 3 patients.

3. Bayesian Estimation of Seizure Likelihood

The recently reported paradigm shift from binary classifications to probabilistic forecasts of seizure risk has been illustrated conceptually in recent works such as Karoly et al. [16], Baud et al. [28, 29]. They discuss the need to consider *pro-ictal* states - states to which we attribute a higher seizure likelihood, that does not necessarily result in a seizure. The likelihood is based on a priori knowledge or assumptions.

In our work, we propose a Bayesian inference based algorithm, namely Bayesian Seizure Likelihood Estimation (BSLE)¹, and evaluate it on intracranial EEG from canines with epilepsy. It is different than the previously mentioned works by estimating the likelihood in an unsupervised way, dismissing the need for accurate seizure event labels.

3.1. Background

The purpose of this section is to introduce preliminary concepts necessary to understand the probabilistic approach.

3.1.1. Problem description

There is a continuous EEG recording system sampling at f Hz (see figure 3.1). We are provided with a dataset $D = \{e_t, a_t\}_{t_0}^{t_m}$. For each time t , $e_t \in \mathbb{R}^{c \times f \cdot T}$ is the observed EEG segment with c -channels of duration T , ending at time t . Also, $a_t \in \{0, 1\}$ is a clinically-approved seizure annotation corresponding to the observed interval $[t - T, t]$. The annotation denotes the expert's best judgement as to whether a seizure event began within the corresponding time-window.

The problem at hand is to find a forecaster f_τ^* with predictive qualities of interest. First, a general definition of seizure forecasting is provided. This setup is made more specific in the evaluation section (3.3).

Definition 3.1 (seizure forecaster with horizon τ). a function of the form

$$\hat{s}_{t+\tau} = f_\tau(e_t, a_{0:t})$$

where $s_{t+\tau}$ is the seizure risk in the interval $[t, t + \tau]$, e_t is the current EEG observation, $a_{0:t}$ is the past event history, and τ is the forecast horizon. The horizon parameter τ can take negative values, to mean detecting seizure events post-observation.

¹code can be found at <https://github.com/noamsgl/msc>.

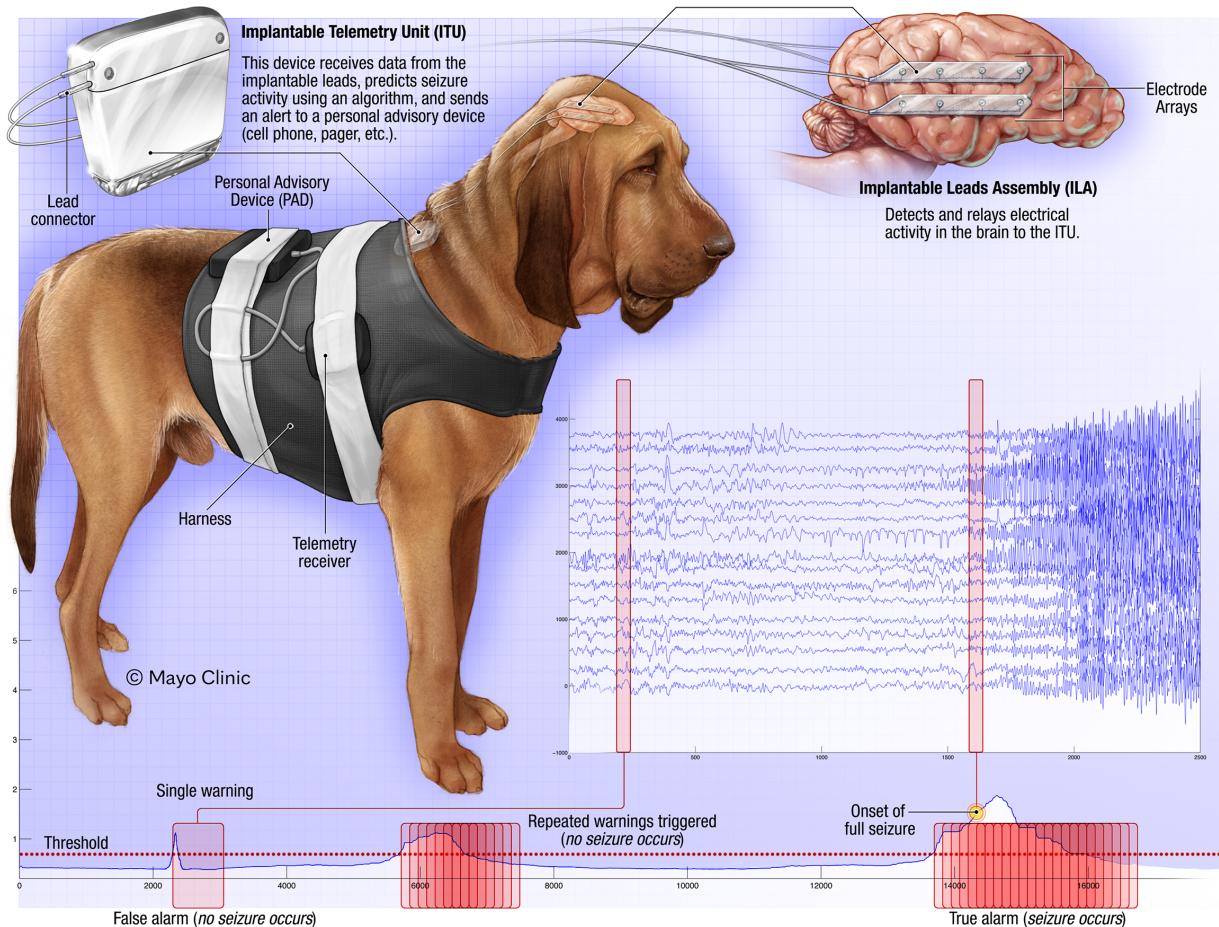


Figure 3.1. | Data collection and required output. From Coles et al. [30]. Used with permission of Mayo Foundation for Medical Education and Research, all rights reserved.

The Canine-Epilepsy dataset was used for empirical evaluation in this chapter, chosen primarily because it is publicly available and the per-subject recording length is long. The EEG is collected and processed by a forecasting algorithm to produce a seizure risk gauge (bottom).

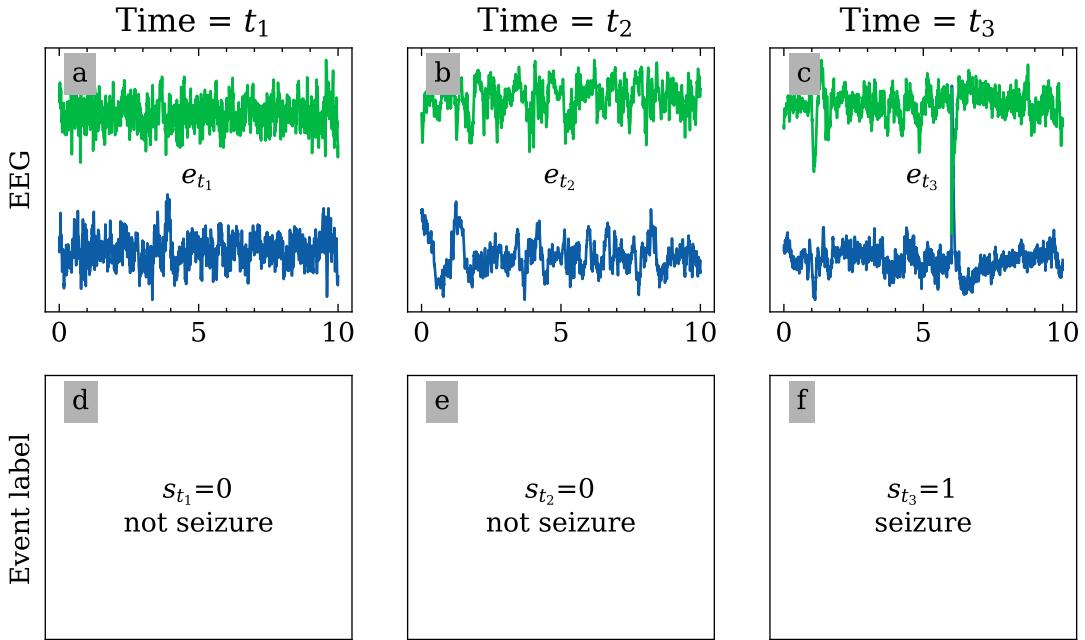
3.1.2. Probabilistic representation of EEG and seizures in time

We capture the notion of EEG observations and seizure events by random variables (r.v.s):

$$e_t \sim E(t) \in \Omega_E = \mathbb{R}^{c \times N} \quad (3.1)$$

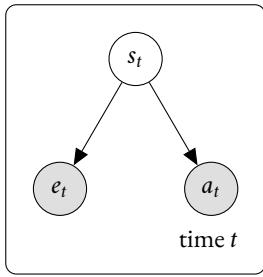
$$s_t \sim S(t) \in \Omega_S = \{0, 1\} \quad (3.2)$$

where Ω_E (Ω_S) is the sample space for the r.v. E (the r.v. S). c is the number of EEG-channels, N is the number of samples recorded (a.k.a. segment length) over a duration T . We define the event that a seizure began within $[t - T, t]$ as $\{S(t) = 1\}$ (see figure 3.2).



3.2. Model specification

Formally, the proposed model is a multilevel probabilistic model, as shown in figure 3.3. To overcome the curse of dimensionality, first the raw time signal is embedded in a low dimensional manifold. Then, Bayesian inference is applied with custom evidence, prior and likelihood functions to infer the probability $\mathbb{P}(s_t | e_t)$.



$e_t | s_t \sim \mathcal{D}_{E|S}$ // EEG embeddings conditioned on seizure variable
 $a_t | s_t \sim \mathcal{D}_{A|S}$ // Annotation conditioned on seizure variable

Figure 3.3. | Probabilistically Modeling Seizure Occurrence, EEG signal and Annotations across Time. The nodes represent random variables. The shaded nodes e and a represent the observed variables EEG and Annotations, respectively. The s node represents a latent variable, whose values can be inferred with Bayes' theorem. The

3.2.1. EEG embedding

A Gaussian process (GP) is a mathematical entity often used in Bayesian time series modeling. It allows flexible modeling of smoothness, periodicity and other interpretable signal properties. Existing algorithms which perform posterior inference over the underlying hyperparameters are available [31, 32]. We embed the EEG data in the low-dimensional manifold of GP hyperparameters (see figure 3.4). These hyperparameter values are then used to represent the EEG in the following procedures. Technical details of the embedding process are provided in appendix A.1.

3.2.2. Bayesian estimation

We wish to construct a model for $\mathbb{P}[S, E, t]$, and then apply it with Bayes' rule to infer the likelihood of a seizure at time t determined by an EEG recording:

$$\text{probability[seizure=S | time=t, EEG=E]} \propto \mathbb{P}[E | S, t] \mathbb{P}[S | t] \quad (3.3)$$

It should be evident that this procedure is general in that each component on the right-hand-side can be estimated independently, and then combined via multiplication.

In the case of epilepsy, handling uncertainty in the face of evidence plays a major role, thus naturally appealing to the mathematical machinery termed Bayes' theorem.

We apply Bayes' theorem to estimate the updated likelihood of a seizure after observing an EEG signal:

$$\mathbb{P}(S | E) = \frac{\underset{\text{likelihood}}{\mathbb{P}(E | S)} \underset{\text{prior}}{\mathbb{P}(S)}}{\underset{\text{evidence}}{\mathbb{P}(E)}} \quad (3.4)$$

For bayesian inference, the likelihood, prior and evidence functions need to be defined and approximated. We now derive the mathematical formulations for each of these, which should complete the description of

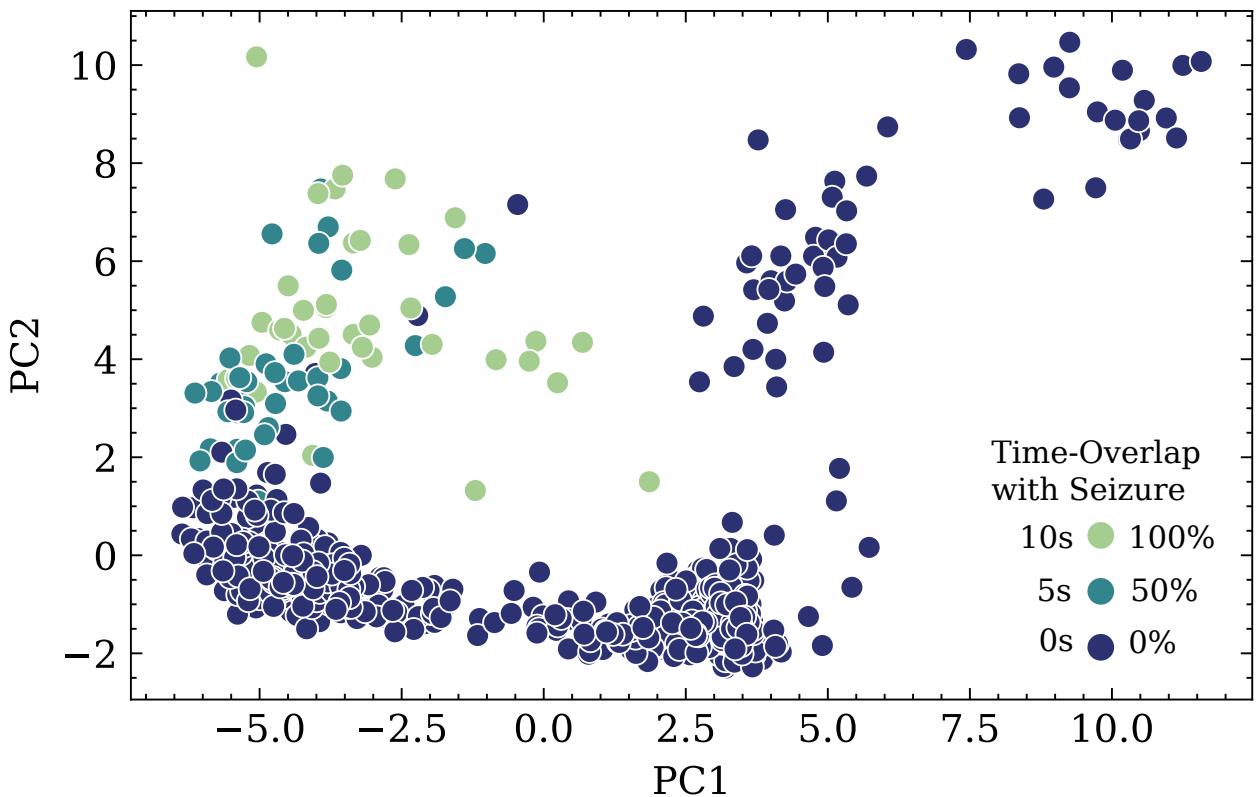


Figure 3.4. | EEG embedding in Gaussian process hyperparameter space. Each point represents a 10-second EEG segment fit with a Gaussian process, and displays the first two principal components of the hyperparameters vector. The color indicates the time of the recording relative to the seizure annotations. The ictal embeddings are found to be distinguishable from the interictal embeddings in the space of GP-hyperparameters.

our inference procedure.

Likelihood Definition

For a patient with epilepsy, we hypothesize that the anomalies in EEG data are inherently more likely to reflect underlying seizure events, because seizures are rare events. First, we define an unsupervised density-quantile novelty score. Then, we establish the seizure likelihood function by adding a threshold parameter.

Density estimation A Gaussian mixture model (GMM) with 4 components and full inter-component covariance matrix is fit to the data in GP-hyperparameter space, to give a density estimation:

Definition 3.2. Density estimation $\hat{p}(e) \approx pdf(E)$

Mathematical details are provided in appendix A.2. See figure 3.5.

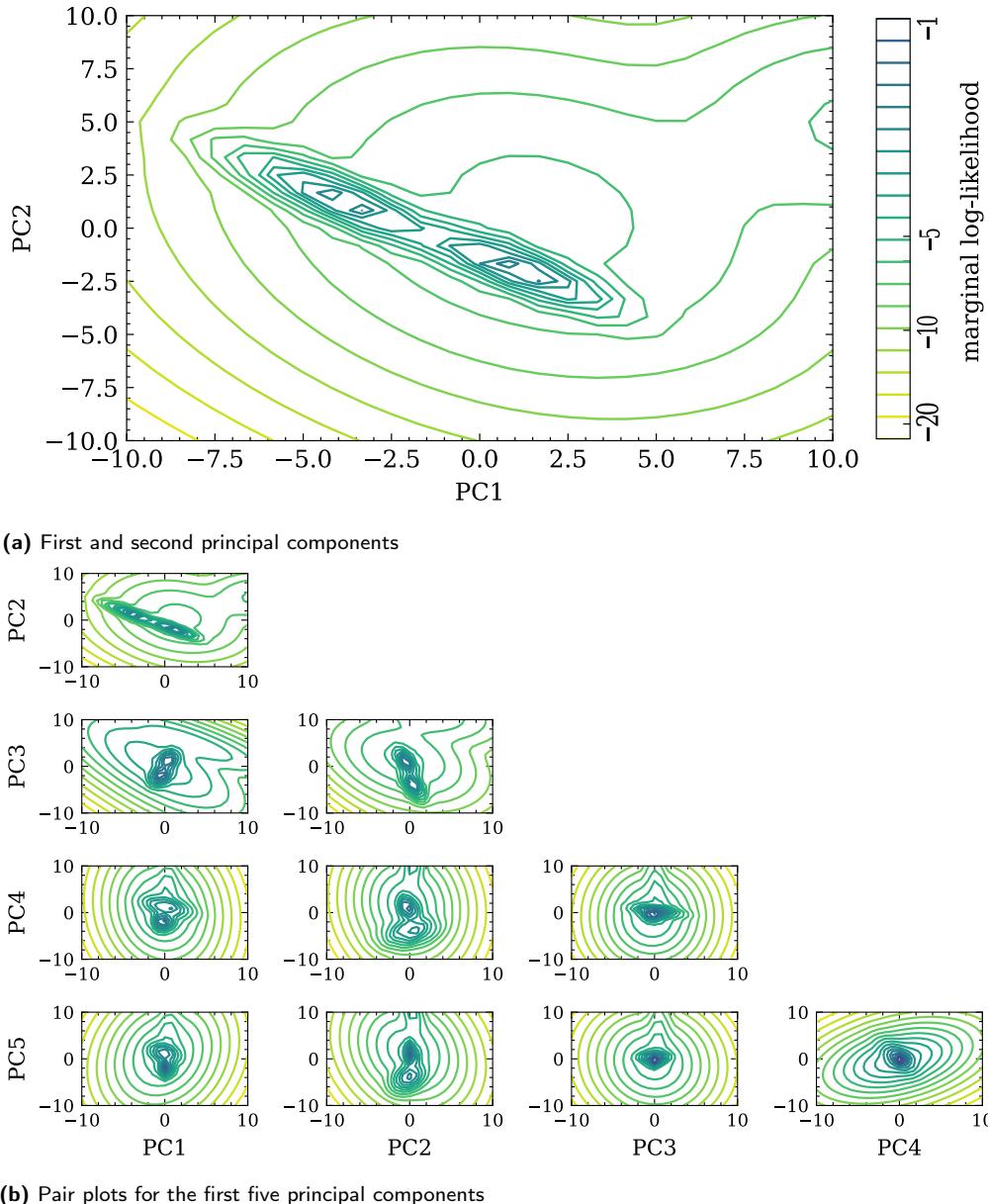


Figure 3.5. | Contour plots for multidimensional density estimation. An estimator $\hat{p}(e_t)$ is computed using a GMM with 4 components fit to the first 5 principal components of the data as represented by GP hyperparameters. The contour plot of $\log \hat{p}(e_t)$ is shown for every pair of principal components by marginalizing over the remaining axes. The regions with lower likelihood are more likely to be classified as seizures.

Density quantile approach to Novelty Scores A Monte Carlo technique, namely the Density Quantile Approach, is used to obtain information about the probability coverage of a given region of the sample space. Given a new sample e , the novelty score is the quantile (also known as a p_{value}) of $\hat{p}(e)$ relative to the random variable E , as approximated by a sample dataset D . This calculation has been shown to converge tractably to the probability density function quantile as the sample size grows [33].

Definition 3.3. Novelty Score

$$\mathcal{Z}(e^{new}) = \frac{|\{e_i \mid \hat{p}(e^{new}) \leq \hat{p}(e_i)\}|}{|\{e_i\}|} \quad (3.5)$$

The essence of the novelty score $\mathcal{Z}(e^{new})$ is that it maps data samples to the interval $[0, 1]$, such that more anomalous samples (i.e. samples from less dense regions) are given larger values.

In accordance with our rarity of seizures hypothesis, a new normal (nonseizure) EEG observation e^{new} will be relatively similar to previously seen observations in $\{e_i\}$, leading to a low novelty score. Conversely, for abnormal (seizure) data, the data will be relatively dissimilar, leading to a high novelty score.

In order to utilize this hypothesis in the BSLE likelihood function, we flip the sign of the S parameter to $\neg S$ meaning not a seizure, thus modeling $\mathbb{P}(E \mid \neg S)$ directly.

Partitioning with a threshold α A new threshold parameter α is introduced to guide the BSLE likelihood function $P(E \mid \neg S; \alpha)$. Using the density estimator $\hat{p}(e)$, we partition the GP-hyperparameter sample space Ω_{GP} into a low density region and a high-density region, and treat each region differently. The threshold parameter α affects the location of the region split.

Definition 3.4 (α -Highest Density Region). Let $f(x)$ be the density function of a random variable X . Then the $100(1 - \alpha)\%$ HDR is the subset $R(f_\alpha)$ of the sample space of X such that

$$R(f_\alpha) = \{x : f(x) \geq f_\alpha\}$$

where f_α is the largest constant such that $\mathbb{P}(X \in R(f_\alpha)) \geq 1 - \alpha$.

Definition 3.5 (BSLE likelihood).

$$\mathbb{P}(e^{new} \mid \neg S; \alpha) = \begin{cases} \mathcal{Z}(e^{new}) & \text{if } e^{new} \in HDR_\alpha \\ 0 & \text{if } e^{new} \notin HDR_\alpha \end{cases} \quad (3.6)$$

Intuitively, the α parameter determines the interplay between auto-rejection of a sample and estimation of its seizure likelihood.

Evidence Definition

Definition 3.6 (BSLE Evidence).

$$\mathbb{P}(E) = \mathbb{P}(\{\text{pdf}(e) \leq \text{pdf}(E)\}) \quad (3.7)$$

Prior definition

The prior function is where the Bayesian modeler induces intentional bias into the inference process. We define and compare two priors, to show that our model is general in this respect.

Weakly supervised cyclical prior This prior encourages forecasters to inhibit a circadian - twenty four hour long - cyclical preference. We construct mathematically the cyclical prior as a linear mixture of 24 circular Gaussian distribution kernels, as in [16].

Definition 3.7 (Circular Gaussian Distribution).

$$f(x|\mu, \kappa; \omega) = \frac{\exp(\kappa \cos(\omega(x - \mu)))}{2\pi I_0(\kappa)} \quad (3.8)$$

In this work, we set $\omega \leftarrow \frac{2\pi}{24}$ to scale the period to 24-hours, and drop it from the notation for brevity in the following text.

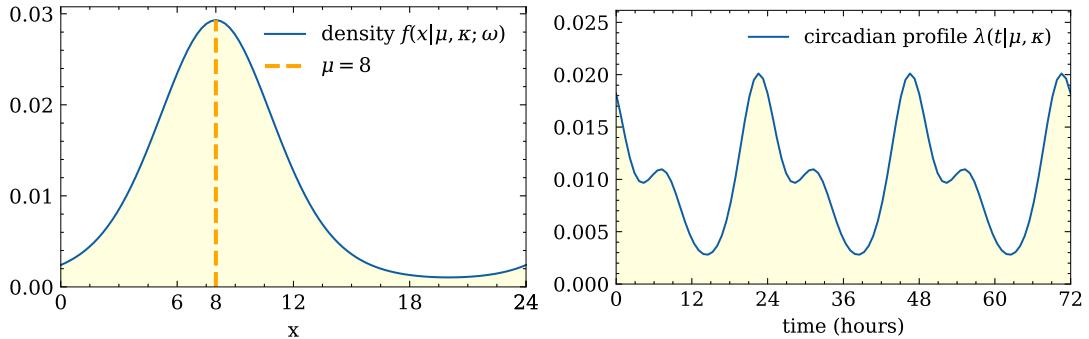


Figure 3.6. | Circular Gaussian (von Mises) distribution and circadian profile. The circular Gaussian distribution is similar to a bell-shaped normal distribution on a circle (left). A mixture of von Mises distributions represents the cyclical seizure base-rate behavior, termed *circadian profile* (right).

Definition 3.8 (Circadian prior).

$$\mathbb{P}(S_t) = \frac{1}{K} \sum_{i=0}^{23} f(t | i, k) \quad (3.9)$$

where K is a normalizing constant evaluated numerically (`np.trapz`).

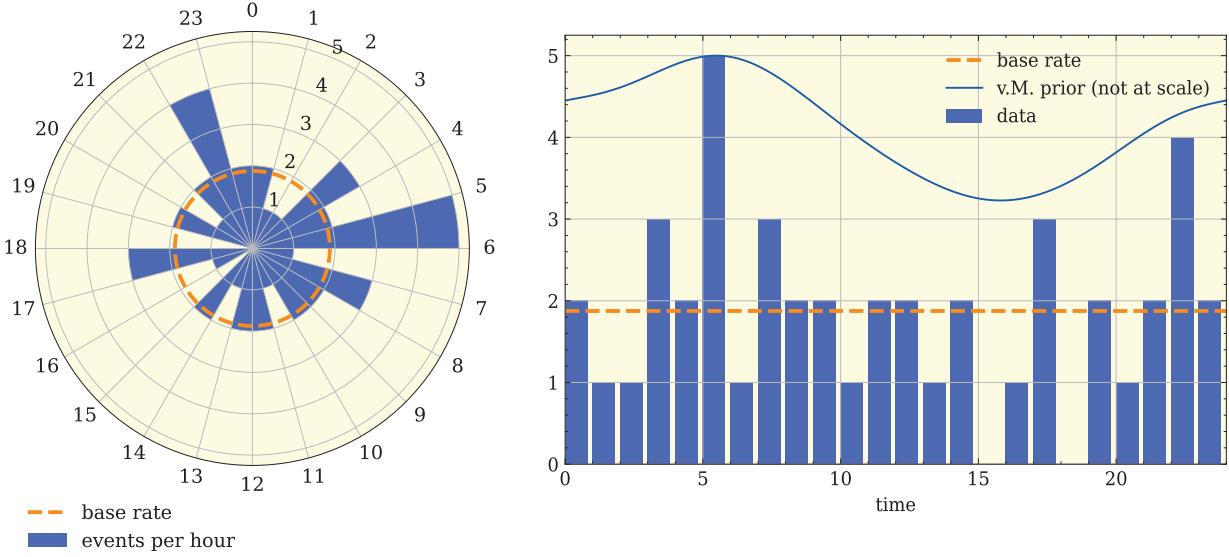


Figure 3.7. | Circadian Seizure Histogram and Circadian Profile. The 45 seizure events of canine I004_A0003_D001 from 465 days are binned by hour-of-the-day into a polar histogram (left). A von Mises mixture density function, with kernel weights given by the circadian histogram, is plotted (right).

Definition 3.9 (Uniform prior). The uniform prior takes a constant value for all time points:

$$\mathbb{P}(S_t) = \lambda_0 \cdot T \quad (3.10)$$

where λ_0 is a constant seizure rate, and T is the observation duration. Without loss of generality, we set $\lambda_0 = \frac{1}{10 \text{ day}}$, $T = 10 \text{ sec}$.

This prior issues equal seizure plausibility to all time points, nullifying the time covariate's effect on the seizure likelihood estimation.

Now, using the fact that $\mathbb{P}(S) + \mathbb{P}(\neg S) = 1$, and having observed an EEG recording E , the a posteriori estimate for the likelihood of a seizure is given by:

$$\begin{aligned} \mathbb{P}(S | E) &= 1 - \frac{\mathbb{P}(\neg S)\mathbb{P}(E | \neg S)}{\mathbb{P}(E)}. \\ &\text{probability of seizure} \\ &\text{given EEG} \end{aligned} \quad (3.11)$$

which completes the description of the BSLE algorithm.

3.3. Empirical results

The previous subsection described the calculations of the likelihood, priors and evidence functions. We also discussed embedding EEG recordings to the space of Gaussian process (GP) hyperparameters. We now show the results of applying these functions to the embedded recordings, as measured by the receiver operating characteristic curves with respect to the provided annotations.

3.3.1. Data

In this chapter we use the Canine-Epilepsy Dataset. The dataset was curated by Davis et al. [34] and made available for the UPenn and Mayo Clinic and Kaggle [21] seizure detection and prediction contests. A canine with natural epilepsy was monitored continuously for 475 days. The monitoring system, consisting of 16-electrode EEG sensors, was implanted in the brain, digitally sampling at the rate of 400 Hz (see figure 3.1). Also, 45 seizures were annotated in the dataset following validation by expert visual review of iEEG data and video recordings. The annotations are given as a sequence of timestamps which mark the start of each observed seizure event.

Train-validation split

We split the whole timeline into an early training set and a later validation set. The training set consists of 3000 samples of EEG, drawn uniformly from the first 200 days of recording. The validation set consists of another 3000 samples, drawn uniformly from day 200 to day 475. For validation purposes only, the 28 annotated seizure recordings were added to the validation set.

Pseudo-prospective study

First, the training dataset is provided to the BSLE estimator in the `fit(eeg, prior_events)` method. If `prior_events` is `None`, the model selects the uniform prior (equation 3.10). Otherwise, a list of seizure timestamps from the validation phase is given, and the model selects the circadian prior (equation 3.9) with kernel weights proportional to the circadian histogram (figure 3.7). The model then calculates the likelihood function and evidence as given by eqs. 3.6 and 3.7.

Only after the `fit(...)` procedure completes, we call `predict_proba(eeg, samples_times)` for the validation phase, which predicts the probability of a seizure, conditioned on the EEG signal, $\mathbb{P}(s_t | e_t)$ for each sample e_t , as given by eq. 3.11. This separation of concerns ensures simulating a real-time setting in which the 'offline' calibration phase is followed by an 'online' detection mode.

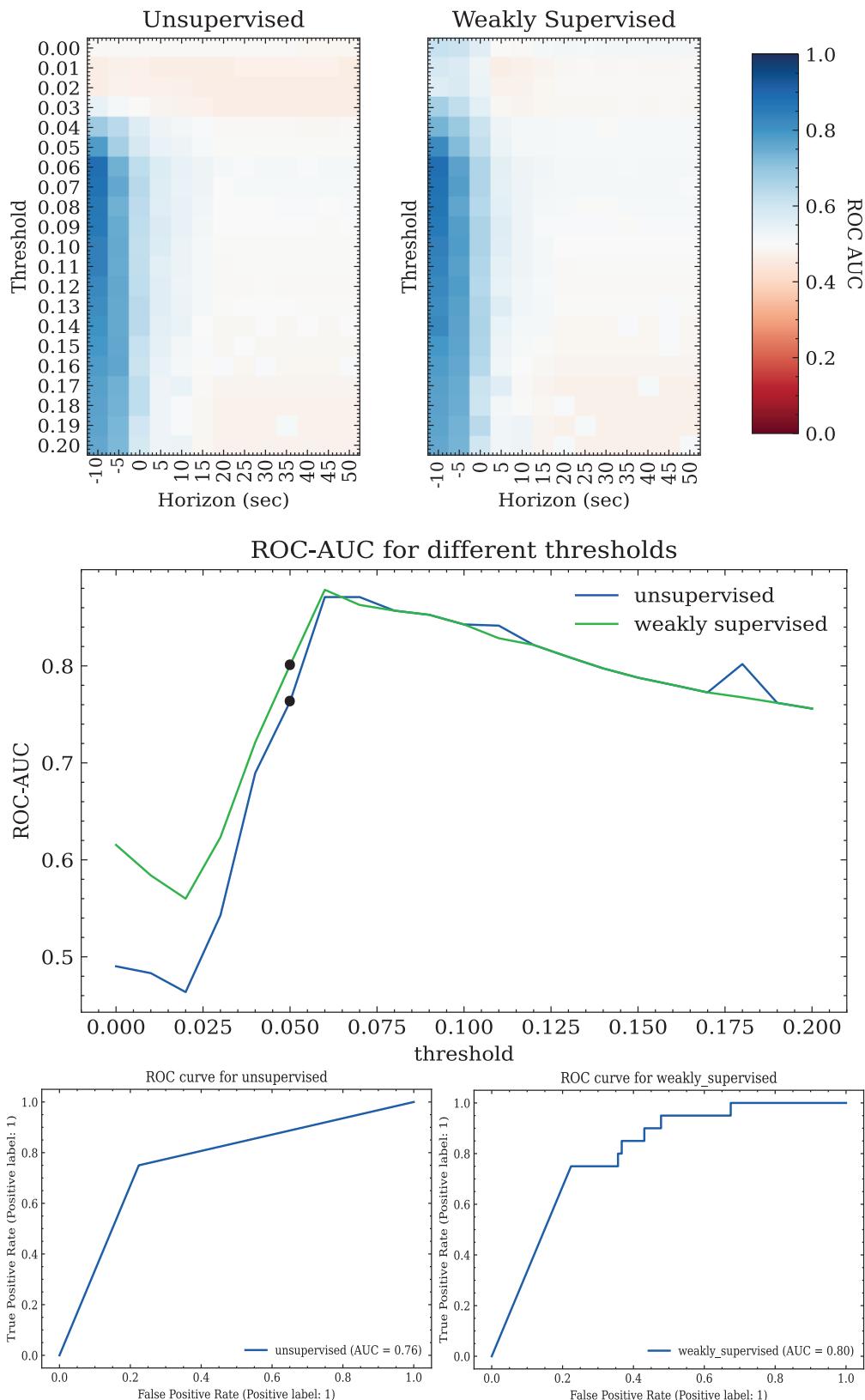


Figure 3.8. | Empirical results. AUC-ROC scores smoothly distributed in the hyperparameter space of both the unsupervised and weakly-supervised models, showing a low risk of overfitting (top row). The ROC-AUC of the weakly supervised model is usually higher than the unsupervised model (middle). An ROC curve is shown for each model when the likelihood threshold parameter is set to $\alpha = 0.05$ (bottom).

4. Conclusion

The methods shown to work in this thesis are just instances of general approaches to seizure prediction. The pattern recognition approach, presented in chapter 2, was selected because of its predominance in the reported literature and its prevalence in top-scoring competition submissions [21]. The probabilistic inference approach, presented in chapter 3, was selected as a less-traveled alternative, which is compared and contrasted to the former approach.

The pattern recognition approach consists of labeling the entire dataset, and then training classifiers to discriminate between preictal and interictal segments. We explore this approach with off-the-shelf classifiers on the Epilepsiae [23] dataset. To the best of our knowledge, chapter 2 is the first report to show that simple classifiers and nonlinear feature sets, such as those found in [20], work also for noninvasive scalp EEG. The Linear SVM achieved above chance performance on all feature sets, with AUC-ROC scores in the range 0.65-0.96. Although the pattern recognition approach has delivered promising results and found widespread public interest in competitions, it faces limitations which inhibit its development into practical appliances. One such limitation is the reliance on expert labels, which does not scale feasibly with the amount of EEG data.

The research question addressed in the second part of the thesis is, how can we formulate the task of seizure forecasting as a Bayesian inference problem? We answer this by proposing such a formulation, namely the Bayesian Seizure Likelihood Estimator, and evaluate it for both prediction and detection. This formulation utilizes the assumptions that seizures are rare, to identify rare events with a high likelihood for a seizure event, without the need for expert labels.

The Bayesian framework is general in the sense that each of the likelihood, prior and evidence functions can be estimated independently, and then combined through multiplication. In the BSLE implementation, the likelihood function is constructed as a density-estimation-based novelty detector. We experiment with two options for the prior function: either constant mean seizure-rate values, for a fully unsupervised model, or a cyclical base-rate prior for a weakly supervised model.

We evaluate both types of models on a long-term iEEG recording from a canine with epilepsy. We discover that our method achieves ROC-AUC scores of 0.88 for both the unsupervised and weakly supervised modes after hyper-parameter tuning, and that for all α -threshold values lower than 0.06, the weakly supervised model has higher ROC-AUC scores than the unsupervised model. Although this framework is suitable in theory for prediction as well as detection by choice of the horizon parameter, the empirical results (figure 3.8) show that the method's ROC-AUC scores fall down to chance levels for horizons larger than 0.5 seconds.

While the uniform prior gives equal seizure probabilities to any block of time, the cyclical prior provides time-varying seizure probabilities. Karoly et al. [16] showed that the circadian seizure histogram improved seizure forecasting in humans. To the best of our knowledge, chapter 3 is the first report to show that the same prior improves seizure forecasting in canines with epilepsy. This is another ever-so-slight indication

that seizures tend to follow cycles, an observation which has been reported many times [35].

Although the method we proposed works well for seizure detection, there are some drawbacks. First, the spatial information of the EEG channels is disregarded, and the selection of two channels to embed was made arbitrarily. Further work should utilize better channel selection and modeling the spatio-temporal qualities of the channels for potentially improved results. Second, the preictal data was not distinguishable in the GP-hyperparameter space from the interictal data. This precluded the possibility of using the method for seizure prediction. However, it is likely that different feature extraction methods, such as the ones used in chapter 2, will preserve this information and could replace the GP embedding procedure.

We hope that further attempts at seizure likelihood estimation will add to our work within the framework of probabilistic modeling. Future research directions should utilize recent advances in probabilistic programming languages and dynamic Bayesian inference. These methods could allow more complex inference procedures such as inferring subject-specific latent hyperparameters with full uncertainty quantification. In addition, the BSLE model could be extended to multiple levels of hierarchy, for example by modeling multiple subjects together or the same subject at different stages of treatment.

Bibliography

- [1] Melinda S Kelley, Margaret P Jacobs, and Daniel H Lowenstein. The ninds epilepsy research benchmarks. *Epilepsia*, 50(3):579, 2009.
- [2] Sonya B Dumanis, Jacqueline A French, Christophe Bernard, Gregory A Worrell, and Brandy E Fureman. Seizure forecasting from idea to reality. outcomes of the my seizure gauge epilepsy innovation institute workshop. *Eneuro*, 4(6), 2017.
- [3] Eric R Kandel, James H Schwartz, Thomas M Jessell, Steven Siegelbaum, A James Hudspeth, and Sarah Mack. *Principles of neural science*, volume 4. McGraw-hill New York, 2000.
- [4] Ingrid E Scheffer, Samuel Berkovic, Giuseppe Capovilla, Mary B Connolly, Jacqueline French, Laura Guilhoto, Edouard Hirsch, Satish Jain, Gary W Mathern, Solomon L Moshé, et al. Ilae classification of the epilepsies: position paper of the ilae commission for classification and terminology. *Epilepsia*, 58(4):512–521, 2017.
- [5] Jonathan Goldstein, Churl-Su Kwon, Michael Harmon, Jeffrey Buchhalter, Alison Kukla, Susan McCallum, Lisa Raman, Susan T. Herman, Brandy Fureman, and Nathalie Jette. Seizure documentation in people living with epilepsy. *Epilepsy & Behavior*, 125:108383, 2021. ISSN 1525-5050. doi:<https://doi.org/10.1016/j.yebeh.2021.108383>. URL <https://www.sciencedirect.com/science/article/pii/S1525505021006442>.
- [6] Florian Mormann, Ralph G Andrzejak, Christian E Elger, and Klaus Lehnertz. Seizure prediction: the long and winding road. *Brain*, 130(2):314–333, 2007.
- [7] Levin Kuhlmann, Klaus Lehnertz, Mark P Richardson, Björn Schelter, and Hitten P Zaveri. Seizure prediction—ready for a new era. *Nature Reviews Neurology*, 14(10):618–630, 2018.
- [8] Leonidas D Iasemidis, Larry D Olson, Robert S Savit, and J Chris Sackellares. Time dependencies in the occurrences of epileptic seizures. *Epilepsy research*, 17(1):81–94, 1994.
- [9] S Kalitzin, D Velis, P Suffczynski, J Parra, and F Lopes Da Silva. Electrical brain-stimulation paradigm for estimating the seizure onset site and the time to ictal transition in temporal lobe epilepsy. *Clinical neurophysiology*, 116(3):718–728, 2005.
- [10] Piotr W Mirowski, Yann LeCun, Deepak Madhavan, and Ruben Kuzniecky. Comparing svm and convolutional networks for epileptic seizure prediction from intracranial eeg. In *2008 IEEE workshop on machine learning for signal processing*, pages 244–249. IEEE, 2008.
- [11] Florian Mormann and Ralph G Andrzejak. Seizure prediction: making mileage on the long and winding road. *Brain*, 139(6):1625–1627, 2016.

Bibliography

- [12] Viktor K Jirsa, Timothée Proix, Dionysios Perdikis, Michael Marmaduke Woodman, Huifang Wang, Jorge Gonzalez-Martinez, Christophe Bernard, Christian Bénar, Maxime Guye, Patrick Chauvel, et al. The virtual epileptic patient: individualized whole-brain models of epilepsy spread. *Neuroimage*, 145: 377–388, 2017.
- [13] Petr Nejedly, Vaclav Kremen, Vladimir Sladky, Mona Nasseri, Hari Guragain, Petr Klimes, Jan Cimbalnik, Yogatheesan Varatharajah, Benjamin H Brinkmann, and Gregory A Worrell. Deep-learning for seizure forecasting in canines with epilepsy. *Journal of neural engineering*, 16(3):036031, 2019.
- [14] Buajieerguli Maimaiti, Hongmei Meng, Yudan Lv, Jiqing Qiu, Zhanpeng Zhu, Yinyin Xie, Yue Li, Weixuan Zhao, Jiayu Liu, Mingyang Li, et al. An overview of eeg-based machine learning methods in seizure prediction and opportunities for neurologists in this field. *Neuroscience*, 2021.
- [15] Jeff Craley, Emily Johnson, and Archana Venkataraman. Integrating convolutional neural networks and probabilistic graphical modeling for epileptic seizure detection in multichannel eeg. In *International Conference on Information Processing in Medical Imaging*, pages 291–303. Springer, 2019.
- [16] Philippa J Karoly, Hoameng Ung, David B Grayden, Levin Kuhlmann, Kent Leyde, Mark J Cook, and Dean R Freestone. The circadian profile of epilepsy improves seizure forecasting. *Brain*, 140(8): 2169–2182, 2017.
- [17] Timothée Proix, Wilson Truccolo, Marc G Leguia, Thomas K Tcheng, David King-Stephens, Vikram R Rao, and Maxime O Baud. Forecasting seizure risk in adults with focal epilepsy: a development and validation study. *The Lancet Neurology*, 20(2):127–135, 2021.
- [18] Khansa Rasheed, Adnan Qayyum, Junaid Qadir, Shobi Sivathamboo, Patrick Kwan, Levin Kuhlmann, Terence O’Brien, and Adeel Razi. Machine learning for predicting epileptic seizures using eeg signals: A review. *IEEE Reviews in Biomedical Engineering*, 14:139–155, 2021.
- [19] JJ Halford, D Shiao, JA Desrochers, BJ Kolls, BC Dean, CG Waters, NJ Azar, KF Haas, E Kutluay, GU Martz, et al. Inter-rater agreement on identification of electrographic seizures and periodic discharges in icu eeg recordings. *Clinical Neurophysiology*, 126(9):1661–1669, 2015.
- [20] Piotr Mirowski, Deepak Madhavan, Yann LeCun, and Ruben Kuzniecky. Classification of patterns of eeg synchronization for seizure prediction. *Clinical neurophysiology*, 120(11):1927–1940, 2009.
- [21] UPenn and Mayo Clinic and Kaggle. seizure detection & prediction challenges, 2014. <https://www.kaggle.com/c/seizure-detection> and <https://www.kaggle.com/c/seizure-prediction>.
- [22] Palak Handa, Monika Mathur, and Nidhi Goel. Open and free eeg datasets for epilepsy diagnosis. *arXiv preprint arXiv:2108.01030*, 2021.
- [23] Matthias Ihle, Hinnerk Feldwisch-Drentrup, César A Teixeira, Adrien Witon, Björn Schelter, Jens Timmer, and Andreas Schulze-Bonhage. Epilepsiae—a european epilepsy database. *Computer methods and programs in biomedicine*, 106(3):127–138, 2012.

Bibliography

- [24] Klaus Lehnertz, Christian Geier, Thorsten Rings, and Kirsten Stahn. Capturing time-varying brain dynamics. *EPJ Nonlinear Biomedical Physics*, 5:2, 2017.
- [25] Michael G Rosenblum, Arkady S Pikovsky, and Jürgen Kurths. From phase to lag synchronization in coupled chaotic oscillators. *Physical Review Letters*, 78(22):4193, 1997.
- [26] Jean-Philippe Lachaux, Eugenio Rodriguez, Jacques Martinerie, and Francisco J Varela. Measuring phase synchrony in brain signals. *Human brain mapping*, 8(4):194–208, 1999.
- [27] Lars Buitinck, Gilles Louppe, Mathieu Blondel, Fabian Pedregosa, Andreas Mueller, Olivier Grisel, Vlad Niculae, Peter Prettenhofer, Alexandre Gramfort, Jaques Grobler, Robert Layton, Jake VanderPlas, Arnaud Joly, Brian Holt, and Gaël Varoquaux. API design for machine learning software: experiences from the scikit-learn project. In *ECML PKDD Workshop: Languages for Data Mining and Machine Learning*, pages 108–122, 2013.
- [28] Maxime O Baud, Jonathan K Kleen, Emily A Mirro, Jason C Andrechak, David King-Stephens, Edward F Chang, and Vikram R Rao. Multi-day rhythms modulate seizure risk in epilepsy. *Nature communications*, 9(1):1–10, 2018.
- [29] Maxime O Baud, Timothée Proix, Vikram R Rao, and Kaspar Schindler. Chance and risk in epilepsy. *Current opinion in neurology*, 33(2):163–172, 2020.
- [30] Lisa D. Coles, Edward E. Patterson, W. Douglas Sheffield, Jaideep Mavoori, Jason Higgins, Bland Michael, Kent Leyde, James C. Cloyd, Brian Litt, Charles Vite, and Gregory A. Worrell. Feasibility study of a caregiver seizure alert system in canine epilepsy. *Epilepsy Research*, 106(3):456–460, 2013. ISSN 0920-1211. doi:<https://doi.org/10.1016/j.eplepsyres.2013.06.007>. URL <https://www.sciencedirect.com/science/article/pii/S092012111300171X>.
- [31] Carl Edward Rasmussen. Gaussian processes in machine learning. In *Summer school on machine learning*, pages 63–71. Springer, 2003.
- [32] Jacob R Gardner, Geoff Pleiss, David Bindel, Kilian Q Weinberger, and Andrew Gordon Wilson. Gpytorch: Blackbox matrix-matrix gaussian process inference with gpu acceleration. In *Advances in Neural Information Processing Systems*, 2018.
- [33] Rob J Hyndman. Computing and graphing highest density regions. *The American Statistician*, 50(2):120–126, 1996.
- [34] Kathryn A Davis, Beverly K Sturges, Charles H Vite, Vanessa Ruedebusch, Gregory Worrell, Andrew B Gardner, Kent Leyde, W Douglas Sheffield, and Brian Litt. A novel implanted device to wirelessly record and analyze continuous intracranial canine eeg. *Epilepsy research*, 96(1-2):116–122, 2011.
- [35] Philippa J Karoly, Vikram R Rao, Nicholas M Gregg, Gregory A Worrell, Christophe Bernard, Mark J Cook, and Maxime O Baud. Cycles in epilepsy. *Nature Reviews Neurology*, 17(5):267–284, 2021.
- [36] CE Rasmussen. Cki williams gaussian processes for machine learning, 2006.
- [37] Sergios Theodoridis. *Machine learning: a Bayesian and optimization perspective*. Academic press, 2015.

Bibliography

- [38] William Falcon and The PyTorch Lightning team. PyTorch Lightning, 3 2019. URL <https://github.com/PyTorchLightning/pytorch-lightning>.

Appendices

A. Density estimation: mathematical derivations

A.1. Embedding time series with Gaussian processes hyperparameters

A Gaussian process (GP) [36] is *a collection of random variables, any finite number of which have a joint Gaussian distribution.*

A Gaussian process $f(x)$ is fully specified by a mean function $\mu(x)$ and a covariance function, or a kernel, $k(x, x')$, by-way-of:

$$m(x) = \mathbb{E}[f(x)] \quad (\text{A.1})$$

$$k(x, x') = \mathbb{E}[(f(x) - m(x))(f(x') - m(x'))] \quad (\text{A.2})$$

And it is denoted:

$$f(x) \sim \mathcal{GP}(m(x), k(x, x')) \quad (\text{A.3})$$

In this work we will take the mean function to be zero.

parameter estimation (inference)

Gaussian processes are commonly used for time series modeling with machine learning. To see why this makes sense, imagine the input x is the time point, and the output $f(x)$ is the time series value at time x . Computationally, this is made feasible by evaluating the function's values at a finite number of points of interest. Using optimization techniques, the model's hyperparameters are inferred to match observed data by maximizing the likelihood function $p(f(x) | \vec{\theta})$ (termed maximum likelihood estimation, or MLE). The learned hyperparameters capture global evolutionary dynamics of the time series.

The Matérn class of covariance functions

The Matérn class of covariance functions is given by:

$$k_{Matern}(x, x') = \frac{2^{(1-\nu)}}{\Gamma(\nu)} (\sqrt{2\nu}d)^\nu K_\nu(\sqrt{2\nu}d) \quad (\text{A.4})$$

Where:

- $d = (x - x')^T \Phi^{-2} (x - x')$ is the distance between x and x' scaled by the *lengthscale* parameter Φ .
- ν is a smoothness parameter. In this work, it is taken to be $\frac{3}{2}$.
- K_ν is a modified Bessel function.

Multitask Gaussian processes

In case $f(x)$ is a vector function, multiple output functions are modeled in conjunction for the same input values, so-called multitask Gaussian process modeling. In this case, given inputs x and x' , and tasks i and j , the covariance between two datapoints and two tasks is given by:

$$k([x, i], [x', j]) = k_{inputs}(x, x') \cdot k_{tasks}(i, j) \quad (\text{A.5})$$

Where k_{inputs} is a standard kernel (e.g., Matérn) that operates on the inputs, and k_{tasks} is a lookup table containing inter-task covariance. This is akin to capturing the inter-channel synchronicity with manually engineered features (see section 2.1.3).

Gaussian process parameters embedding

Inference of Gaussian process (GP) parameters is a well-documented approach to modeling time-series data [31]. The extension to multitask GPs enables modeling of multivariate time-series, such as the case of multi-lead EEG signals.

For each EEG segment x , the preprocessing steps include:

- Normalizing x by subtracting the mean and dividing by the standard deviation of the training set.
- initializing a GP model with zero mean, a scaled Matérn-1.5 kernel, and a rank-1 multitask covariance kernel.
- Optimizing the model's parameters to obtain a maximal marginal log-likelihood (details in appx. B).

The optimized model's parameters θ are persisted and used henceforth to represent the original EEG segment x .

A.2. Gaussian Mixture Models

Gaussian mixture models [37] are used to model the distribution of an unknown set of vectors $\{x\} \subseteq \mathbb{R}^l$ as a linear combination (i.e., a mixture) of different Gaussian distributions, that is,

$$p(x) = \sum_{k=1}^K p_k p(x | k; \zeta_k) \quad (\text{A.6})$$

where $\{\zeta_k\}$ parametrize the individual Gaussian distributions:

$$p(x | k; \zeta_k) = p(x | k; \mu_k, \sigma_k) = \mathcal{N}(x | \mu_k, \sigma_k) \quad (\text{A.7})$$

Fitting the model (e.g. via expectation maximization) provides an approximation $\hat{p}(x | k; \zeta_k)$ of the dataset's underlying pdf, which is an estimate of the data-distribution of the GP-hyperparameters $P(E | D)$

B. GP embedding: training details

For dimensionality reduction of an observed EEG segment, we performed exact inference of the GP parameters maximizing the observation likelihood, using the GPyTorch and PyTorch Lightning frameworks [32, 38].

More formally, in fitting the Gaussian processes to the EEG samples we carried out exact Type-II Maximum Likelihood Estimation for each sample. This means optimizing the model’s hyperparameters (mean module, covariance module, etc.) w.r.t maximization of the *marginal log likelihood* (MLL) of the given data \mathbf{E}, \mathbf{t} :

$$\theta \leftarrow \underset{\theta}{\operatorname{argmax}} p_f(\mathbf{E} \mid \mathbf{t}) = \int p(\mathbf{E} \mid f(\mathbf{t})) p(f(\mathbf{t}) \mid \mathbf{t}) df \quad (\text{B.1})$$

where $f \sim \mathcal{GP}(\mu, K)$ is the modeled signal before adding the homoscedastic Gaussian noise, and θ is the set of parameters to be optimized. See code for implementation details.

GP & inference (training) configuration details		
	Parameter	Value
GP params	mean module	zero mean
	covariance module	scaled Matérn-1.5 kernel
	task covariance rank	1
	number of tasks	2
	likelihood (noise model)	Gaussian (homoscedastic)
Training params	optimizer	Adam
	learning rate	0.01
	max. number of epochs	1000
	patience (early stopping)	8

Table B.1. The GP parameters and training parameters used in our experiments.

תקציר

מודלים חישוביים של התקפים באפילפסיה מ-EEG יאפשרו למכשירים לבושים להתריע בפניהם משתמשים לפני תחילת התקף. התמקדות קהילתית בשיטות זיהוי דפוסים השיגה תוצאות מבטיחות בהבנה בין התקפים מוקדמים לתפקוד מוח תקין, כפי שמייצג על ידי וקטורים של EEG. מכיוון שסימון התקפים אינו זמין בקנה מידה גדול, הצלחתן של טכניקות זיהוי דפוסים אלו תהיה מוגבלת על ידי תוויות מוחץ לתקציב ככל שהנתונים נגדלו.

אנו מציעים חלופה בייסיאנית למסוגים לאזוי התקפים וחיזוי. שיטה זו חסכונית בסימוניים ומשיגה ציון ROC-AUC 88.0 עם אפס תוויות במשימת זיהוי התקפים. גרסה בהנחיה חלשה המותה למקצים הצירקדיים משפרת את האזוי אצל כלבים. השיטה שלנו מורכבת משני שלבים: ראשית, התפלגות EEG משוערת ומקרה לרצפים חריגיים סבירות גבוהה יותר להתקפים. לאחר מכן, נוסיף ידע אפרירורי המבוסס על מעתני הזמן לשיפורים מונחים.

שיטת חישוביות בחיזוי התקפי אפילפסיה

מחקר לשס מילוי חלקי של הזוריות לקבלת תואר
"מגיסטר למדעים"

מגיש
נועם סיגל

העבודה נעשתה בהנחיית
פרופ' אורן שרייקי
וד"ר דוד טולפין

המחלקה למדעי המחשב

תאריך	חתימת המחבר
תאריך	אישור המנהה
תאריך	אישור מנהה
אישור י"ר ועדת מוסמכים	תאריך

טבת ה'תשפ"ג

באר שבע

שיטות חישוביות בחיזוי התקפי אפילפסיה

מחקר לשם מילוי חלקו של הדרישות לקבלת תואר
"מגיסטר למדעים"

מניגש
נועם סיגל

העבודה נעשתה בהנחיית
פרופ' אורן שרייקי
וד"ר דוד טולפין



הוגש למחלקה למדעי המחשב באוניברסיטת בן-גוריון בנגב

טבת ה'תשפ"ג

באר שבע