Automated detection and prediction of seizures using probing neurostimulation

A dissertation presented

by

Senan Ebrahim

to

The Division of Medical Sciences

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Neurobiology

Harvard University

Cambridge, Massachusetts

May 2019
Automated detection and prediction of seizures using probing neurostimulation

Abstract

To study and prevent seizures in patients with epilepsy, clinical neurophysiologists seek effective methods to detect seizures, and ideally predict them for timely intervention. EEG monitoring is the clinical gold standard, while single pulse electrical stimulation (SPES) has emerged as a modality for active probing of brain states.

Seizure detection on EEG data is employed in neurocritical care, epilepsy diagnosis and management, and novel therapies such as closed loop stimulation. A detector with both high sensitivity and specificity is necessary for clinical use. We introduce a generalized linear model built from a set of 141 custom features for classification of seizures in continuous EEG. In 16 rats with epilepsy exhibiting 1012 labeled seizures, we built a pooled classifier with an AUROC of 0.995.

We also aim to automate multiple additional EEG labels in neurocritical care settings. We developed a robust method using 592 features extracted from the EEG data of 97 ICU patients. An affinity propagation (AP) method was used to generate 30-50 clusters for each patient; clinical EEG experts labeled clusters by observing the medoids. We observed a 60-fold reduction in expert labeling time, without a significant change in interrater agreement.

Seizure prediction by analyzing continuous EEG is a therapeutic goal for closed loop seizure preemption. We developed pooled and individualized predictors using the following methods: (1)
support vector machines (SVM); and (2) multilayer perceptrons (MLP). We then assessed model performance using epileptic rat data. MLP yielded the highest AUROC of 0.88 on our pooled dataset of 1012 rodent seizures.

We finally studied whether SPES into the hippocampal focus of rats with focal epilepsy yields novel predictive features, as compared to EEG monitoring. In the induced seizures for three subjects, we found multiple features across time and frequency domains, including evoked HFOs, that significantly change in preictal periods. In the two subjects with multiple spontaneous seizures, we trained SVM classifiers that perform at AUROCs of 0.94 and 0.98.

These results offer new insights into the mechanisms underlying seizure initiation, and may help improve diagnostic and therapeutic approaches for patients suffering from focal epilepsy.
# Contents

Title Page................................................................................................................................. i

Copyright .................................................................................................................................. ii

Abstract .................................................................................................................................... iii

Table of Contents ....................................................................................................................... v

Table of Figures and Tables ....................................................................................................... vii

Acknowledgements ................................................................................................................... x

1. Background & Significance ....................................................................................................... 1

1.1. Epilepsy Pathophysiology & Models .................................................................................. 2

1.2. EEG & Signal Processing ................................................................................................... 5

1.3. Featurizing EEG Data ......................................................................................................... 7

1.4. Seizure Detection Literature Review .................................................................................. 10

1.5. The Human Ictal-Interictal EEG Continuum ..................................................................... 13

1.6. Seizure Prediction Literature Review .................................................................................. 16

1.7. Single Pulse Electrical Stimulation in Epilepsy ................................................................. 20

2. Automated Seizure Detection .................................................................................................. 23

2.1. Experimental and Data Collection Methods ....................................................................... 24

2.2. Seizure Detection Analytical Approach ............................................................................. 25

2.3. GLM Performance in Class Separability and Latency ....................................................... 29

2.4. Discussion of Seizure Detection Performance .................................................................. 33

3. Automated EEG Labeling in Human Patients ..................................................................... 39
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1. Affinity Propagation Clustering Approach</td>
<td>40</td>
</tr>
<tr>
<td>3.2. EEG Clustering Results</td>
<td>47</td>
</tr>
<tr>
<td>3.3. Discussion of EEG Clustering Performance</td>
<td>49</td>
</tr>
<tr>
<td>4. Automated Seizure Prediction</td>
<td>52</td>
</tr>
<tr>
<td>4.1. Seizure Prediction Approach Methods</td>
<td>53</td>
</tr>
<tr>
<td>4.2. Seizure Prediction Performance</td>
<td>56</td>
</tr>
<tr>
<td>4.2.1. SVM Prediction Results</td>
<td>56</td>
</tr>
<tr>
<td>4.2.2. MLP Prediction Results</td>
<td>58</td>
</tr>
<tr>
<td>4.3. Discussion of Prediction Model Results</td>
<td>60</td>
</tr>
<tr>
<td>5. Stimulation for Feature Enrichment in Seizure Prediction</td>
<td>63</td>
</tr>
<tr>
<td>5.1. Probing SPES in Rodents Methods</td>
<td>64</td>
</tr>
<tr>
<td>5.2. Predictive SPES Response Features</td>
<td>67</td>
</tr>
<tr>
<td>5.3. SVM Performance With SPES</td>
<td>78</td>
</tr>
<tr>
<td>5.4. Discussion of SPES Featurization and Classification Results</td>
<td>79</td>
</tr>
<tr>
<td>6. Conclusions</td>
<td>83</td>
</tr>
<tr>
<td>6.1. Summary of Findings</td>
<td>84</td>
</tr>
<tr>
<td>6.2. Implications and Clinical Applications</td>
<td>85</td>
</tr>
<tr>
<td>6.3. Future Directions</td>
<td>86</td>
</tr>
<tr>
<td>Appendices</td>
<td>89</td>
</tr>
<tr>
<td>A.1. Code Documentation</td>
<td>90</td>
</tr>
<tr>
<td>A.2. Supplementary Figures</td>
<td>91</td>
</tr>
</tbody>
</table>
Figures and Tables

Table 1-1: Features computed on EEG signals for input into classifier models.................................9

Figure 2-1: A seizure detection paradigm for training ictal versus non-ictal EEG classifier..............28

Figure 2-2: Automated seizure detection performance for generalized linear model.........................30

Figure 2-3: Ictal and non-ictal EEG are highly separable in the PC space.........................................31

Figure 2-4: Pooled GLM classifier detecting a seizure with high specificity and low latency..........32

Figure 2-5: Extrapolation of method to large single-channel Wilcoxon multifocal epilepsy dataset........33

Figure 3-1: Development of a method integrating EEG features across domains.............................41

Table 3-1: EEG features computed on Westover dataset for clustering........................................42

Figure 3-2: An unsupervised clustering model using AP on CPD-BoW applied to sample data........44

Figure 3-3: Diagram of BoW method............................................................................................45

Figure 3-4: A graphical user interface for efficient cluster-based labeling of EEG...........................46

Table 3-2: Annotation time cost (minutes) per patient.................................................................47

Figure 3-5: Interrater agreements between the model and the experts and within the experts.........48

Figure 4-1: Seizure prediction methods applied to our dataset.........................................................54

Figure 4-2: Seizure prediction performance by multiple SVM classifiers........................................57

Figure 4-3: Separability of preictal and interictal windows in the PC space for Subject 1.................58

Figure 4-4: Seizure prediction performance by multiple MLP classifiers.........................................59

Figure 5-1: Featurizing normal and preictal SPES responses.............................................................68

Table 5-1: Features calculated on SPES responses........................................................................69

Figure 5-2: Comparison of feature distribution for three induced seizures from three subjects......71

Figure 5-3: Comparison of feature distribution for 39 spontaneous seizures from Subject 1...........74

Figure 5-4: Comparison of feature distribution for 10 spontaneous seizures from Subject 2............76
Figure 5-5: Seizures terminate earlier in the SPES dataset than the ECoG cohort..........................78

Figure 5-6: Method performance of SVM trained on SPES features for two subjects.........................79

Supplementary Figure A-1: Schematic of rodent epilepsy monitoring unit (EMU)...............................91

Supplementary Figure A-2: Hyperparameter optimization for seizure detection...............................92

Supplementary Figure A-3: Diurnal seizure clustering varies between subjects...............................92

Supplementary Figure A-4: Autoencoding of predictive features by a three-layer CNN.....................93
Prof. Syd Cash has truly been a one-of-a-kind mentor to me. He envisioned the work in this thesis with me several years ago, and patiently guided me to realizing that vision. His creativity, warmth, and humor made it a joy to work with him, irrespective of the ups and downs of the research process. I will cherish the time we spent together at his whiteboard. He is a role model to whom I will always look up, and for whom I will always be grateful.

I would also like to thank the talented Prof. Marcio Moraes, who became family through our shared love of science. Without his technical guidance, we would not have been able to execute any of the experiments presented in this thesis. I am also very grateful to two colleagues in this work: Nicolas Fumeaux and Constantin Krempf. Both are brilliant scientists whose partnership was essential to our success in these projects. As a bonus, both are also wonderful people who made it fun to stay late, present at conferences, and come up with new ideas.

I cannot possibly express my gratitude for Dr. Yangling Chou, who put in hundreds of orders, managed our grant applications, and continually supplied us with hard disks, graphics cards, and delicious candy. Brian Coughlin has been an invaluable colleague, offering crucial technical assistance on electronics, and teaching me in the process. I am very grateful for the continued mentorship from my incredible mentors Dr. Eyal Kimchi, Dr. Angelique Paulk, Dr. Rina Zelmann, and Dr. Sunil B. Nagaraj. I am also thankful for the early mentorship of Dr. Gio Piantoni, Dr. JB Eichenlaub, Jason Naftulin, and Mia Borzello. The undergraduate students I worked with in lab have taught me so much in different ways: Adesh Kadambi, Aafreen Azmi, Emma Rogge, and Jen Xu. Mentoring them has been a privilege. I thank Kyle Pellerin and Maurice Abou Jaoude for teaching me about sports and for making lab fun.

I would like to express my gratitude to the professors who served on my committees for their insightful advice: Prof. Kevin Staley, Prof. Jan Drugowitsch, Prof. Ziv Williams, Prof. Venkatesh Murthy, Prof. John Guttag, Prof. Mark Kramer Prof. Chris Harvey, Prof. David Cox, Prof. Rick Born, and Prof. Gary Yellen. I would like to thank the entire labs of Prof. Karen Wilcox, Prof. Charles Lieber, and Prof. Michael Cima for being exemplary collaborators and teaching me about their fields. I am also grateful to Prof. Matthew Frosch, Prof. Rebecca Betensky, and Prof. Sebastien Haneuse for giving me my start in teaching with their courses in programming and biostatistics. I am grateful for the support of Dr. Loren Walensky, Amy Cohen, and the entire Harvard MD/PhD Office, and the Harvard Program in Neuroscience, particularly Karen Harmin.

I would like to thank the Quincy House community for giving me a wonderful home for the past ten years, particularly Lee and Deb Gehrke and the Tutors who supported me throughout my undergraduate and graduate studies. I am especially grateful to Greg Lee and Yan Chen from the Quincy Dining Hall for their delicious food and kind friendship over the past ten years.

I would like to thank my friends & family, especially my amazing mom, who gifted me a love of science, and my amazing brothers Ali & Hassaan, who have always been in my corner. Finally, my gratitude to Eliza, the love of my life, who has taught me so much about life itself.
Chapter 1

Background & Significance
1.1. Epilepsy Pathophysiology & Models

Epilepsy is a debilitating neurological disease that affects 50 million people worldwide. Medical management is successful in roughly two-thirds of these patients, while the remaining 30% may be considered candidates for surgery. Surgery is frequently contraindicated due to involvement of eloquent cortex, and even among those that undergo surgical resection, only 40-50% are seizure free at 10 years. Overall, one quarter of patients with epilepsy are insufficiently treated by current means, the majority of whom have focal epilepsies. Mesial temporal lobe epilepsy (mTLE), in which pathological activity spreads from a mesial temporal focus to surrounding cortex, is the most common form of epilepsy in humans.

The brain is fundamentally a network of functional nodes which interact to varying degrees of synchronicity. Network dynamics in the epileptic brain may transition from normal to pathological via a distinct, detectable, preictal state. Synchronicity properties of the network may fundamentally change during this state. Actively probing neural network synchronicity with neurostimulation is a novel method for predicting seizures. The overarching goal of this study is to build robust predictive models for seizures in order to improve therapeutic possibilities such as responsive neurostimulation for patients.

Epilepsy is one of the most common neurological diseases, with significant morbidity due to unpredictable, recurrent seizures. The majority of patients with any kind of epilepsy have indicated the need for predictive technologies as their top research priority. Due to the complexity of the underlying networks, epilepsies of one class, or even of one particular genetic mutation, can present clinically with great heterogeneity across patients. Mesial temporal lobe epilepsy (mTLE), both the most common and most refractory form of epilepsy, is
characterized by seizures emanating from a temporal lobe focus and spread across cortex.\textsuperscript{2} In mTLE, mossy fiber sprouting in the CA3 pyramidal cells is considered a histopathologic correlate of epileptogenesis at the circuit level, but the functional circuit rearrangement causing seizure initiation remains undetermined.\textsuperscript{14,15} mTLE is correlated with complex behavioral outputs including depression, memory loss, and the personality changes of Geschwind syndrome.\textsuperscript{16–18}

Responsive neurostimulation (RNS) is often an additional treatment option for medically refractory mTLE, in which if a seizure brain state is detected, it is suppressed by electrical stimuli delivered to the appropriate location.\textsuperscript{19} The brain’s response to electrical stimulation can also be used to characterize the generation of a seizure (ictogenesis). We can leverage this response data in real time to estimate the probability of impending seizures and accordingly trigger interventions to reduce seizure occurrence. These concepts are explored further in Chapter 1.6.

Seizures result from a stochastic loss of complexity and increase in synchronization of neural activity, in the setting of a net excitatory balance.\textsuperscript{7,20} The dynamic state of neural networks changes both on subsecond timescales, particularly at high frequencies, and minute-to-hour timescales.\textsuperscript{21} The latter timescale is of primary interest to us in predicting seizure initiation in this study, while the former is of interest in to studies focused on modeling seizure propagation and termination.

Robust translational animal models of human mTLE are essential for studying mechanisms and testing therapies. Rats given intrahippocampal microinjections of kainic acid develop a treatment-resistant form of focal epilepsy in a similar manner to human patients.\textsuperscript{22} These rats
experience focal onset seizures that typically generalize in a tonic-clonic phenotype. Additionally, these epileptic rats exhibit hippocampal lesions that closely resemble the dissected hippocampi of human patients with mTLE.  

Given the diversity of epilepsy syndromes, we considered the various types of seizures and seizure-like events in both humans and model organisms. We use the clinical definitions of seizures determined by the American Epilepsy Society, which include different temporal thresholds for different seizure types. For a generalized tonic-clonic seizure, the minimum seizure duration is ten seconds, a definition we use throughout this work. For additional events in animals with seizure-like phenotypes, we consider the clinical definitions of seizures and apply them rigorously to the animal model, before deciding on a label.

While the behavioral readout in animal models of epilepsy is widely considered analogous to human seizures, we are mindful of the deceptive similarities between behaviors that can appear when analyzing behavioral video data. Therefore, we employ Racine behavioral staging only as a reference for seizure validation, rather than as a predictive feature. In some studies of the past twenty years, spike-wave discharges accompanied by behavioral immobility in rodents have been considered to model absence seizures. However, follow-up studies indicated that these spike-wave events with apparent immobility were disruptible with tactile stimulus, in contravention of the definition of a seizure.

We ultimately determine the validity of our chosen animal model for testing interventions by observing similarities in behavioral and physiological features, as well as consistently stereotyped responses to current established therapies. Chemoconvulsant injections into the
hippocampus show numerous physiologic and histologic hallmarks of human mTLE, as compared to kindling or genetic models. Of all the published chemoconvulsant models, we chose to utilize the kainic acid intrahippocampal model because its seizure frequencies are the most reliable between rats, and the least variable in time, once spontaneous seizures emerge, as evidenced by Spearman tests. 

There are multiple excitatory loop structures in the hippocampal circuitry that can be altered to pathologically cause or therapeutically limit epilepsy. In our study, we will focus on the recurrent collaterals among CA3 pyramidal cells that increase regional excitation and promote intrahippocampal network synchronization. However, how these changes promote the synchronization characteristic of seizures is not fully understood. We will use electrical stimuli as described in Chapter 1.7 to probe synchronicity among hippocampal micro-circuits and the conditions that drive them to synchronize pathologically.

1.2. EEG & Signal Processing

Electroencephalography (EEG) was introduced in humans by Hans Berger in 1929, confirming Richard Caton’s earlier intracortical recordings in rabbits by showing clear waveform electrical activity emanating from the brain. Berger’s foundational studies established the definition of alpha waves (8-13 Hz), associated with a wakeful but relaxed behavioral state with the eyes closed, as well as beta waves (13-30 Hz), associated with attention and activity. EEG voltage deflection has been documented to result from a relatively small proportion of cortical neurons in any given area that are highly synchronized; other active but less synchronized neurons are of marginal impact. EEG has since become the standard of care in epilepsy, as well as other neuropathologies, for first-line diagnostics and monitoring.
Electrocorticography (ECoG), also called intracranial EEG (iEEG), was developed in the 1950s by Jasper & Penfield to identify epileptogenic zones in patients with epilepsy. ECoG is used clinically in patients with intractable epilepsy for brain mapping prior to resective surgery due to its relatively high signal-to-noise ratio (SNR), high temporal resolution on the order of milliseconds, and moderate spatial resolution on the order of millimeters. The ECoG field potentials generally measure cortical pyramidal neuron activity. These signals are conveyed through the meninges up until the dura, and then collected by a subdural lead. An advantage to this system as compared to conventional scalp-based EEG is that the signal is no longer conducted through the cranium, which attenuates the signal, lowering the SNR. Additionally, in the animal model setting, conventional scalp-based EEG is impractical, leading to experimental advantages of using iEEG for long-term recordings.

In order to achieve submillisecond and submillimeter resolution, particularly in deep structures of the brain, the chief recording modality in use is the local field potential (LFP). LFP results from the placement of depth electrodes into any nervous tissue of interest. LFP waveforms reflect an integration of the proportional inputs of multiple sources within brain tissue. Generally, the further the distance of the source to the recording electrode, the less it is reflected in voltage deflections of the LFP. LFP activity primarily encapsulates and synthesizes local: (1) sodium and calcium currents in synaptic activity; (2) fast sodium-driven action potentials or 'units'; (3) calcium-mediated spikes; (4) synchronous voltage-gated currents. LFPs can also include slow shifts on the order of seconds, the so-called Bereitschaftspotential, believed to be mediated by synchronized afterhyperpolarizations. In this work, we remain mindful of these intracellular dynamics underlying LFP, but we focus on featurizing to reflect an intercellular or network level of activity, since circuit level neural activity is the salient abstraction for seizure initiation in epilepsy.
Electromyography (EMG) is another recording modality sometimes deployed in the clinical monitoring of epilepsy. In particular, phenotypes of myoclonic absences, hyperreflexic epilepsy, and epilepsies with sleep involvement have all been analyzed using EMG. EMG used in combination with EEG can be informative for seizure detection approaches. In this work, we collected EMG data from rodents undergoing 24/7 EEG monitoring largely to confirm expected correlations of tonic-clonic motor activity with Racine behavioral staging derived from video recordings for behavioral/motor output. We maintain our primary analytical focus on the ECoG and LFP recordings, and featurizing them appropriately for classification or clustering.

In this thesis, we will utilize multiple recording modalities relevant to probing brain states in epilepsy, including ECoG, LFP, and EMG, all subtypes of biophysical, extracellular, electrographic data. Processing of these electrographic data includes denoising as a first step to elevate the SNR. Due to ambient AC electrical noise, the use of a band-stop filter at 60 Hz (or other AC frequencies in other countries) is universally employed. Additional filters at the harmonics of 120 Hz and 180 Hz are also often applied in clinical settings and past studies. Therefore, we use all three of these filters in our EEG analysis pipeline. Additionally, electrographic signals are often detrended or normalized to isolate the signal of interest. We selectively utilize both of these approaches in Chapters 2-6 as well, to prepare the EEG data appropriately for featurization.

1.3. Featurizing EEG Data

There have historically been two main approaches to featurizing neural time series data for classification problems: (1) manual encoding into hand-crafted features and (2) auto-encoding into features optimized through statistical optimization. In this section, we consider the various featurization approaches that have led to successful classification performance. Here we
discuss the manually encoded features that have successfully rendered highly performing classifiers, a subset of which we employ in this work. Then, we briefly review approaches with autoencoded features that are abstracted from the data by a neural network.

Manual features of interest have been employed in studies of automatic EEG clustering, seizure detection and seizure prediction. The algorithms for these approaches are covered in depth in Chapter 1.4, 1.5, and 1.6 respectively. Here, we review particular features of interest to problems of EEG classification and clustering, extracted from the literature in these areas. These features are summarized in Table 1-1.

We utilize multiple linear features from the time domain, including RMS, coastline and skewness. Fourier decomposition enabled analysis of time-series data in the spectral or frequency domain.\(^{55-57}\) We utilize several features in the frequency domain, enabling a study of brain oscillations. Cortical oscillations have been observed in multiple mammalian species as having behavioral and functional correlates. As explored in Chapter 1.2, alpha oscillations in the 8-13 Hz range are known to have a coordinating role throughout the mammalian cortex, as well as a suppression effect.\(^{34,58,59}\) Theta oscillations in the 4-8 Hz range have been repeatedly established as generated by the hippocampus and elevated during memory tasks.\(^{60-62}\) Gamma oscillations in the 25-35 Hz range were found to be associated with sensorimotor integration in Rhesus monkeys.\(^{63}\) In the setting of human pathophysiology, abnormal oscillations are known to correlate with multiple diseases including Parkinson’s disease, depression, and epilepsy.\(^{64,65}\)
Table 1-1: Features computed on EEG signals for input into classifier models.

<table>
<thead>
<tr>
<th>Feature</th>
<th>Description</th>
<th>Domain</th>
</tr>
</thead>
<tbody>
<tr>
<td>RMS</td>
<td>Root-mean square of signal(^4,66)</td>
<td>Time</td>
</tr>
<tr>
<td>Coastline</td>
<td>Absolute sum of signal derivative(^67)</td>
<td>Time</td>
</tr>
<tr>
<td>Skewness</td>
<td>Asymmetry of signal probability distribution(^4)</td>
<td>Time</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>Flatness of signal probability distribution(^4)</td>
<td>Time</td>
</tr>
<tr>
<td>Autocorrelation function</td>
<td>Linear correlation of signal with itself on lag(^4)</td>
<td>Time</td>
</tr>
<tr>
<td>Hjorth parameters</td>
<td>Activity, mobility, complexity of EEG signal(^68)</td>
<td>Time</td>
</tr>
<tr>
<td>Maximal cross-correlation</td>
<td>Correlation of one channel signal with another on lag(^4)</td>
<td>Time</td>
</tr>
<tr>
<td>Non-linear energy</td>
<td>Energy content of linear oscillator(^69)</td>
<td>Time</td>
</tr>
<tr>
<td>Shannon entropy</td>
<td>Entropy of signal – (\sum p_i \log (p_i))(^70)</td>
<td>Time</td>
</tr>
<tr>
<td>Renyi entropy</td>
<td>Entropy (\frac{1}{1-\alpha} \log \sum_i p_i^\alpha) for quadratic ((\alpha = 2))(^70)</td>
<td>Time</td>
</tr>
<tr>
<td>Spikes</td>
<td>Event thresholded by amplitude, duration, derivative(^67)</td>
<td>Time</td>
</tr>
<tr>
<td>Fractal dimension</td>
<td>Similarity of signals at multiple scales(^71)</td>
<td>Time</td>
</tr>
<tr>
<td>Phase synchronization index</td>
<td>Mean phase coherence between two signal channels(^72)</td>
<td>Time</td>
</tr>
<tr>
<td>Band power</td>
<td>Power spectral density by FFT band(^4)</td>
<td>Frequency</td>
</tr>
<tr>
<td>Normalized band power</td>
<td>Band power normalized by total power(^4)</td>
<td>Frequency</td>
</tr>
<tr>
<td>Spectral edge frequency</td>
<td>Minimum frequency at which 90% of power is retained(^4)</td>
<td>Frequency</td>
</tr>
<tr>
<td>Coherence</td>
<td>Linear synchronization of signals by band(^4)</td>
<td>Frequency</td>
</tr>
<tr>
<td>Spectral entropy</td>
<td>Signal complexity binned by frequency(^73,74)</td>
<td>Frequency</td>
</tr>
<tr>
<td>State entropy</td>
<td>Entropy as a ratio of low and high bands(^73,74)</td>
<td>Frequency</td>
</tr>
<tr>
<td>Response entropy</td>
<td>Entropy as a ratio of low and high bands(^73,74)</td>
<td>Frequency</td>
</tr>
<tr>
<td>High frequency oscillations (HFOs)</td>
<td>Event thresholded for signal filtered in [80, 200] Hz(^75)</td>
<td>Frequency</td>
</tr>
</tbody>
</table>

In addition to the manual encoding approach on which we rely throughout this work, due to its higher performance, there is a growing field of applying autoencoding networks to featurize EEG data. However, given the limited size of datasets for seizure prediction, these approaches rarely
outperform hand-crafted features. For instance, Truong and colleagues trained a 3-layer CNN to autoencode features for seizure prediction, resulting in lower performance on the same dataset than the CNN trained on manually encoded features by Hussein and colleagues.\textsuperscript{76,77} We attempted to autoencode features from the denoised EEG data with a CNN, as shown in Supplementary Fig. A-4 and observed limited separability of features (Supplementary Fig. A-4D). For the rest of our analyses as presented in this work, we opted to manually encode features for our analyses with features like those in Table 1-1, in order to maximize performance, as well as interpretability.

1.4. Seizure Detection Literature Review

Responsive neurostimulation (RNS) is an increasingly effective therapeutic option in the management of intractable epilepsy syndromes. RNS can be improved by performing in a closed loop fashion, in turn relying on algorithmic detection of seizures in two ways: (1) to reliably stimulate as early as possible once a seizure is detected; (2) to predict seizures before their occurrence, which requires training a classifier on labeled seizure and preictal data. In either application, clinicians are faced with prolonged EEG recordings that generate substantial amounts of data, resulting in a time-intensive manual labelling process. Therefore, the management of intractable epilepsy by responsive closed loop paradigms will require a highly sensitive and specific automatic seizure detector.

Given the cost of a false negative (missed seizure) in the clinical setting of epilepsy, most devices operate at high sensitivity. With this constraint, previously published automated detectors trained on insufficient data yield a prohibitively low positive predictive value (PPV), with a false positive rate (FPR) on the order of 1/hour.\textsuperscript{78–80} To address this challenge, we
introduce a generalized seizure detection model for automated evaluation of large EEG datasets.

In order to conduct our study, we sought to create a dataset with over 1,000 labeled seizures to sufficiently train our algorithms for maximal performance. Publicly available labeled EEG datasets from human patients are limited to hundreds of seizures, including the CHB-MIT, Temple, Bonn, and Duke datasets. Furthermore, each subject in the dataset rarely has more than ten seizures recorded, due to the clinical rather than research focus of the data collection. This limitation has resulted in the creation of classifiers that perform well on internal test sets but do not externally validate. Advancement of the field of seizure detection to become clinically impactful will rely on the creation of larger EEG datasets with labeled seizures. Therefore, we created our own dataset using a rodent model of focal epilepsy, in which we could record 24/7 for months at a time to collect all seizures.

Rodent models of epilepsy are indispensable for probing disease circuits and testing novel therapies due to their genetic and structural similarities to the human brain. These models provide a controlled experimental framework unavailable in human patients, which can be employed to build a large dataset of seizures. This dataset can in turn be leveraged to investigate hypotheses concerning seizure initiation, buildup, spread, and termination. Here, we use the kainic acid unilateral focal epilepsy model in rats, which reliably models mesial temporal lobe epilepsy (mTLE). We constructed a large dataset of over 1,000 labeled seizures as described in Chapter 2.1, which enabled us to effectively assess different seizure detection algorithms.

Existing approaches to seizure detection focus extensively on feature engineering. Several methods use a single feature type such as a power spectrogram for feature input. In our
approach, we utilize twenty distinct signal processing feature classes in the time and frequency domains, identified in previous studies as being relevant to seizure physiology. These features are linear and nonlinear, and are drawn from both the time and frequency domains. Basic single-channel features such as RMS and coastline computed in the time domain are broadly descriptive of seizures and other discharges indicative of net excitatory balance.\textsuperscript{66,89} Time domain features such as Hjorth parameters capture changes particular to EEG signals that may be pertinent to seizure detection.\textsuperscript{68} Additional features drawn from the frequency domain have been shown to be highly effective to detect seizures due to their characteristic spectrographic signatures, including band powers and spectral edge frequency.\textsuperscript{66,87} Finally, multi-channel features such as cross-correlation and phase synchronization index can specifically represent synchronization of activity across networks.\textsuperscript{72} See Table 1-1 for a complete summary of all features used in our study, with references to their relevance for seizure detection and prediction.

Classification algorithms previously applied to these computed features have ranged from simply thresholding on a PCA of the features to neural networks with complex architectures.\textsuperscript{86,90–92} These algorithms are generally compared to one another by using a unified metric of performance such as the area under the receiver operating characteristic (AUROC). In previous studies, performance has not been uniformly reported (e.g. as AUROC), and many datasets are proprietary, making precise performance comparisons challenging. However, we have identified two seizure detection studies on EEG data with an AUROC over 0.98. One used a convolutional neural network (CNN) trained on a combination of handcrafted and autoencoded features, yielding an AUROC of 0.983.\textsuperscript{93} Another used a random forest classifier trained on manually featurized envelopes of wavelet coefficients to produce the highest performing AUROC of 0.995.\textsuperscript{94} As a caveat, this performance was obtained on a curated highly distinct generalized
seizure type visible on a single channel, which may reduce the applicability to other seizure types.⁸⁹

We assessed several of these methods on our own dataset for performance, and ultimately employed a generalized linear model (GLM), a highly interpretable statistical approach that performs well with regularization for feature-rich data.⁹⁵,⁹⁶ We used the validated assumption of a binomially distributed response variable of seizure occurrence. We also optimized the number of features and the coefficient for the L2 ridge penalty applied to the regression, as has been done in previous work.⁹⁷

One particular challenge of seizure detection is the extremely low prevalence of ictal time in proportion to interictal time. Classifiers by default equally penalize misclassification in either direction, which results in biased performance in favor of the most prevalent class. In the setting of epilepsy, this bias manifests as a high false positive rate for seizure alarms, making detectors of limited clinical utility.⁹⁸ We address this imbalance by selectively sampling to attain a higher ratio of ictal to interictal data. We additionally labeled datasets as completely as possible, validating our approach presented in Chapter 2 by one month of data by reading through the entire EEG to identify borderline events that may have been missed by our classifier. This laborious process of labeling seizures accurately inspired an investigation into borderline ictal-interictal events, and how to automatically identify them.

1.5. The Human Ictal-Interictal EEG Continuum

In our study of seizure detection, we realized the significant challenge entailed in manually labeling all events. Our discussion of relevant work thus far has focused principally on clinically relevant epilepsy syndromes with various types of seizures with behavioral correlates. However,
in contemporary clinical medicine, it is increasingly recognized that a significant fraction of critically ill patients in the intensive care unit (ICU), even those with no history of epilepsy, exhibit nonconvulsive seizures (NCS): seizures with little or no clinical manifestations. Additionally, patients with epilepsy often exhibit NCS of which they have no awareness, but which are observed on EEG during neuromonitoring. NCS can cause neuronal injury or worsen existing injury, and are related to poor neurologic outcomes for patients. Due to the clinical significance of these events in multiple neurological care settings, in our rodent studies, we choose to also model NCS that meet electrographic criteria for seizures, regardless of a visible behavioral correlate. Therefore, we include subclinical NCS in our definition of seizure, and aim to detect them as part of our seizure detection study in Chapter 2.

Seizure-like patterns are frequently observed in patient EEGs, known as ictal-interictal continuum (IIC) patterns. These IIC patterns have feature commonalities with seizures, but do not meet the clinical requirements set by the American Epilepsy Society’s International Classification of Epileptic Seizures. IIC events are associated with increased risk of seizures and poor outcomes in critically ill patients. According to American Clinical Neurophysiology Society (ACNS) standard classification schema, IIC patterns are a group of rhythmic and periodic EEG patterns, including Periodic Discharges (PD) and Rhythmic Delta Activity (RDA). IIC patterns can be further categorized as Lateralized (L) or Generalized (G) based on whether the patterns present in a single (L) or in both (G) hemispheres.

Certain IIC patterns are correlated with an increased risk of seizures. PD, particularly GPD, constitute an independent risk factor for worse prognosis, including long-term seizure morbidity. In current clinical practice, physicians must use heuristics to assess the risk that a given EEG pattern is predictive of poor prognoses, against the risks and side effects associated with the available treatments. In order to facilitate a more objective, rigorous approach, the
relationships between the presence, frequency and duration of IIC patterns need to be studied systematically in both humans and model organisms to better understand the continuum and its clinical significance.

It can be challenging to capture NCS or IIC in routine clinical EEGs, since these recordings only typically last 20 to 30 minutes. Prolonged continuous EEG monitoring (cEEG) is therefore important for capturing NCS and IIC patterns as well as representative baseline. However, it remains a complex and time-consuming task for the electroencephalographer to interpret the large amount of cEEG data, and it is invariably the bottleneck in both research and clinical workflows. Therefore, the first essential step towards our seizure prediction studies is to automate the labeling of the cEEG.

Previous studies attempted to divide the EEG into segments by detecting event boundaries, timepoints at which the pattern changes in some feature space.\textsuperscript{105,106} Subsequently, the EEG segments were clustered based on a full set of features extracted from each segment. However, the resulting clusters were uncorrelated with specific EEG patterns. More recent studies have tried to automatically classify preset segments in fixed durations of EEG data. However, these analyses were either carried out on a limited a number of pattern types, or were used for scoring the entire EEG based on combined predetermined thresholds for each segment.

In Chapter 3, we aim to apply unsupervised machine learning methods to achieve efficient pre-clustering of NCS and IIC patterns in prolonged cEEG recordings. We utilized the affinity propagation (AP) clustering method, introduced by Frey and Dueck for datasets demanding larger numbers of clusters.\textsuperscript{107} Unlike k-means clustering, AP determines an optimal number of clusters by maximizing measures of similarity between points in the same cluster. AP clustering also renders an exemplar for each cluster, and typically runs two orders of magnitude faster.
than other methods. We hypothesize that AP clustering facilitates efficient interactive labelling of prolonged EEG recordings by experts.

1.6. Seizure Prediction Literature Review

The goal of predicting seizures is generally the implementation of a stimulatory preemption device or early warning system, as over 90% of patients with refractory epilepsy report the need for such types of devices. In current responsive neurostimulation (RNS) clinical paradigms, it is apparent that intermittent stimulation in some way inhibits seizures. However, it is unclear the extent to which this effect is mediated by direct ictal inhibition in a temporally proximal mechanism, versus the extent to which it is caused by long-term modulation of the network induced by stimulation. In one recent study of eleven patients with focal epilepsy who received an RNS implant, direct ictal inhibition was found to have no significant effect, while indirect long-term modulation was significant. Therefore, at least for some RNS devices, there is evidently no therapeutic significance to the current closed loop functionality. High power stimulation applied more selectively, only at times when seizures are likely imminent, may resolve this conundrum to make RNS truly effective as a closed loop for short-term inhibition.

Seizure prediction is predicated on the existence of a distinct preictal state, during which network dynamics are distinct from the interictal state. Studies of focal epilepsy consistently support such a unique state, which can be identified by transforming EEG into signal features including signal correlation, coherence, energy, phase synchronization and dynamical entrainment (Lyapunov exponent). Over the past decade, classifiers differentiating preictal, ictal, and interictal states have been constructed using these and many other features from a variety of EEG datasets, as discussed in Chapter 1.3. An individualized classifier takes as input multiple features of data from a single subject, and maps the data points to labeled
categories such as preictal and interictal, in order to identify all seizures in a subject. A pooled classifier pools the input data from multiple subjects to train the classification function, then classifies data from those subjects into the designated classes. Pooled and individualized predictive classifiers present distinct advantages: an individualized classifier will yield higher performance, while a pooled classifier would allow physicians to leverage a larger dataset of seizures to treat a single patient. The pooled classifier can also be tuned as additional data is collected from the patient.\textsuperscript{115}

One fundamental challenge evident in previous predictive classification studies has been the limited duration of human EEG recordings available, leading to overfit models that could not scale to longer recordings.\textsuperscript{116–118} Even for larger EEG datasets that incorporate dozens of patients, rarely do patients have more than ten seizures captured for analysis, due to limited durations spent by patients in monitoring settings.\textsuperscript{119–121} The largest number of seizures available in a labeled, validated public human dataset until 2018 was 183 seizures, labeled from 24 patients, in the EPILEPSIAE database.\textsuperscript{81} The Temple University Corpus of EEG data, fully released in 2016, was unlabeled until 2018, at which time 900 labeled seizures were validated. Crucially, these seizures come from 315 different patients.\textsuperscript{81,119,120} With a mean of less than three seizures per patient, and significant variability in epilepsy types between patients, prediction approaches have performed poorly via CNN and LSTM approaches, unlike in the setting of detection as addressed in Chapter 1.4.\textsuperscript{83} Therefore, despite ongoing progress towards open sharing and validation of labeled EEG data, the prediction problem remains significantly data-limited in human patients. Long-term studies with 24/7 EEG monitoring in model organisms such as rats have demonstrated the essential capacity to collect an order of magnitude more data per subject.\textsuperscript{86,122} As a result of this limitation, in this study, we endeavor to use animal models to create custom data sets with appropriate features and sizes to render high-
performing predictive models.

Here we review some key studies that informed our chosen methods for seizure prediction. In previous studies, performance has not been uniformly reported (e.g. as AUROC or AUPR). Particularly older studies tend to measure performance with a typical single sensitivity in the range of (0.8,0.99) and report its corresponding specificity or false positive rate (FPR). However, in the last five years, more studies are reporting the AUROC as a threshold-agnostic measure of performance. In an ideal setting, we would be able to apply the past published methods to the exact, corresponding datasets to recapitulate the AUROC; however, many datasets remain proprietary, obviating this possibility and making precise performance comparisons challenging.

In 2011, Park and colleagues returned the first robust, validated individualized (patient-specific) classifiers by which to predict seizures better than chance, by using support vector machines (SVM). They reported a sensitivity of 0.975 with a corresponding FPR of 0.27/hour. Since then, various other neural networks have rendered pooled and individualized seizure detectors with AUROCs > 0.8 that perform with a sensitivity over 95% and a specificity over 90%. Pooled models built with multilayer perceptrons (MLP), a feedforward neural network, were found to perform with a sensitivity of 0.73 with a false positive rate of 0.28/hour in one early study. In another study, an SVM trained on EEG featurized with a combination of handcrafted and autoencoded features performed with a reported accuracy of 0.862.

Additionally, a significant caveat of many published seizure prediction studies is the lack of a standard buffer period between the end of a preictal period and the start of a seizure. Of the
fifty publications pertaining to seizure prediction reviewed for this summary, only seven explicitly state the use of a buffer period, ranging between ten and thirty seconds. Notably, these classifiers with buffer periods generally have lower AUROCs than some others; the highest performance among these with buffer periods, AUROC = 0.84, comes from Hussein and colleagues’ pooled classifier CNN trained on the iEEG NeuroVista data. 

The authors of other works performing with AUROCs over 0.84 do not cite the use of any buffer period in their methodologies. Therefore, it is possible that previously mentioned classifiers with AUROC > 0.90 such as the LSTM described by Tsiouris and colleagues are being trained without a buffer between the preictal and interictal periods. These classifiers might potentially be early detectors rather than predictors of seizures. While this is also a relevant problem for patients with epilepsy and is assessed in Chapter 2, in Chapter 4 we will be focused on prediction. We consider an AUROC of 0.84 to be the state of the art for valid seizure prediction.

Prediction horizons (the durations of preictal periods) in these studies vary between 3 minutes and 3 hours. A previous 2016 worldwide Kaggle competition to predict seizures in human and canine data (winning AUROC = 0.82) used a buffer of thirty seconds and preictal horizon of one hour. Due to the field increasingly using this reference point for comparison, we also used these buffer and horizon values as our defaults for prediction, as outlined in Chapter 4.

For predictive algorithms to eventually be deployed locally on closed-loop stimulation devices for epilepsy, the runtimes must be subsecond or faster on these lightweight devices. Therefore, another consideration when selecting a predictive paradigm was the complexity of the feature space. Ideally, we will create feature spaces that can be rendered with relatively short runtimes. In this study, we generally focused on maximizing performance, agnostic to runtimes, but we consider the attendant computational complexity for future use in closed-loop systems. We
hypothesize that appropriate feature engineering will maximize predictive performance for multiple algorithms.

1.7. Single Pulse Electrical Stimulation in Epilepsy

If indeed a neurologically distinct preictal state precedes every seizure, we hypothesize that it can be optimally identified through active probing. Continuous EEG data passively recorded from patients or model organisms have found limited success in delivering scalable predictive algorithms, as addressed in Chapter 1.6.\(^4\) Therefore, in addition to using passive electrographic data, in this work, we also analyze the response to electrical stimulus for seizure prediction, an essential innovation to fulfill the objective of preictal classification.

Neurostimulation in humans is clinically applied for multiple indications. The most established of these is therapeutic deep brain stimulation (DBS) for movement disorders and psychiatric disorders.\(^{135–137}\) However, neurostimulation has also been effectively deployed diagnostically, a modality in which it is more analogous to our anticipated use for seizure prediction.\(^{138,139}\) Trebuchon and Chauvel consider the ability of electrical stimulation to intentionally trigger seizures during presurgical brain mapping.\(^{138}\) To date, neurostimulation has not been successfully used to probe brain states dynamically in the setting of epilepsy for the purpose of seizure prediction, without actively inducing the seizure.

We considered multiple stimulation paradigms that have been used to probe brain networks. Low-frequency electrical stimulus (LFES) has been used to measure phase-locking as a measure of hypersynchrony.\(^{140}\) However, 4 Hz LFES has been demonstrated to permanently alter network dynamics to increase seizure susceptibility.\(^{141}\) Inversely, irregular pulse trains have
an anticonvulsant effect on the network.\textsuperscript{142,143} To avoid these pitfalls, we instead employ single pulse electrical stimulation (SPES) to investigate network dynamics in rodent models of epilepsy.\textsuperscript{85} Other stimulus paradigms will be used in future closed loop studies as interventional stimulation. However, in the present study of probing neurostimulation presented in Chapter 5, only SPES paradigms are used.

SPES has also been used extensively to identify epileptogenic cortex in human patients, with multiple studies indicating that localization of the seizure focus is more likely to be achieved at higher spatial resolutions by SPES than passive EEG source localization efforts.\textsuperscript{144–147} SPES has also been employed to predict post-resective surgical outcomes in humans with epilepsy.\textsuperscript{148} SPES in epilepsy has primarily been applied intracortically, with stimulus delivered into one part of the cortex, and the response, a cortico-cortical evoked potential (CCEP), recorded in another.\textsuperscript{149}

SPES has previously been used in the hippocampus to evoke potentials in patients with refractory epilepsies.\textsuperscript{150} These potentials were then appropriately featurized to spatially map the hippocampal-cortical networks.\textsuperscript{151,152} This approach has also been employed in vitro and in animal models, where a wider range of stimulation types can be utilized than in patients.\textsuperscript{153,154} Based on this foundation in spatial mapping, we apply SPES in order to temporally probe the hippocampal-cortical epileptogenic circuitry. Analogously to CCEPs, after each stimulus, the intrahippocampal evoked potential (IHEP), hippocampal-hippocampal evoked potentials (HHEP) or hippocampal-cortical evoked responses (HCEP) is analyzed for likely epileptogenicity.

Having selected SPES as our probing modality, we considered optimizing the stimulus parameters of amplitude, frequency, phase, and period. We found relatively wide ranges of
SPES parameters in use in the literature: ranges of 0.05 to 5 mA amplitude, 0.01 to 1.5 Hz approximate frequency, and 0.05 to 1 millisecond durations.\textsuperscript{155-161} Intrahippocampal SPES is known to correlate with functional effects such as memory impairment in patients when amplitudes are over 5 mA, a relatively low threshold.\textsuperscript{162} We employed a unilateral stimulation paradigm, probing in only the left hippocampal focus to observe propagation effects of that stimulus contralaterally as HHEPs. However, our experimental approach allows bilateral stimulation, which we could deploy in the future in a closed-loop paradigm, as bilateral interventional stimulation is significantly more effective at limiting seizures in rats.\textsuperscript{163,164}

In this work, we endeavor to appropriately featurize the response to SPES for seizure prediction. A study by Boulogne and colleagues found that the latency to maximum CCEP response, with a simple threshold applied at 100 ms, was predictive of spatial epileptogenicity.\textsuperscript{165} In particular, cortex exhibiting delayed responses had other associated pathologies of epileptogenesis. Mouthaan and colleagues found similar results by employing a variety of features in the time and frequency domains, including maximum and minimum responses, latencies to those responses, and various band powers.\textsuperscript{166} High frequency oscillations (HFOs) in particular have been identified as markers of spatial epileptogenicity, and have been evoked in human patients with epilepsy receiving SPES, as well as animal disease models.\textsuperscript{153,167-169} In Chapter 5, we investigate whether time-domain frequency-domain features, including HFOs, contribute to a temporally predictive feature space.
Chapter 2

Automated Seizure Detection
In this chapter, we investigate a highly performing method by which labeled seizures can be automatically detected. I am a co-first author on the work presented in this chapter, which includes valuable contributions from my colleagues: Nicolas F. Fumeaux, Brian F. Coughlin, Adesh Kadambi, Jen X. Xu, Aafreen Azmi, Maurice Abou Jaoude, Sunil B. Nagaraj, Kyle Thomson, Thomas Newell, Cameron S. Metcalf, Karen Wilcox, Eyal Y. Kimchi, Marcio F. D. Moraes, Sydney S. Cash.

2.1. Experimental and Data Collection Methods

All procedures were performed in accordance with institutional and national guidelines for animal care and use for research purposes, and the study protocol was approved by the Massachusetts General Hospital Institutional Animal Care and Use Committee (IACUC). We implanted young (aged 2-3 months, n = 15) wild-type male Sprague-Dawley rats with surface electrodes, EMG pads, and intrahippocampal depth electrodes bilaterally. We induced anesthesia for surgery via nebulized isoflurane (1-3%). We drilled burr hole craniotomies for implantation of ECoG and LFP electrodes. We implanted the rats with standard electrodes for recording of neural activity from the One Channel Electrode System (PlasticsOne, Inc., Roanoke, VA, USA). In each subject’s brain, we placed a screw electrode for ECoG overlying left parietal cortex (AP = -3mm, ML = -3mm) and depth monopolar electrodes for LFP bilaterally into hippocampal CA3 (AP: -5.3mm, ML: ±4.5mm, DV: 6mm). We also placed a guide cannula made from a 23G needle alongside the left hippocampal depth electrode. We inserted an EMG pad electrode underneath the trapezius muscular plane of the neck. Finally, we placed our reference electrode in the frontal bone and the ground electrode in the occipital bone.

Following complete recovery from the surgery (after a minimum of 3 days), microinjections of kainic acid (400 nL of 2.0 g/L in 0.9% saline) were administered into left hippocampal CA3 via
the guide cannulae that were surgically fixed. The microinjection was made through a custom injector made of fused silica tubing fitted to the cannula. These kainic acid microinjections induced behavioral status epilepticus within fifteen minutes of infusion, which was self-limited within 3 hours. Following recovery, rats developed spontaneous recurrent seizures approximately 20 days post-SE.

We developed a rodent epilepsy monitoring unit (EMU) to monitor the physiology and behavior of our subjects. We recorded EEG and video from these subjects 24/7 for 3 months, with 99.8% uptime overall (animals had to be occasionally removed from the EMU for monitoring and care). We developed a custom set of hardware and software tools to acquire and analyze video and EEG data. We used the OpBox system (KimchiLab.org/opbox) for EEG recording, and a script using the OpenCV and skvideo Python libraries for video recording. We acquired EEG signals at 1 kS/s and video at 30 fps. A complete diagram of this experimental setup is shown in Supplementary Fig. A-1. The resultant dataset from these experiments are henceforth referred to as the Cash focal epilepsy dataset.

2.2. Seizure Detection Analytical Approach

We analyzed the EEG data by processing and computing features in the time and frequency domains, as enumerated in Table 1-1. We identified events of interest including seizures, epileptiform discharges, and interictal spikes by thresholding line length of the left parietal EEG at two standard deviations above mean. We reviewed the output from this non-specific event detector, and labeled specific events according to EEG morphology, including seizures. As our ground truth, we used the clinical definition of a seizure as an evolving, rhythmic, high-amplitude signal across multiple EEG channels, lasting at least ten seconds. Latency was calculated as the time from the start of electrographic seizure, as delineated by a trained
electroencephalographer, to the time when the seizure probability returned by our model crosses the threshold.

We validated select seizures by observing the corresponding video on an as-needed basis for EEG traces that were potentially representative of a new seizure type for each animal. We observed multiple convulsive seizures for every subject on video. In order to fully capture the distribution of seizures needing detection, we included seizures rated Racine 0-5, with Racine 0 being those seizures that did not produce tonic-clonic activity of any kind. However, we did not include spike-wave absence seizures or other epileptiform events with interruptible behavioral correlates, in order to ensure reproducibility of our methods.

In this study, we assess two different models of seizure detectors: (1) pooled; and (2) extrapolated. (1) In the pooled model, we include data from all subjects equally weighted in the training dataset, and measure performance on the hold-out test dataset, also drawn from all subjects. This approach approximates the clinical scenario of simultaneously recording seizures from many patients and analyzing them in aggregate to detect future seizures for all patients. (2) The extrapolated model is a leave-one-out test in which we train on data from all subjects except one, and test on this subject. This approximates the most likely clinical application, in which a new patient’s data is analyzed with a detector trained on other patients (which can later be tuned by incorporating weighted data from that patient).

We obtained an additional, larger and more varied rodent seizure dataset from the Wilcox lab, which enabled us to test this approach in a dataset with higher variance and lower bias than the Cash dataset, in which all animals had identical unilateral foci in the left hippocampus. The Wilcox dataset includes 2883 seizures from 96 rats with multifocal epilepsy, as induced by systemic injections of kainic acid. The procedures pertaining to animal use and care for this
model were previously outlined by Vargas and colleagues. These rats were recorded continuously with video-EEG for weeks at a time, using one EcoG channel. Seizures were manually annotated by expert reviewers by the same process and definitions described above. We trained a pooled generalized linear model for the Wilcox multifocal epilepsy dataset by the same methods previously described for the Cash focal epilepsy dataset.

Fig. 2-1 demonstrates our general experimental and analytical approach to classifying ictal versus non-ictal data. We built a custom MATLAB-based graphical user interface (GUI) that enabled a human EEG labeler to review and label the EEG and spectrogram of events of interest, as well as view the video and other salient information (see Appendix A.1 for access and documentation). We had a single expert label all seizures in a one month period by looking at the entire month of EEG data, and validated those labels with two other experts. Subsequently, thresholding on a single feature such as coastline or band power was used to isolate events of interest in the remaining seizure time. We observed that this first-pass thresholding method captured all seizure events in the fully validated month of data. Despite its relatively low specificity (PPV on the order of 0.1), which would be prohibitive for clinical use, this first-pass allowed us to label data faster by an order of magnitude as a first-pass.

We labeled 1027 seizures from 19 rats using this method. Rats with five seizures or fewer that died or met humane endpoint criteria within one month were excluded from the analysis, resulting in a final dataset of 1012 seizures from 16 rats. Once the seizure segment was identified as a seizure by labeling, we extracted these periods of EEG data, and processed them by detrending and applying notch filters at 60 Hz, 120 Hz and 180 Hz. We computed the features shown in Table 1-1 on these periods, divided into ten second windows, which is the minimum duration for a seizure. In the non-ictal data, we included the entire hour of data preceding every seizure, with a thirty second buffer. We also included one hour of data selected
from ten randomly chosen interictal segments of one hour in length, each of which was at least two hours away from a seizure. Thus, our sampled data was representative of all relevant non-ictal states. For our pooled dataset, we used 3054 ictal windows from 1012 seizures and 9.96E5 non-ictal windows from the non-ictal (interictal and preictal) segments. Feature values for all windows were standardized by period. We used corresponding video data to determine behavioral seizure status, as well as Racine stage the seizure segments.

**Figure 2-1:** A seizure detection paradigm for training ictal versus non-ictal EEG classifier. (A) An electrographic recording of a spontaneous seizure in a rat with focal epilepsy, corresponding to Racine 5 behavior observed on video. Channels displayed are left parietal ECoG (blue), left hippocampal CA3 LFP (orange), right hippocampal CA3 LFP (green), and trapezius EMG (red). (B) The ECoG spectrogram corresponding to this seizure. (C-D) Analogous to A-B but for an interictal (non-ictal) period of EEG for the same subject. (E) Flowchart of machine learning approach employed for training a generalized linear model, including cross-validation for optimal hyperparameters.

We calculated the features shown in Table 1-1 for each ten second window. We computed 141 features using three channels (ECoG, left hippocampal LFP, and right hippocampal LFP) and seven frequency bands. These bands were as follows: 05.-4 Hz (delta), 4-8 Hz (theta), 8-12 Hz (alpha), 12-16 Hz (spindle), 16-25 Hz (beta), 25-50 Hz (gamma), and 70-100 Hz (high gamma). Feature values for all windows were standardized by period. Corresponding video data confirmed behavioral seizure status ranging from Racine 0 to Racine 5 for all animals.
Our machine learning approach is described in Fig. 2-1E. We first split the dataset into 80% training data and 20% hold-out test data. We then implemented 5-fold cross-validation, using grid search to identify the optimal hyperparameters. When splitting the feature windows, we grouped data from within the same ictal or non-ictal segments together. The loss function employed for hyperparameter tuning was 1-AUROC to equally weight sensitivity and specificity. We utilized PCA for dimensionality reduction and visualization of the feature space. Final performance was determined by the loss computed on the hold-out 20% of the data. We used this paradigm to train GLMs with a ridge penalty, with an assumption of a binomial distribution, due to the fixed number of independent binary classification events in our system. For the GLM, the hyperparameters in the grid search were the number of principal components used for classification and the ridge coefficient, as have been used previously to optimize classifiers. The results of the GLM hyperparameter grid search are shown in Supplementary Fig. A-2.

2.3. GLM Performance in Class Separability and Latency

We first trained the classifier on data pooled from all subjects. As shown in Fig. 2-2A, the pooled classifier achieved an AUROC of 0.995 on test data. At a sensitivity of 0.99, the classifier rendered a specificity of 0.911. Operating at this threshold, with the mean seizure frequency of 13.5 seizures/week, renders a mean overall PPV of 0.35. Per Fig. 2-2B, given multiple sensitivities in the range of (0.9, 0.99), on the 95% confidence interval (CI) of prevalence in our dataset, the weekly FDR is bounded by (6.2, 487) false detection events per week, based on the prevalence 95% CI. This weekly FDR yields the equivalent of an hourly FDR in the range of (0.0369, 2.89). The optimal hyperparameters from the grid search for the pooled classifier were lambda of 0.01 and 50 feature PCs. As can be seen in Fig. 2-2C, the mean AUROC of our extrapolated classifier, in which no data from the test subject was included in training, was 0.962. The optimal hyperparameters from a grid search conducted for each extrapolated
classifier were a lambda in the range of (0.01, 1) and retention in the range of (5, 50) feature PCs with this regularization.

Figure 2-2: Automated seizure detection performance for generalized linear model.
For this analysis, 3054 ictal windows from 1012 seizures and 9.96E5 non-ictal windows from the non-ictal (interictal and preictal) segments were selected from 16 rats and pooled for this analysis. (A) Method performance by ROC for a pooled generalized linear model with ridge penalty, trained on 80% of all subject data pooled, and tested on the remaining hold-out 20% (test AUC = 0.995). (B) Performance visualized as false positive seizures detected per week, at three sensitivity levels chosen for clinical relevance. Vertical red lines indicate boundaries of 95% confidence interval for prevalence based on our dataset. (C) Method performance by ROC for an extrapolated generalized linear model with ridge penalty, trained on all data from all subjects except one, and tested on all data from the hold-out animal. ROC for each animal is shown in gray; mean ROC is shown in blue (mean test AUC = 0.962). (D) Performance visualized for extrapolated classifier as false positive seizures detected per week, at three sensitivity levels chosen for clinical relevance. Vertical red lines indicate boundaries of 95% confidence interval for prevalence based on our dataset.
To visualize the class separability of ictal versus non-ictal EEG in our highly correlated feature space, we applied a PCA to all features. Fig. 2-3 shows the data plotted on the first three principal components, with similar results for ground truth and for our classifier label, shown as a probability without thresholding. These PCA visualizations reflect the high degree of separability of ictal and non-ictal EEG data along the axes constructed, both in the ground truth label and the classifier posterior probability. This visualization validates the effective nature of the feature space in separating ictal and non-ictal EEG data.

Figure 2-3: Ictal and non-ictal EEG are highly separable in the PC space. PCA was performed on the EEG features described above for data pooled from four randomly selected animals. The EEG segments are then plotted in the space of the first three principal components. (A) True classes of each data point in the testing set, where non-ictal windows are represented by a blue cross, and ictal windows by a red circle. (B) Posterior probabilities superimposed as colors on crosses for non-ictal windows, and circles for ictal windows.

We sought to determine whether our model, in which the temporal order of ictal windows is indeterminate, continues to exhibit high levels of performance as an early seizure detector, such that it might be used for therapeutic interventions. To assess latency at higher resolution, we also trained a model on five second windows. The mean latency at the resolution of five second
windows was under five seconds, as over 80% of seizures were detected within the first five seconds. For the representative sample seizure shown in Fig. 2-4A, the overall seizure probability clearly increases specifically within the first ten seconds of the labeled seizure, as indicated in Fig. 2-4C. As observed in Fig. 2-4B, certain PC features are specifically predictive of this particular seizure; the model combines them to detect the seizure with high probability per Fig. 2-4C.

**Figure 2-4:** Pooled GLM classifier detecting a seizure with high specificity and low latency. The pooled classifier shows performance with latency under five seconds for most seizures, one representative sample of which is shown here. (A) The electrographic data is shown with the sample seizure segment highlighted in gray. (B) Standardized values for all PC features plotted in time. (C) Seizure probability computed by the model plotted in time.

Having established the performance of the GLM on our dataset, we then sought to apply our method to the Wilcox multifocal epilepsy dataset with related but distinct characteristics as previously described and summarized in Fig. 2-5A-B. We used this dataset to calculate the subset of univariate features from Table 1-1, that is those requiring only a single channel (i.e. excluding bivariate features like coherence and cross-correlation). We then applied the same
grid search of hyperparameters to train multiple GLMs and obtain the highest performing one. The performance is displayed as an AUROC of 0.963 in Fig. 2-5C.

Figure 2-5: Extrapolation of method to large single-channel Wilcox multifocal epilepsy dataset. (A) Seizure distribution of the Cash dataset (1012 seizures from 16 rats with focal epilepsy, four channel EEG). (B) Seizure distribution of the Wilcox dataset (2883 seizures from 96 rats with multifocal epilepsy, single channel EEG). (C) Method performance by ROC for a newly trained GLM with ridge penalty for the Wilcox dataset.

2.4. Discussion of Seizure Detection Performance

Our results suggest that a GLM using the features enumerated in Table 1-1 constitutes a robust method for the detection of seizures with low latency. Our pooled classifier performs at the current state of the art. As previously explored, Bose and colleagues reported an AUROC of
0.995 for a pooled classifier using a random forest method, which is a benchmark for the current state of the art.\textsuperscript{94} We have matched this performance with our pooled classifier AUROC of 0.995, as shown in Fig. 2-2A, using a simpler method with more interpretable features and reduced computational complexity. Our model’s FDR is comparable to those of previous classifiers being studied for clinical applications.\textsuperscript{78,79,173} Therefore this generalized linear model is suitable for applications in both research and clinical settings.

The performance of our extrapolated (leave-one-out) classifier exceeds the current state of the art detector, trained on a dataset of 38 neonatal human patient EEGs, rendered by Temko and colleagues with an AUROC of 0.961.\textsuperscript{174} The same group later validated their approach with a new dataset, showing comparable performance in a novel clinical setting with neonates.\textsuperscript{175}

The clinical relevance of an extrapolated classifier lies in that a new patient, with no data yet obtained, can effectively have their seizures automatically detected. While we have exceeded the state of the art for a leave-one-out extrapolated classifier, and this level of performance is relevant for research use, we believe the field has yet to meet the clinical standard for an extrapolated classifier. Assuming an operating clinical sensitivity of 0.90, our pooled classifier would render fewer than ten false alarms per week, as shown in Fig. 2-2B. However, our extrapolated classifier operating at a sensitivity of 0.90 would render hundreds of false alarms per week, as shown in Fig. 2-2D. One additional caveat of our state of the art performance is that the considerable variability between seizures in human mTLE patients may differ from that in rodent models of focal epilepsy.\textsuperscript{22,32}

In comparison to previous studies reliant on complex algorithms, the successful training of our model was contingent on the large number of labeled seizures appropriately featurized. While we optimized a pooled generalized model with 50 PCs computed from 141 features, these
features are easily computable. This lightweight computational load lends itself well to clinical applications of seizure detection reliant on low power devices.

We anticipate that performance of the extrapolated classifier in particular may be further improved by inclusion of additional animals to fully capture the distribution of focal epilepsies. Datasets with hundreds of subjects consistently train classifiers with high performance for both pooled and extrapolated classifiers, for both human and rodent EEG data. While we observed somewhat lower performance for seizure detection in the single-channel Wilcox dataset (AUROC of 0.963), we hypothesize that using multiple channels to record those seizures would yield additional multi-channel features that will improve performance to make it comparable to our pooled classifier. Therefore, our analyses of both the Wilcox multifocal epilepsy dataset and the Cash focal epilepsy dataset lend credence to the importance of bivariate features in seizure detection.

We have demonstrated the separability of ictal and non-ictal classes via PCA to further illustrate the robust nature of the feature space. Being formulated in an unsupervised fashion without class labels (as opposed to LDA, for instance), the first three principal components demonstrate the utility of the feature space itself. Just as the separability of ictal and non-ictal segments is readily apparent in the visual representation of Fig. 2-3, the GLM we trained can readily classify with the high degree of accuracy reflected in Fig. 2-2.

Our assessment of seizure detection latency rendered a comparable distribution to the state of the art. This distribution is at a resolution of five second windows; reducing the window size from five seconds to one second would increase computational complexity and lose low frequency features, leading to poorer performance. In this study, we prioritized maximal performance in seizure classification; in other applications such as responsive neurostimulation
therapy, we can use smaller windows to facilitate rapid detection at the expense of performance.

Artifacts inherent in EEG data have often been cited as a significant barrier to classifier performance for seizure detection, as well as other machine learning applications. In this study, we include artifacts in segments of both classes according to the visible morphology, unless they obscured observation of the underlying signal by an expert on both EEG and spectrogram, in which case they were eliminated from the dataset. Thus our classifier will be robust to a real-world closed-loop application in which artifact will regularly be superimposed on both ictal and non-ictal EEG data.

While we aimed to model mTLE with the kainic acid intrahippocampal rat model, which we selected for its reduced variability, we recognize that many other epilepsies such as multifocal epilepsy are not as easily modeled. In particular, the creation of a single unilateral seizure focus in the left hippocampus yields a relatively stereotyped epilepsy syndrome with limited variability in seizure phenotype. As shown in Fig. 2-5, a GLM trained on a subset of these features can still perform in a more variable disease model such as multifocal epilepsy; the performance is expected to improve if additional channels are used. Ultimately, it remains to be seen whether the performance we observed in our data set will be validated by data from human patients with various epilepsies.

Human generalized focal tonic-clonic seizures in mTLE are appropriately modeled by the kainic acid intrahippocampal rat model, although individual features of the ictal and non-ictal segments will have distinct distributions from those of the animal model. However, the same approach described here could be used with a sufficiently large human dataset to train a new model on human data with similar performance. An additional challenge in the clinical setting is the
collection of sufficient high quality recordings of seizures from each subject. While we utilized
54.0 ± 81.6 seizures per subject in the Cash dataset and 30.0 ± 48.4 seizures per subject in the
Wilcox dataset, in human patients, it is difficult to collect data from such a large number of
seizures per subject, due to clinical priorities tending towards immediate seizure control. Given
the higher degree of interpatient variability in seizures than the intersubject variability found in
our chosen model, having a sufficiently large sample from each patient is even more crucial to
performance. In the clinical setting with its attendant variability, a feature-rich approach like ours
with regularization will yield robust performance. Additionally, with the increased availability of
longer term ambulatory recordings from implanted systems leading to larger sample sizes per
patient, the clinical impact of seizure detection software such as ours will continue to grow.\textsuperscript{180,181}

An additional caveat is that we used all seizures, including Racine 0 nonconvulsive
electrographic seizures. Including electrographic subclinical seizures was done to reduce bias
for detector performance by maximizing the variability of seizures represented, despite the
attendant reduction in interrater agreement regarding the definition of a Racine 0
(nonconvulsive) electrographic seizure.\textsuperscript{182} Detection of subclinical nonconvulsive seizures has
clinical utility as well because of long-term effects on morbidity and mortality.\textsuperscript{101} However, this
poses challenging questions as to the definition of the ground truth seizure for a detector to
classify. This question is exacerbated in the setting of the rat model, where immobile “seizure-
like” states correlate with spike-wave discharges.\textsuperscript{183} These spike-wave discharges were
frequently observed in our subjects. We classified them as part of the non-ictal class rather than
as seizures, due to our corroboration of the interruptibility of these events via tactile or auditory
stimulus, contrary to the definition of seizures in human patients.\textsuperscript{28}

We envision multiple possible directions to build upon this analysis. Firstly, this classifier can be
used to implement a closed loop interventional paradigm to assess seizure control. With an
appropriate stimulatory paradigm, a significant reduction in seizure burden is expected.$^{184,185}$ Secondly, the classifier can be altered to be applied to datasets with different seizure characteristics, such as extremely infrequent seizures, which would require more heavily weighting the seizure class. Finally, this classifier can be applied as part of a fully automated seizure prediction pipeline, wherein seizures are first automatically labeled at very high accuracy, and then a predictive model is trained using those generated labels as ground truth.

In summary, our method automatically detected all seizures in a preclinical epilepsy model with a high degree of sensitivity and specificity, with performance exceeding the current state of the art. This detection algorithm will significantly reduce the need to manually review EEG data to identify seizures, allowing the field to leverage larger EEG datasets from subjects with epilepsy to analyze seizure dynamics. Additionally, it opens the door to a computationally simple method for low-power real-time detection of seizures for neuromodulatory intervention or for seizure forecasting and warning systems.
Chapter 3

Automated EEG Labeling in Human Patients
In this chapter, we describe a method by which EEG data can be rapidly categorized with limited human supervision. I am a co-first author on the work presented in this chapter, which includes valuable contributions from my colleagues: Jin Jing, Emile d'Angremont, Sahar Zafar, Aline Herlopian, Eric S. Rosenthal, Mohammad Tabaeizadeh, Justin Dauwels, M. Brandon Westover.

3.1. Affinity Propagation Clustering Approach

As discussed in Chapter 1.5, we sought a method by which to appropriately cluster EEG segments into clinically relevant label groups. Unlike for classification problems such as seizure detection, the unsupervised problem of descriptive EEG labeling requires fewer seizures to necessarily be included in the dataset. Additionally, multiple labels beyond the ictal binary are useful to clinicians describing the interictal continuum of their patients for diagnostic purposes. These additional labels describing the IIC patterns addressed in Chapter 1.5 are more clearly defined in humans than in rodent models.

Therefore, for this study, a human EEG dataset was constructed from approximately 24 hours of EEG recorded from each of 97 patients in the ICU setting. The data was received in a deidentified format, with all PHI removed before use for this analysis, in accordance with HIPAA and Partners IRB requirements. This dataset is courtesy of M. Brandon Westover and will be referred to as the Westover dataset. The cEEG data for each patient was converted to longitudinal bipolar montage and downsampled to 200Hz, since no high frequency features were used. Digital bandpass filtering between 0.5Hz to 40Hz was applied to de-noise the data before further analysis.
To include contextual information of the surrounding EEG at various scales, we computed the features shown in Table 3-1 within windows of four different lengths, centered on the two second central interval to which we aim to ultimately assign a label, as shown in Fig. 3-1A. In addition, as shown in Fig. 3-1B-C, four regional average spectrograms (LL: Left Lateral, RL: Right Lateral, LP: Left Parasagittal, and RP: Right Parasagittal) were computed, in order to integrate both spectral and spatial domain knowledge before feature extraction.

Figure 3-1: Development of a method integrating EEG features across domains. Our feature space integrates information from multiple domains to maximize accuracy of labels. (A) Temporal features are captured by the original time series, windowed as shown around a central two second interval. (B) Spatial map of EEG leads on patients grouped into regions. (C) Spectral domain reflected on regional spectrograms displayed.
As can be seen in Table 3-1, for each spatial location and temporal scale, we extracted a variety of features that describe each two second EEG interval; features included classic measures such as line length, kurtosis, entropy, nonlinear energy operator activation, relative power, power ratios, and power kurtosis.

**Table 3-1**: EEG features computed on Westover dataset for clustering.

<table>
<thead>
<tr>
<th>Feature Name</th>
<th>Summary Statistics</th>
<th>Domain</th>
</tr>
</thead>
<tbody>
<tr>
<td>Line length</td>
<td>Total line length</td>
<td>Time</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>Total kurtosis</td>
<td>Time</td>
</tr>
<tr>
<td>Shannon entropy</td>
<td>Absolute value</td>
<td>Time</td>
</tr>
<tr>
<td>Nonlinear energy operator</td>
<td>Mean, SD</td>
<td>Time</td>
</tr>
<tr>
<td>δ, θ, α, and β kurtosis</td>
<td>Total kurtosis</td>
<td>Frequency</td>
</tr>
<tr>
<td>δ, θ, α, and β relative power</td>
<td>Mean, min, SD, 95th percentile</td>
<td>Frequency</td>
</tr>
<tr>
<td>δ/θ, δ/α, and θ/α ratios</td>
<td>Mean, min, SD, 95th percentile</td>
<td>Frequency</td>
</tr>
</tbody>
</table>

Accounting for all four spatial regions, all four temporal scales, and the 37 different spectral and temporal features, we extracted 592 features to describe each two second EEG interval. We hypothesize that this rich feature space will sufficiently differentiate all patterns encountered in NCS and IIC events in cEEGs from ICU patients. PCA, with 95% variance retained, reduced the dimensionality for each feature array.

A changepoint is a time instant at which some statistical property of a signal changes abruptly. Changepoint detection (CPD) is a general method to identify these abrupt changes in a time series.\textsuperscript{186,187} The property in question can be the mean of the signal, its variance, or a spectral characteristic, among others. To implement CPD, we employed a parametric global method, implemented in the MATLAB Signal Processing Toolbox. The CPD algorithm chooses a point
and divides the signal into two sections. Then it computes an empirical estimate of the desired statistical property for each section. At each point within a section, CPD measures how much the property deviates from the empirical estimate, and adds the deviations for all points. After that, the algorithm adds the deviations section-to-section to find the total residual error.

In addition, the algorithm varies the location of the division point until the total residual error attains a minimum. In this study, the chosen statistic is the mean of the total power of EEG. In this case, CPD minimizes the total residual error from the best horizontal level for each section. Given a signal \([x_1, x_2, ..., x_N]\), if there are \(K\) changepoints to find, then the objective function to minimize is given by:

\[
J(K) = \sum_{r=0}^{K} \sum_{i=k_r}^{k_{r+1}-1} (x_i - \langle x \rangle_{k_r}^{k_{r+1}-1})^2 + \beta K
\]

where \([k_1, k_2, ..., k_N]\) are the indices of \(K\) changepoints, with \(k_0\) and \(k_{K+1}\) defined as the first and last samples in the signal respectively. The mean operator is given by:

\[
\langle x \rangle^a_b = \frac{1}{a-b+1} \sum_{i=b}^{a} x_i
\]

while \(\beta K\) represents the penalty term added to avoid overfitting. CPD rejects the inclusion of additional changepoints if the reduction in residual error does not meet the threshold. To perform the minimization, CPD leverages a recursive optimization algorithm based on dynamic programming with early abandonment. As shown in Fig 3-2B, the sampled data from the Westover dataset is segmented appropriately by CPD.
Figure 3-2: An unsupervised clustering model using AP on CPD-BoW applied to sample data. The CPD algorithm breaks EEG into segments that are relatively homogeneous between changepoints. (A) EEG spectrogram shown for a randomly selected region. (B) First, CPD delineates segments. (C) Mapping of each feature vector associated with each two second EEG interval to one of 100 words. As a result, each segment obtained from CPD is represented as a sequence of words (“sentence”). For each segment, we calculate the histogram of words or “Bag of Words” (BoW). For details on this method see Fig. 3-3. (D) Finally, we use the chi-squared AP method discussed in Chapter 1.5 to cluster the EEG segments based on their corresponding word histograms.

A bag-of-words (BoW) model (also known as a “term-frequency counter”) records the number of times that words appear in each document of a collection. In this study, we analogously consider EEG recordings as a special type of “text,” with pattern vectors extracted for each consecutive two second EEG interval as the elementary “words.” For each patient, we learn a dictionary of words consisting of the most representative EEG intervals or exemplars. Those exemplars are identified as the cluster medoids by unsupervised k-medoids clustering with k=100 (chosen empirically) on the reduced feature space after applying PCA. See Fig. 3-3 for a visual outline of this BoW method. As shown in Fig. 3-2C, a BoW representation is created from each segment, which is then used as the input into a chi-squared AP method. The results of the clustering are shown on a sample of the Westover dataset in Fig. 3-2D.
Dimensionality Reduction
Features $\mathbb{R}^{592} \rightarrow$ PCA $\rightarrow \mathbb{R}^{100}$ dimensions

Dictionary Learning
Vector quantization ($k=100$)

Figure 3-3: Diagram of BoW method.
PCA is used to reduce the 592-feature space to 100 features. Then BoW clusters are learned via unsupervised from the new feature space via k-means clustering, with $k=100$. Clusters w1, w2, w3, w4 are shown of all $k=100$ clusters.

To facilitate rapid annotation, instead of labelling each consecutive interval via the traditional brute force method, our novel approach is to label only the medoids of the clusters that result from the AP, and apply the label to every member segment within each cluster. As shown in Fig. 3-4, a MATLAB-based graphical user interface (GUI) was developed to enable interactive rapid labelling by EEG experts. Our approach consists of having the expert label the cluster medoids, then apply the label given to each medoid to all segments belonging to the same cluster. The different EEG patterns that we aimed to distinguish were “Seizure,” and the most common IIC patterns discussed in Chapter 1.5, namely “LPD,” “GPD,” “LRDA” and “GRDA.” An “Other” class was added as well to cover all other conditions including baseline/background EEG and artifacts. Our GUI included a two dimensional embedding map computed using t-SNE for data visualization and exploration. This view assisted the labeler in visualizing the clusters.
In our GUI, 14 seconds of EEG from 24 hours recording is shown at a time in the window on the right, displayed in groups of electrodes from the left and right lateral (LL, RL), left and right parasagittal (LP, RP), and central regions. The regional average (LL, RL, LP, RP) spectrograms containing this EEG time point (as marked by the dash lines) are displayed on the left, with the changepoint detection results at the bottom. The unsupervised clustering membership assignment is illustrated by the horizontal color bar below the CPD panel, as determined by the CPD-BoW-AP steps. The colors given in the horizontal bar are assigned based on the average total power from all members in that cluster. The higher the power values (usually correlates with severity of the EEG patterns), the darker the color. Above the spectrograms, the t-SNE embedding map retains the intrinsic structure/relationship of samples in feature space; points that are close to each other in this embedding map are very likely to be similar patterns in feature space. Each scattered point in this map corresponds to a 592-dimensional feature vector extracted from a two second EEG interval.

Three EEG experts labeled the center two second interval of the medoid of each cluster of all patients, using the proposed GUI. Hereafter, all clusters were labeled according to the visually scored medoid for all three experts. Once labels were applied to the clusters by all three experts, we applied the following merging procedure to reduce label noise, in accordance with clinical practice. For each of the following three pairs of labels, if two experts agreed on one label and the third expert had the other, the third expert’s label was changed to agree with the other two: (Seizure, LPD); (Seizure, GPD); and (Seizure, LRDA).
To assess the validity of our approach using the clusters as belonging to shared classes, we leveraged a supervised approach in which we trained on cluster labels (auto-labels) and tested on human-applied medoid labels. We applied a Gaussian kernel SVM to show that the clusters validly separate different EEG patterns. For this purpose, we randomly selected 100 two second EEG intervals for each label, excluding the medoid that was shown to the experts, and used the PCA-reduced feature arrays of these intervals as training data. The center two second intervals of each cluster medoid, which were all visually scored by the experts, were used as the testing data. We did this analysis for each patient and for all three experts separately. We hypothesized that the pairwise agreement between the model and each expert would be at least as high as the agreement between each expert and the other experts.

### 3.2. EEG Clustering Results

The cEEG recordings had a mean duration of 30.19 ± 3.84 hours, and the BoW-based AP clustering of each recording resulted in a mean number of clusters of 27 ± 11. Table 3-2 shows the time needed by the experts to label the full approximately one day of data by only labeling the medoid of each cluster presented in the GUI.

**Table 3-2: Annotation time cost (minutes) per patient.**

<table>
<thead>
<tr>
<th>Source</th>
<th>Mean ± SD</th>
<th>Median</th>
<th>Interquartile Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Expert 1</td>
<td>4.99 ± 4.83</td>
<td>3.63</td>
<td>(2.23, 5.58)</td>
</tr>
<tr>
<td>Expert 2</td>
<td>2.08 ± 1.38</td>
<td>1.55</td>
<td>(1.12, 2.94)</td>
</tr>
<tr>
<td>Expert 3</td>
<td>8.85 ± 3.33</td>
<td>8.40</td>
<td>(6.74, 11.34)</td>
</tr>
<tr>
<td>Overall</td>
<td>5.31 ± 4.44</td>
<td>3.95</td>
<td>(1.68, 7.67)</td>
</tr>
</tbody>
</table>

Fig. 3-5 indicates that there were no significant differences between the groups. The mean pairwise interrater percentage agreement between the model and each expert was tested by
Welch t-test without any statistically significant result (p>0.05). This finding indicates that the features that were randomly selected from each cluster were informative; the model was able to predict the score of an expert at least as well as would the label of another expert. Thus, it is likely that the clusters were formed in such a way that they truly distinguish different pattern types.

Figure 3-5: Interrater agreements between the model and the experts and within the experts. Raincloud plots of the percentage agreement between the model (M) and the experts (1-3) pairwise, as well as between experts (1-3) pairwise.

Our automated clustering method performed with a mean pair-wise interrater agreement of 72.4 ± 19.4 percent, when averaged across all patients, as shown in Fig. 3-5. Similarly, the
average pairwise interrater agreement between experts was 67.0 ± 23.4 percent. There is no significant difference between these distributions either as determined by t-test (p > 0.05).

### 3.3. Discussion of EEG Clustering Performance

We present an unsupervised clustering method to categorize cEEG data into a limited number of clusters, which can rapidly be annotated by an EEG expert with an easy to use GUI. This allows experts conducting EEG labeling to quickly identify the presence of seizures and clinically relevant IIC patterns in patients undergoing neuromonitoring.

The validity of the clustering was assessed by the agreement of the clustering with the experts. If the cluster rendered was pure, i.e. if the whole cluster indeed belongs to one specific pattern type, the agreement would be high. Inversely, if there were a large variety of pattern types within a cluster, the agreement would be low. As can be seen in Fig. 3-5, the agreement between the experts and the automatic clustering is at least as good as the interrater agreement of all three experts. The lack of significant difference in any pairwise test supports the validity of this method: interrater agreements between the automated method and the human experts and those calculated among human experts are comparable in performance.

Previous studies of EEG data labeling have generally yielded similar human expert interrater agreements, and lower agreements with automated methods. Most studies report a subset of Gwet’s AC1, percent agreement, and Cohen’s kappa. When including IIC events along with seizures, interrater agreements reported by others fall to 62-66 percent agreement.\(^{191,192}\) This drop is due to the relatively high prevalence of borderline events that meet multiple clinical criteria. The majority of studies conducted using experts labeling EEG data from the ICU or EMU including both seizure and multiple IIC labels have returned kappas in the range of 0.50-
Azuma and colleagues conducted a related study in the setting of sleep and rendered a kappa of 0.63 for expert labeling various EEG abnormalities including fast waves, slow waves, and extended spindles. While Gaspard and colleagues report the highest percentage agreement in the field of 0.93 (kappa of 0.91), their study includes seizures but excludes IIC rhythmic patterns such as periodic discharges and rhythmic activity, which we labeled in our study. IIC rhythmic patterns like GPD are close in interrater agreement, reportedly at a kappa of 0.81 in another study. As Wusthoff and colleagues demonstrated, when measuring expert interrater agreements, they average at kappa of 0.93 for seizures but as low as 0.20 for shorter rhythmic discharges. Therefore, we consider 66 percent agreement to be a performance benchmark for interrater agreement in labeling of all IIC and seizure patterns, which our mean agreement exceeds for both our experts and automated method.

Our automated clustering method exceeds previously reported performance of automatic systems on cEEG from the ICU. Experts using NeuroTrend, a closed source solution by Encevis, initially recorded a machine-expert kappa of 0.38 (compared to 0.79 for inter-expert) and later a Gwet's AC1 of 0.57. Our order of magnitude reduction in time was also comparable to that reported by Koren and colleagues. A multiple feature classifier trained to predict a binarized ‘reactivity’ label assigned by EEG experts, performed with a pair-wise machine-expert AC1 ranging from 0.65-0.7.

To further improve interrater agreement for both manual labeling, and our clustering approach, we can modify the visualization of EEG data. For instance, displaying neonatal EEGs with amplitude integration to labelers has proven to increase interrater agreement. It is possible that similar EEG manipulations can yield improvements in interrater agreement in the ICU.
space, though they may also mask subtle events to which an IIC label belongs, voiding the performance improvement by eliminating borderline events. While our GUI presented a novel t-SNE map alongside familiar representations of the EEG data as voltage traces and spectrograms, additional novel methods for EEG visualization have emerged to make EEG interpretation easier for human experts. Whether these methods can improve interrater agreement remains to be seen.

Labeling the cEEG data with our method took the experts 60 times less time than manual labeling: the median time per patient was less than 4 minutes. This shows that this method is fast and easy to use. Apart from saving the critical care neurologist a significant amount of time, with this method, a large labeled dataset can be rapidly generated. This dataset can in turn be reliably used for relating the different pattern types to patient outcome in a supervised manner.

We are building on previous work that has demonstrated the clinical utility of an EEG clustering approach. We anticipate that this study and future ones like it will significantly improve clinical workflows for neurointensivists, who currently work to manually label large quantities of data. Our clustering method performs robustly enough that clinicians labeling data will only have to label a representative subset of a patient’s data and have the algorithm apply the labels across the dataset. In addition to saving clinician time, we preserve accuracy by eliminating long labeling sessions and allowing experts to evaluate exemplary EEG segments in more detail. The resultant performance yields state of the art interrater agreement with other human expert labelers.
Chapter 4

Automated Seizure Prediction
In this chapter, we investigate methods by which seizures can be accurately predicted. I will be a co-first author on the work presented in this chapter, which includes valuable contributions from my colleagues: Emma Rogge, Constantin Krempp, Nicolas F. Fumeaux, Aafreen Azmi, Brian F. Coughlin, Adesh Kadambi, Sunil B. Nagaraj, Eyal Y. Kimchi, Marcio F. D. Moraes, Sydney S. Cash.

4.1. Seizure Prediction Approach Methods

In our approach to seizure prediction, we used both conventional machine learning and deep learning methods. We utilized well-established binary classification frameworks for seizure prediction as outlined by Mormann and colleagues in two reviews of the field spanning the past decade. As reviewed in Chapter 1.6, various prediction horizons and buffers have been used. In our case, we chose thirty seconds as the buffer and one hour as the horizon, consistent with the plurality of studies in the last five years, and the 2016 Kaggle competition framework. We sampled one hour segment of interictal data for every hour segment of preictal data (one per seizure). We ensured that interictal data was at least two hours away from any seizure, and then randomly sampled from a uniform distribution.

The general machine learning paradigm utilized for training models for seizure prediction is reflected in Fig. 4-1A. The feature set computed on the preictal and interictal windows is the same full feature set described in Table 1-1 and utilized for seizure detection in Chapter 2. We considered and applied several methods that had performed with high AUROCs, as summarized in Chapter 1.6. In this chapter, we focus on the use of two classification methods:

1. Support vector machines (SVM) use a kernel transform to achieve nonlinear classification of data. SVM is non-probabilistic, rendering a categorical output (i.e.
preictal versus interictal binary). In this chapter, as in Chapter 3, we utilize a Gaussian kernel. Hyperparameters to optimize include number of features, the box constraint, and the kernel scale. The box constraint weights the cost function for misclassification, while the kernel scale provides a distribution basis implied by the kernel.\textsuperscript{204}

(2) Multilayer perceptrons (MLP) constitute a feedforward neural network which uses an input layer, a hidden layer, and an output layer by which we can assess class label. We employed a basic architecture with one hidden layer, the sole hyperparameter optimized was the number of neurons in the hidden layer. Fig. 4-1B diagrams the network.

![Figure 4-1: Seizure prediction methods applied to our dataset.](image)

**Figure 4-1:** Seizure prediction methods applied to our dataset.  
(A) This flowchart shows our general machine learning paradigm used for seizure prediction, including cross-validation for optimal hyperparameters. (B) MLP schematic diagram for three-class classifier. Hidden layer optimized by grid search to have between 100 and 150 neurons.

Both of these methods have been previously applied to seizure prediction, with AUROCs reported in the range of (0.7, 0.8); however, multiple authors observe performance generally
being limited by the data and features.\textsuperscript{125,205–209} We utilize these methods because they are readily generalizable and maximally leverage optimal feature engineering for relatively small feature spaces (<1000 features), such as our hand-crafted feature space.\textsuperscript{204}

Additionally, we use the MLP approach to simultaneously identify a basis that is relevant to seizure prediction and detection, framing the analysis as a three-class problem (preictal versus interictal versus ictal). Performance is then assessed per class in a one-versus-all manner. For the MLP ROCs, we also compute both a micro-average and a macro-average. A macro-average includes equal weights from all classes in the average. Inversely, a micro-average carries over prevalence imbalance by not equally weighting classes in an average. Both are useful summary statistics for a MLP multi-class classifier.

We considered prediction for multiple datasets, including the datasets in Chapter 2 and 3, and other public datasets, ultimately deciding to use the Cash focal epilepsy dataset from Chapter 2 due to the high number of seizures overall (1012) and the number per subject (mean of 54.0 ± 81.6). While the dataset provided by our collaborators in the Wilcox lab had a comparable mean of 30.0 ± 48.4 seizures per subject, and an even larger total number of seizures (2883), that EEG data was collected from a single channel. Our four channel system (three EEG and one EMG) provided the opportunity to assess synchronization across brain regions as discussed in Chapters 1.2 and 2.1, which would not be possible with the Wilcox dataset. These bivariate features measuring synchronization are known to be highly predictive as discussed in Chapter 1.6. Therefore, all analyses presented in this study are conducted from our Cash focal epilepsy rodent dataset described in Chapter 2.

In this chapter, we assess two different types of seizure predictors: (1) pooled; and (2) individualized. (1) In the pooled model, we include data from all subjects equally weighted in the
training data, and measure performance on the test data, also drawn from all subjects. This approach approximates the clinical scenario of simultaneously recording seizures data from many patients and analyzing them in aggregate to predict future seizures for all patients. (2) We also create an individualized predictor for each subject, in which we train and test only on data from that single subject. Similarly, we ensure that the training data precedes the test data. This approach approximates the device-level clinical application, in which a large quantity of a single patient’s data is analyzed, with a predictor trained to predict seizures from only that patient.

4.2. Seizure Prediction Performance

In this section, we assess the ability of the classifiers we trained using the SVM and MLP methods described in Chapter 4.1 to perform on the large focal epilepsy dataset described in Chapter 2. We generally assessed performance primarily as AUROC, which is increasingly the standard for performance comparison.\textsuperscript{98,129,134,203}

4.2.1. SVM Prediction Results

As shown in Fig. 4-2, we trained both a pooled SVM, and twelve individualized SVMs for subjects with at least ten spontaneous seizures. The pooled SVM was tested to perform at a mean AUROC of 0.70 on the pooled data set, with 93 PCs resulting from the grid search. Individualized SVMs classifiers performed with considerably higher AUROCs with a mean of 0.81 ± 0.11, one example of which is shown in Fig. 4-2B. The optimal number of feature PCs varied between 70 and 110.
Figure 4-2: Seizure prediction performance by multiple SVM classifiers. (A) ROC for pooled SVM classifier trained on features from Table 1-1 computed on the rodent focal epilepsy dataset. (B) ROC for sample individualized SVM classifier trained for Subject 1 of the rodent focal epilepsy dataset (AUROC = 0.86). Mean AUROC of individualized classifiers was 0.81 ± 0.11.

In order to visualize the separability of the preictal and interictal classes, we plotted all featurized windows in the space of the first three PCs. The results of this visualization can be seen in Fig. 4-3 for the individualized classifier for Subject 1 (see Fig. 4-2B for corresponding ROC). The ground truth class labels (Fig. 4-3A) align closely with estimates rendered by the SVM for likelihood of belonging to class 1 (preictal).
PCA was performed on the EEG features described in Table 1-1 for data from Subject 1. The EEG segments are then plotted in the space of the first three principal components. (A) True classes of each data point in the testing set, where non-ictal windows are represented by a blue circle, and ictal windows by a yellow cross. (B) Posterior probabilities superimposed as colors on crosses for non-ictal windows, and circles for ictal windows.

4.2.2. MLP Prediction Results

We trained the multi-class predictive MLP model using a mini-batch size of 512 and 150 neurons in the hidden layer, found to be optimal on grid search. With the MLP approach shown in Fig. 4-1B, we attained an AUROC of 0.88 for preictal classification (one versus all), as plotted in Fig. 4-4A. The micro-average AUROC for the pooled MLP was 0.95, and the macro-average AUROC was 0.92. Because ROC curves indicate the true positive and false positive rates agnostic to relative prevalence, and the dataset is heavily skewed towards preictal and interictal data as compared to ictal data (representative of actual prevalence of events) we also visualize these three-way results as a PR curve. The corresponding PR curve is shown in Fig. 4-4B,
demonstrating robustness to the prevalence imbalance in the dataset. The pooled classifier loss reached convergence before 500 iterations for every fold of the 5-fold cross-validation, as shown for example in Fig. 4-4C. Finally, individualized MLP predictors performed with comparable AUROCs with a mean of 0.88 ± 0.10, as shown for an example of Subject 1 in Fig. 4-2D. The individualized ROCs were not significantly different between SVM and MLP as determined by a two-tailed t-test (p = 0.12). The MLP classifier had a predictor AUROC of 0.93, a micro-average AUROC of 0.96, and a macro-average AUROC of 0.95.

Figure 4-4: Seizure prediction performance by multiple MLP classifiers. (A) ROC for pooled MLP classifier trained on features from Table 1-1 computed on the rodent focal epilepsy dataset. AUROC for preictal versus all is 0.88. (B) Corresponding PR for MLP classifier described in (A). AUPR for preictal (class 1) versus all is 0.76. (C) Representative sample fold-loss for MLP classifier described in (A). (D) ROC for sample individualized MLP classifier trained for Subject 1 of the rodent focal epilepsy dataset (AUROC = 0.93). The mean AUROC for preictal classifier was 0.88 ± 0.10.
4.3. Discussion of Prediction Model Results

We trained pooled and individualized predictive classifiers that met or exceeded the state of the art as measured by AUROC. Perennial challenges of seizure prediction include limited seizures per patient, limited numbers of patients, and tradeoffs between interpretability and prediction performance. We addressed these problems by creating an ideal dataset, as reviewed in Chapter 2. We then created and evaluated two methods to predict seizures, with an emphasis on performance. Through these methods, we sought to define a distinct, classifiable preictal state.

We have demonstrated the separability of preictal and interictal classes via PCA to further illustrate the robust nature of the feature space (see Fig. 4-3). This visualization supports the existence of such a distinct preictal state, while the first three principal components demonstrate the utility of the hand-crafted feature space itself. The individualized SVMs can readily classify interictal and preictal EEG data with the high degree of accuracy reflected in Fig. 4-2, with a mean AUROC of 0.81 ± 0.11, on par with the state of the art at AUROC = 0.84. The individualized MLPs also performed similarly as shown in Fig. 4-4 with 0.88 ± 0.10.

The MLP approach renders a highly performing pooled classifier with an AUROC = 0.88, performing above the previously established state of the art for a pooled classifier (AUROC = 0.84). The low variance of the fold-wise AUROCs suggests that we are avoiding overfitting in this approach. Additionally, the similarity of the micro-average and macro-average suggest that the performance is robust to differences in preictal versus interictal versus ictal relative prevalences. The three-output MLP uniquely enabled us to frame the problem of seizure prediction as a multi-class problem, and include seizure data in the training set, thereby representing the full class space of EEG: all EEG is either preictal, interictal, or ictal. The MLP
performed with a higher AUROC than the state of the art, but other neural network architectures may achieve yet higher performances, particularly with additional data.

Two further methods currently being investigated are long short-term memory (LSTM) networks, which have previously been studied in the context of seizure prediction, and Sequence Transformer Networks (STN) models, which have not yet been deployed in our field. In the past year, STNs have been adapted from the natural language processing field to successfully predict clinical outcomes such as mortality with an AUROC of 0.851. As with others in our field attempting these methods, our performance continues to be data-limited, despite having created the largest available seizure dataset with over 1,000 seizures. In order to overcome this constraint, we are working to obtain a larger dataset from collaborators.

While the pooled SVM renders a low performing pooled classifier with AUROC = 0.70, it is useful as a reference point to compare future work in other feature spaces, for instance that presented in Chapter 5. For individualized classifiers, both MLP and SVM approaches yielded predictive classifiers above the state of the art benchmark, with all mean AUROCs > 0.80. These individualized classifiers are recommended to be the new state of the art, and will be used as a basis for comparison for future work.

The predictive capability of the pooled SVM in particular may have been limited by a few factors that are worth discussion. Firstly, as a kernel-based method, SVM is relatively susceptible to overfitting. This is particularly true in a high variance setting, such as our pooled SVM, which includes data from twelve different rats. Although this classifier has less bias than any other predictive modality, the variance is prohibitively high, resulting in relatively low performance. Therefore, the resultant performance is significantly better for individualized classifiers. Secondly, SVM tends to be the preferred approach for smaller datasets, such as the
individualized dataset on the order of 10-100 seizures, as compared to the pooled dataset with over 1,000 seizures, where deep learning methods begin to excel. Finally, unlike MLP, SVM does not have an optimal multiclass implementation in use, despite the development of the directed acyclic graph SVM (DAGSVM) to address this challenge.\textsuperscript{214}

For future studies of individualized classifiers in seizure prediction, SVM appears sufficient to achieve state of the art performance. For larger datasets of pooled data with higher variance, MLP is more robust and can achieve state of the art performance. These methods demonstrate two useful performance benchmarks for future comparison as the field continues to leverage novel methods to continually improve our ability to predict seizures.
Chapter 5

Stimulation For Feature Enrichment in Seizure Prediction
In this chapter, we investigate novel stimulation methods to improve performance in seizure prediction. I will be a co-first author on the work presented in this chapter, which includes valuable contributions from my colleagues: Constantin Krempp, Brian F. Coughlin, Nicolas F. Fumeaux, Adesh Kadambi, Aafreen Azmi, Rina Zelmann, Angelique C. Paulk, Marcio F. D. Moraes, Sydney S. Cash.

5.1. Probing SPES in Rodents Methods

In order to assess the efficacy of SPES response features in seizure prediction, we developed a novel paradigm with probing SPES in our established rat model of mTLE. All procedures were performed in accordance with institutional and national guidelines for animal care and use for research purposes and the study protocol was approved by the Massachusetts General Hospital institutional animal care and use committee (IACUC). We devised the electrical stimulus paradigm for a new trial cohort of animals after careful consideration of: (1) animal health and safety; (2) translational relevance to human patients; (3) scientific validity; and (4) statistical principles. For our probing stimulus, we used single pulse electrical stimulation (SPES), which has been previously used in rodents and humans to map epileptogenic cortex and predict epilepsy progression.\textsuperscript{144,147,148,215}

We implanted young (aged 2-3 months, n = 4) wild-type male Sprague-Dawley rats with surface electrodes, EMG pads, and intrahippocampal depth electrodes bilaterally, as well as intracortical depth electrodes on the left (side of future chemoconvulsant injection). We induced anesthesia for surgery via nebulized isoflurane (1-3%). We drilled burr hole craniotomies for implantation of EEG and LFP electrodes. We implanted the rats with custom electrodes with appropriate impedances (ranging from 60 to 100 kOhms); impedances were measured weekly post-implantation to ensure they continued to range within a factor of two of their original measurement. We manufactured these electrodes using stainless steel wires with teflon coating.
(127 micron inner diameter / 203 micron outer), molded and fixed to proper stereotaxic coordinates (California Fine Wire Company, Grover Beach, CA).

In each subject, during a single surgery, we placed subdural screw electrodes overlying parietal cortex bilaterally (AP = -3mm, ML = ±3mm). We also placed one screw electrode overlying left frontal cortex (AP: 2.5mm, ML: 2mm). We inserted depth quads of electrodes bilaterally into hippocampal CA3 (AP: -5.3mm, ML: ±4.5mm, DV: 6mm). We also inserted a depth quad of electrodes into left entorhinal cortex (AP: -6mm, ML: 4.5mm, DV: 8mm). We additionally placed a guide cannula made from a 23G needle alongside the left hippocampal depth electrode. We implanted an EMG pad electrode underneath the trapezius muscular plane of the neck. Finally, we placed our reference electrode in the frontal bone and the ground electrode in the occipital bone.

Following complete recovery from the surgery (after a minimum of 3 days), microinjections of kainic acid (400 nL of 2.0 g/L in 0.9% saline) were administered into left hippocampal CA3 via the guide cannulae that were surgically fixed. The microinjection was made through a custom injector made of fused silica tubing fitted to the cannula. These kainic acid microinjections induced behavioral status epilepticus within fifteen minutes of infusion, which was self-limited within 3 hours. Following recovery, rats developed spontaneous recurrent seizures approximately 20 days post-SE.

We utilized a single pulse electrical stimulation (SPES) paradigm to probe evolving hippocampal dynamics in epilepsy. Post-implantation, prior to the induction of epilepsy, we continuously delivered a biphasic 0.2 ms pulse with a 0.5 mA amplitude via a bipolar electrode in the left hippocampus. These characteristics have been determined based on a survey of existing studies in rodents and humans as undetectable, painless, without adverse events, yet sufficient
to probe the local network.\textsuperscript{148,156,216} We observed in all four rats that this stimulation was able to evoke a visible electrographic response (see Fig. 5-1), but not so powerful as to disrupt normal electrographic function and behavior or cause adverse effects such as pain, dysfunction, or awareness of stimulation. This SPES was delivered at intervals randomly chosen from a uniform distribution in the range of three to five seconds.

We built a new rodent epilepsy monitoring unit (EMU) with stimulation capability to monitor the physiology and behavior of our subjects. We recorded sixteen EEG channels and video while delivering SPES continuously on two left hippocampal depth leads from these subjects 24/7 for 3 months, with 99.9\% uptime overall (animals had to be occasionally removed from the EMU for monitoring and care). We used the Intan RHS2000 hardware stack (Intan Technologies, Los Angeles, CA), and developed a custom set of software tools to acquire and analyze video and EEG data. We used a custom version of the open-source Intan RHS2000 system software (intantech.com/products_RHS2000.html) for EEG recording, modified to accommodate continuous stimulation triggering by our specific method. We simultaneously used a script using the OpenCV and skvideo Python libraries for video recording. We continuously acquired EEG signals at 20 kS/s and video at 30 fps.

In addition to capturing at least one day of baseline EEG before kainic acid injection and the SE seizure induction itself, we aimed to capture all spontaneous seizures over three months with continuous SPES. The SPES responses captured are variously IHEPs, HHEPs and HCEPs, as enumerated in Chapter 1.7. In this chapter, we refer to these various responses collectively as SPES responses. While we analyzed them separately due to each channel of the 16 having its own, the same approach was used for all SPES responses. The seizure labeling process employed for this dataset was identical to that described in Chapter 2.2, employing low thresholds tested to capture all events of interest. The same criteria for seizures as described in
Chapters 2-3 were used for this labeling process, notably including the ten second minimum duration.

Finally, to build a prediction model from the spontaneous seizures, an individualized SVM was trained for each subject, according to the same paradigm presented in Chapter 4.1, but using the new feature set described in Chapter 5.2. Due to the comparable prevalence of preictal and interictal data in this dataset, we utilized ROC curves as a performance measure.

5.2. Predictive SPES Response Features

The use of SPES in rodents presents the unique opportunity to identify response features that are predictive of seizures. The stimulation parameters that evoke these features will inform the design of future closed-loop interventions in both rodents and human patients. From this trial cohort of four rats, we conducted a preliminary analysis of the induced seizure on three subjects, and a predictive analysis of three months of spontaneous seizures in two subjects. Data from one animal (Subject 4) has not yet been collected, and data from another (Subject 3) was only partially collected - a delay in data collection caused by two malfunctioning headstage chips, which have both since been replaced so that the spontaneous seizure data from these subjects can be collected and analyzed in the future. Therefore, the analyses presented below are conducted on seizure data collected from three subjects (Subjects 1-3), with two of them (Subjects 1-2) having multiple spontaneous seizures recorded and analyzed.

We recorded 39 seizures from Subject 1, and 10 from Subject 2, while capturing data with 99.9% uptime to ensure the validity of interictal labels. The limited loss of data results from when animals had to be disconnected periodically for animal care purposes (no seizures occurred during this time). Subject 3 is only included in the study of the single acute status
epilepticus (SE) seizure during induction. Subject 3 is excluded from the predictive classification study because rigorous analysis of spontaneous seizure prediction requires successful 24/7 data capture to ensure accuracy of preictal versus interictal labeling.

Fig. 5-1 shows the raw signal trace of a SPES response both preictally (5-1A) and interictally (5-1B) in eight of the sixteen channels recorded. To analyze the SPES response data, we first isolated a response window of fixed duration after the stimulus in the range of (100, 1000) ms, with a buffer of 3ms after the stimulus delivery, as shown in Fig. 5-1. We then detrended all response windows by subtracting the mean signal of a baseline segment of windows. All data was downsampled to 2 kS/s due to the lack of need for high frequency features. The mean signal of each channel was calculated for one hundred randomly chosen segments in a non-ictal segment. This signal mean for each channel was then subtracted from each response window in the subsequent segments. In this way, we were able to remove the response artifact and amplifier settle components of the recorded signal. We then featurized the data as discussed in Chapter 5.2. Code documentation for this work can be found in Appendix A.1.

Figure 5-1: Featurizing normal and preictal SPES responses.
Raw data recorded from Subject 2 on eight different representative channels from both preictal and interictal time segments while a 1 Hz stimulus is delivered. (A) During a representative preictal segment prior to an induced seizure, a response window 3ms following the seizure is sampled and detrended before featurization. Here, a window of 100ms is shown for illustration, delineated by vertical black lines. This sampling procedure is repeated for windows between 100ms and 1ms. (B) Same as (A), but for a representative interictal segment. Hcp: Hippocampus; Ent: Entorhinal Cortex, L: Left, R: Right.
The features used for SPES responses were purposefully different from those in Chapters 2-4. The reason for this is three-fold: (1) the features that might be maximally predictive for seizures are likely to be different; (2) SPES response potentials have a response window measured relevant to the stimulus time, which renders unique features like latencies; (3) keeping the feature space relatively bounded will improve the likelihood that we can adapt this approach into closed-loop applications. Table 5-1 indicates which features were calculated for the SPES data. These eleven univariate features were each computed on all sixteen channels, leading to a complete feature vector of 176 features for each SPES response window.

**Table 5-1: Features calculated on SPES responses.**

<table>
<thead>
<tr>
<th>Feature Name</th>
<th>Summary Statistics</th>
<th>Domain</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maximum voltage</td>
<td>Mean, SD</td>
<td>Time</td>
</tr>
<tr>
<td>Minimum voltage</td>
<td>Mean, SD</td>
<td>Time</td>
</tr>
<tr>
<td>Latency to maximum voltage (argmax)</td>
<td>Mean, SD</td>
<td>Time</td>
</tr>
<tr>
<td>Latency to maximum voltage (argmin)</td>
<td>Mean, SD</td>
<td>Time</td>
</tr>
<tr>
<td>Delta power (1-4 Hz)</td>
<td>Mean band power</td>
<td>Frequency</td>
</tr>
<tr>
<td>Theta power (4-8 Hz)</td>
<td>Mean band power</td>
<td>Frequency</td>
</tr>
<tr>
<td>Alpha power (8-13 Hz)</td>
<td>Mean band power</td>
<td>Frequency</td>
</tr>
<tr>
<td>Beta power (13-25 Hz)</td>
<td>Mean band power</td>
<td>Frequency</td>
</tr>
<tr>
<td>Gamma power (25-50 Hz)</td>
<td>Mean band power</td>
<td>Frequency</td>
</tr>
<tr>
<td>High Gamma power (70-110 Hz)</td>
<td>Mean band power</td>
<td>Frequency</td>
</tr>
<tr>
<td>HFO power (130-170 Hz)</td>
<td>Mean band power</td>
<td>Frequency</td>
</tr>
</tbody>
</table>

As a first-pass analysis of the earliest data collected, we investigated for the three induced SE seizures, one of which was recorded from each of Subjects 1-3, whether the 4-feature time-domain space delineated a preictal state. This analysis served as a sanity check to ascertain that our experimental stimulus paradigm was indeed effectuating relevant SPES response
features. In this analysis, due to the absence of interictal data, the mean baseline pre-injection response window of a sample of 100 windows was used as the baseline for detrending of all other windows, both preictal and baseline. As demonstrated in Fig. 5-2, certain features showed significant differences in their preictal versus baseline distribution, as determined by Welch t-test with a Bonferroni correction. Fig. 5-2A shows the significantly different preictal and interictal distributions of maximum response voltage response from an entorhinal depth LFP channel. Additionally, the temporal trend towards an increase in maximum response voltage in this channel preictally is shown in Fig. 5-2B. As can be seen on the colormap of the time-domain features in Fig. 5-2C, most features similarly retained significance in most channels.
Figure 5-2: Comparison of feature distribution for three induced seizures from three subjects. 

(A) Preictal (blue) versus baseline (green) maximum response voltage for an entorhinal LFP channel (channel 1) from Subject 2 plotted as a histogram. 

(B) Preictal (blue) versus baseline (green) maximum response voltage for the same entorhinal LFP channel (channel 1) from Subject 2 plotted versus time. The red line indicates the start of the relevant induced seizure. 

(C) Colormap of feature versus channel significance matrix by Welch t-test with Bonferroni connection. Significance as shown in red indicates that for all three animals this channel and feature were significantly different.
In a subsequent and more comprehensive analysis, we used the labeled spontaneous seizures from Subjects 1 and 2 to assess both time and frequency domain features. We recorded and labeled 39 spontaneous seizures from Subject 1, and 10 from Subject 2. In this analysis, we used a one second window with a 3ms buffer from the stimulus delivery. Then the mean response window of a sample of 100 interictal windows was used as the baseline for detrending all other windows, preictal and interictal. Next, for all detrended response windows, we calculated the features shown in Table 5-1. We then executed pairwise comparisons of the preictal and interictal feature distributions by Welch t-test, optimal here because of the unequal variances and sample sizes. We applied a Bonferroni correction (alpha level of 0.01) based on the number of comparisons. Significance on this test indicates that there is a very low probability (below threshold alpha of 0.01) that these two sample distributions of SPES responses (preictal and interictal) came from the same underlying population.

Fig. 5-3 shows the results of this feature significance analysis for Subject 1, while Fig. 5-4 similarly shows these results for Subject 2. In Fig. 5-3A, constructed from the analysis of Subject 1, we see that most channels are significant for most features. Fig. 5-4A shows the same results in Subject 2. Fig. 5-3B-G highlight six examples of significantly different distributions per Welch t-test in Subject 1, with Fig. 5-3B-G doing the same for Subject 2. These were chosen from the matrix for illustrative purposes to show changes in temporal characteristics, multiple spectral bands, and HFOs.

As can be seen in Fig. 5-3B-C, minimum voltage of SPES response, shown for channel 2 (left entorhinal LFP) and channel 15 (left hippocampal CA3 LFP), is bimodally distributed. In the preictal distribution of Fig. 5-3B, a new mode appears, indicating that some significant portion of the time, the preictal entorhinal HCEP distribution takes on a third range of values that is never seen during interictal windows. In the IHEP distribution shown in Fig. 5-3C, the change between
distributions takes the form of a significant rightward shift on the distribution, indicating a lower absolute valued minimum voltage response preictally. This finding is recapitulated in Fig. 5-4B for a HCEP distribution of minimum response voltage on channel 0 (left entorhinal LFP). A similar rightward shift is observed on Fig. 5-4C for the maximum response voltage in the same channel HCEP.

An additional point of interest in the SPES response feature space is that there is a significant change in the delta and theta power bands. While delta power appears to be left shifted, theta power appears to be right shifted on both HCEP (Fig. 5-3D-E) and IHEP (Fig. 5-4D-E). Finally, as shown in Fig. 5-3F and 5-4G, there is a significant increase in HFO power preictally on IHEP. This effect is also visible on left-sided HCEP, including at the left frontal ECoG leads (Fig. 5-3G), and entorhinal cortical depth leads (Fig. 5-4F).
Figure 5-3: Comparison of feature distribution for 39 spontaneous seizures from Subject 1. (A) This colormap for a single animal (Subject 1) was created on the basis of all 39 seizures, with preictal and interictal data sampled as described. Six histograms from positive t-tests are highlighted for illustrative purposes from the summary colormap, as marked in white. Channels are as follows: 0-3: left
Figure 5-3 (continued): entorhinal LFP; 4: EMG; 5-8: right hippocampal CA3 LFP; 9: left parietal ECoG; 10: right parietal ECoG; 11: left frontal ECoG; 12-15: left hippocampal CA3 LFP. (B) Distribution of the minimum voltage feature in a left hippocampal CA3 LFP channel during preictal versus interictal windows. (C) Distribution of the minimum voltage feature in a left entorhinal LFP channel during preictal versus interictal windows. (D) Distribution of the delta (1-4 Hz) power feature in a left entorhinal LFP channel during preictal versus interictal windows. (E) Distribution of the theta (4-8 Hz) power feature in a left entorhinal LFP channel during preictal versus interictal windows. (F) Distribution of HFO (130-170 Hz) power feature in a left hippocampal CA3 LFP channel during preictal versus interictal windows. (G) Distribution of HFO (130-170 Hz) power feature in the left frontal ECoG channel during preictal versus interictal windows.
Figure 5-4: Comparison of feature distribution for 10 spontaneous seizures from Subject 2.

(A) This colormap for a single animal (Subject 2) was created on the basis of all 10 seizures, with preictal and interictal data sampled as described. Six histograms from positive t-tests are highlighted for
Figure 5-4 (continued): illustrative purposes from the summary colormap, as marked in white. Channels are as follows: 0-3: left entorhinal LFP; 4: EMG; 5-8: right hippocampal CA3 LFP; 9: left parietal ECoG; 10: right parietal ECoG; 11: left frontal ECoG; 12-15: left hippocampal CA3 LFP. (B) Distribution of the minimum voltage feature in a left entorhinal LFP channel during preictal versus interictal windows. (C) Distribution of the maximum voltage feature in a left entorhinal LFP channel during preictal versus interictal windows. (D) Distribution of the delta (1-4 Hz) power feature in a left hippocampal CA3 LFP channel during preictal versus interictal windows. (E) Distribution of the theta (4-8 Hz) power feature in a left hippocampal CA3 LFP channel during preictal versus interictal windows. (F) Distribution of HFO (130-170 Hz) power feature in a left entorhinal channel during preictal versus interictal windows. (G) Distribution of HFO (130-170 Hz) power feature in a hippocampal CA3 LFP channel during preictal versus interictal windows.

While there was no statistically significant difference in seizure prevalence between this cohort and the Cash focal epilepsy EEG dataset described in Chapter 2.2, as measured by a Welch t-test (p = 0.86), we observed possible early termination of a few seizures due to the SPES while observing primary data. Fig. 5-5 illustrates the significant difference between the ECoG dataset from Chapter 2 and the SPES dataset in terms of seizure duration. We pooled the data from all animals (without normalizing by subject, so subjects with more seizures are weighted proportionately). The duration of seizures is significantly different in the SPES dataset, as compared to the ECoG dataset (Fig. 5-5A). A Welch t-test on the duration of labeled seizures in each dataset did render a statistically significant reduction in seizure duration in the SPES dataset (p<0.01).

The distribution is clearly visibly lower in Fig. 5-5A, with a threshold at 10 seconds due to our clinical definition of a seizure as lasting a minimum of ten seconds. We observed in both Subject 1 and Subject 2 that a beta sub-band of 12-20 Hz was the dominant seizure frequency. In Fig. 5-5B, it is evident that the power is disrupted by the stimulus and takes over one second to recover. Therefore, it is possible that SPES still results in early seizure termination as it disrupts the beta synchronicity of circuits in a seizure, particularly as it is delivered every three to five seconds.
Figure 5-5: Seizures terminate earlier in the SPES dataset than the ECoG cohort.
(A) This histogram compares the distributions of seizure durations between the SPES dataset used in Chapter 5 and the ECoG dataset used in Chapter 2. The mean seizure duration for the SPES dataset is 17.4 ± 11.0 seconds, while the mean seizure duration for the SPES dataset is 38.2 ± 15.1 seconds, a significant difference as measured by a t-test (p<0.01).
(B) Peristimulus mean power plotted around SPES for each channel in Subject 1. Stimuli are delivered only into the left hippocampus as a bipolar pulse on channels Stim 0/1. Channel legend abbreviations are as follows. EMG: electromyography; Rec: channel used for recording only; Stim: channel capable of stimulation.

Having described the SPES feature space in the context of spontaneous recurrent focal seizures, we now proceed to the training of models to predict seizures in this feature space.

5.3. SVM Performance With SPES

To address the ultimate question of whether or not in aggregate these response features can meaningfully train a classifier to predict seizures, we trained an SVM model on a per-subject basis (individualized classifier). An SVM classifier was trained using one-second SPES response windows, featurized as described in Chapter 5.2. These windows were drawn from 245 minutes of interictal data, and two minutes of preictal data per seizure, with a minimum one second buffer period. Each minute had an average of 14 response windows, due to our setting of the stimulus frequency as randomly chosen from a uniform distribution in the range of (0.20, 0.33) Hz, per Chapter 5.1. For Subject 1, this consisted of 1055 windows of preictal data and
3362 windows of interictal data, sampled around 39 seizures. For Subject 2, this consisted of 10 seizures leading to 260 preictal and 3362 interictal windows.

**Figure 5-6**: Method performance of SVM trained on SPES features for two subjects. (A) Individualized classifier for Subject 1, trained on 80% of data with optimized hyperparameters and tested on held-out 20%. ROCs are shown for each of the five cross-validation folds. Mean AUROC is 0.94. (B) Same as (A) but for Subject 2. Mean AUROC is 0.98.

Fig. 5-6 shows the performance of an SVM trained on the SPES data according to the same paradigm described in Chapter 4. Once again, the optimal hyperparameter of the number of features was sought by grid search in the range of (60, 120). The 5-fold validation loop generated five ROCs for each classifier. The AUROC for all folds for Subject 1 was in the range of (0.94, 0.95), as shown in Fig. 5-6A, with a mean AUROC of 0.94. The AUROCs for the folds of Subject 2 were in the range of (0.97, 1.0), with a mean at 0.98.

### 5.4. Discussion of SPES Featurization and Classification Results

This chapter investigated the hypothesis that the response to SPES can be appropriately featurized to predict seizures. The time-domain features that were routinely found to be significantly different in distribution in both the induced seizure and spontaneous seizures fit with previous characterization of CCEP response latency.\textsuperscript{149,165} Our identification of the increase in theta power intrahemispherically on both IHEP and HCEP is consistent with previous findings of
pathologic theta oscillations in patients with epilepsy. Furthermore, evoked HFO power was found to be significantly different in most channels, fitting the clinical understanding of HFOs as potentially predictive of seizures. Our findings of temporal feature distinctions are also consistent with spatial characterizations of evoked HFOs. These spectral features collectively demonstrate highly powered aberrant synchronous activity that is evoked at multiple frequencies by a single pulse.

To our ultimate clinical objective of maximizing performance for predicting seizures, we built two SVMs, one for each subject with spontaneous seizures captured. Even with a relatively small amount of data from each of the two subjects, i.e. less than fifty seizures, individual classifiers are performing with high levels of sensitivity and specificity, as shown in Fig. 5-6. The individualized classifier AUROCs rendered (0.94, 0.98), resulting in orders of magnitude less false positives than the current state of the art based on passive EEG (AUROC = 0.84). These individualized AUROCs also exceeded the means of both SVMs (AUROC = 0.81 ± 0.11) and those of the MLPs (AUROC = 0.88 ± 0.10) on the passive EEG dataset in Chapter 4, despite the relatively low number of seizures (39 for Subject 1, 10 for Subject 2). This makes our study highly relevant to future clinical applications, if the effect persists in larger and more varied cohorts of animals with epilepsy.

We are also mindful of the statistically underpowered nature of this study on the SPES dataset, given the low number of animals, as well as relatively low number of seizures. We continue to employ our rigorous definition of seizures in this section, resulting in the elimination of seizure-like events of 8-10 second durations. For adequate statistical power to conduct a generalized study of seizure prediction with SPES, additional experiments must be completed on more subjects. However, this proof of concept study still rigorously investigates on a per subject basis whether or not SPES features are informative to prediction of multiple seizures.
One important caveat of the study of the induced seizures, the results of which are displayed in Fig. 5-2, is that the induction of status epilepticus by injection of kainic acid is distinct from the endogenous ictogenic mechanisms of the epileptic brain. However, these processes are analogous in many ways, including electrographic, histologic, functional, and behavioral characterizations.\textsuperscript{219–221} Therefore, the acute status epilepticus event, although predictable in occurrence, displays significant differences in our feature space during preictal periods. This study is included here in order to assess the validity of the feature space across three animals, since Subject 3 could not be included in the spontaneous seizure analysis.

Another caveat of our main feature significance analyses, presented in Fig. 5-3 and Fig. 5-4, is that the feature distributions are generally not normally distributed. In cases where population distribution cannot be assumed to be normal, statistical theory generally recommends the use of nonparametric tests. However, in practice, a Welch t-test is robust to the assumption of normality as long as variables are independent, as is the case here.\textsuperscript{222,223} Additionally, keeping all tests two-tailed mitigates potential increases in Type I error as shown by Monte Carlo simulations.\textsuperscript{223} Therefore, for our features, parametric tests for significance remain practically valid to compare preictal and interictal data.

As shown in Fig. 5-5, we identified an unintentional effect of SPES in disrupting seizures. Due to only having two animals with spontaneous seizures recorded during SPES, it is not yet possible to make exact recommendations for calibrating the stimulation in a closed-loop setting. However, in principle, the amplitude and frequency of probing stimulus can be reduced such that the SPES is unable to disrupt a seizure or otherwise perturb the network, while interventional stimulus can be set to comparable parameters as were interrupting the seizure.
As discussed further in Chapter 6.3, future experiments should focus on minimizing the number of stimuli further. In addition to reducing the inadvertent effect of seizure termination, this approach will mitigate other potential long-term confounding effects of neurostimulation that reduce epileptogenicity over time, without regard to seizure prediction. While we did not observe these effects, others have reported them as significant. However, the stimulation should be delivered often enough that a robust sample of preictal response windows is available, such that seizures will not be missed. Therefore, pursuant to the analysis in Chapter 5.2, we recommend future studies operate in the range. Additionally, we recommend continued use of the “jitter” or sampling from a uniform random distribution bounded within 25% of the target frequency to avoid entraining the network, while ensuring consistent stimulation.

Finally, with an eye to closed-loop implementation of SPES-based prediction, we observed that the runtimes to calculate these simple features for one-second response windows measure in microseconds on an Intel Core i7-4790 3.6 GHz Quad-Core Processor, and when appropriately downsampled, only require 10MB of RAM for one hour of featurized data. This contrasts to the computationally intensive feature spaces described in Chapters 2-4 of this work, as well as in previous studies as described in Chapter 1.3-1.6. Therefore, implementation in a closed-loop experimental paradigm as well eventual deployment in a mobile device setting become feasible with this approach. Additionally, the high speed of computation theoretically enables sub-millisecond orders for interventional stimulatory horizons.
Chapter 6

Conclusions
6.1. Summary of Findings

Collectively, our findings in all chapters reveal the importance of feature engineering. Whether our approach was supervised or unsupervised, and focused on prediction or detection, the crux for high performance was in the feature space.

In Chapter 2, we constructed a generalized linear model for seizure detection in a rodent model of mTLE. Our method used a custom set of features to automatically detect seizures rodents with focal epilepsy a high degree of sensitivity and specificity, with performance matching the current state of the art. Our method also generalized to another dataset, the Wilcox dataset, which modeled multifocal epilepsy rather than mTLE.

In Chapter 3, we established a method to show that cEEG data can be validly clustered into a small number of distinct patterns by applying a novel method. Our results suggest that long EEG recordings can be rapidly annotated by experts at least 60 times faster than unaided manual review.

In Chapter 4, we constructed support vector machines and multilayer perceptron (MLP) classifiers that can individually predict seizures at performance levels comparable to the state of the art. Additionally, the pooled MLP classifier exceeds the state of the art. These results suggest that the feature space we used uniquely defines a distinct preictal state that can be detected. This feature space is further optimally refined by the MLP for classification.

Finally, in Chapter 5, we applied SPES to the hippocampal focus in the robust animal model of mTLE. We identified several SPES response features in multiple channels that are significantly different preictally as compared to interictally. Aggregated in an SVM, these features render
highly specific and sensitive classifiers of SPES data into preictal and interictal classes for seizure prediction.

**6.2. Implications and Clinical Applications**

This project was undertaken due to the significant clinical burden of epilepsy, and the potential positive ramifications of a seizure predictor that works for patients. The implications of this work are: (1) developing a new modality for seizure prediction by active SPES probing; (2) developing sensitive and specific classifiers meeting clinical standards; and (3) informing future research methodologies. Here we consider the clinical and research applications of each Chapter.

This detection algorithm presented in Chapter 2 will significantly reduce the need to manually review EEG data to identify seizures, allowing the field to leverage larger EEG data sets from subjects with epilepsy to analyze seizure dynamics. Additionally, it opens the door to a computationally simple method for low-power real-time detection of seizures for neuromodulatory intervention or for seizure forecasting and warning systems. However, whether this detection approach can scale to clinical applications remains to be seen.

Using our automatic EEG labeling method as demonstrated in Chapter 3 on the Westover dataset, we are currently in the process of labeling >30TB of EEG data from 2,000 ICU subjects. The resulting EEG data will provide sufficient data to train deep neural network models to automatically detect NCS and IIC patterns. This rich data will also allow us to gain a deeper understanding of the clinical consequences of NCS and IIC events, and how the consequences depend on the attributes of different NCS and IIC patterns. This EEG labeling method is also now being investigated in its application to a similarly large clinical neonatal EEG dataset.
While our MLP seizure prediction algorithms yielded comparatively high AUROCs of 0.88 (higher than the state of the art), the clinical relevance of this finding remains to be studied. This caveat is particularly true for the pooled classifier, as the variability between patients with mTLE in terms of predictability is likely greater than that exhibited in our well-controlled animal model. However, this MLP classification approach should be attempted on a large human dataset with relatively stereotyped focal epilepsies to elucidate clinical validity.

As we discussed in Chapter 1, RNS as a therapeutic modality in epilepsy currently does not utilize the signal evoked by stimulation for any kind of predictive purpose. By our analysis of the SPES dataset, we have demonstrated that at least for certain animals (n = 2), with certain types of focal epilepsy, spontaneous recurrent seizures are highly predictable in the SPES response feature space we describe in Table 5-1. These studies should be extended in animal models, and if clinically and ethically appropriate, explored in patients with refractory focal epilepsies.

**6.3. Future Directions**

The fields of seizure detection and seizure prediction are currently framed largely as binary classification problems; studies like ours will continue to be of primary research use in this research context. However, in the future, we and others in the field envision that a probabilistic framework will be optimal for clinical applications. In future studies, we plan to reframe some of our analyses in the seizure forecasting framework recently proposed by Freestone and colleagues. We have synthesized a seizure dataset that captured all seizures from 19 rats for six months, leading to multiple analytical projects. It is well established that seizures in many patients temporally cluster. We aim to study seizure clustering and subtyping, as well as analyzing the temporal pattern of seizures over the timescale of days. We plan to integrate that information into the predictive model, featurized as time since last seizure and seizure count in the
preceding 24 hours. We hypothesize that this approach will improve the performance of our prediction methods.\textsuperscript{227}

A preliminary temporal clustering analysis on a subset of seizures from a subset of animals is shown in Supplementary Fig. A-3. We found that results are highly variable between animals, as they are in human subjects with focal epilepsy. Some subjects have seizures that appear to cluster during the day, as represented by fKH16, while others are clustered more at night as shown for fKH14. While rats are mostly nocturnal animals, their sleep cycle is notably shorter than humans, making sleep-wake correlations challenging to study.\textsuperscript{228} However, rats are more active at night, making those with daytime seizure clustering phenotypes possibly analogous to patients with nocturnal seizures. Due to the high rate of sudden death in nocturnal epilepsy, the temporal clustering of seizures remains an important area of future investigation.\textsuperscript{229}

In the seizure prediction space, we also hypothesize that with a larger dataset, greater improvements in performance will be ultimately unlocked by autoencoders. However, as shown in Supplementary Fig. A-4, in order to successfully implement these approaches for seizure detection or prediction, to the point where they outperform the feature engineering described in Chapter 1.3, will an order of magnitude more data than is in any available seizure prediction dataset, including our own. In particular, our group will endeavor to create LSTMs and STNs that perform on large, variable datasets as described in Chapter 4.3. The Wilcox lab is in the process of labeling over 10,000 seizures from hundreds of rats with multifocal epilepsy, and will partner with our group to conduct these autoencoding studies once the dataset is available. Ongoing efforts to share data between EEG labs are crucial to future progress in this field as the limited number of seizures is continually the bottleneck to performance.
One of the most clinically promising frontiers explored in this work is the continued investigation of SPES and its utility in predicting seizures. Subsequent studies should further optimize the stimulation parameters and featurization approach based on our preliminary work, and aim to investigate whether our findings hold a larger number of animals. SPES should be delivered often enough to probe the synchronicity of brain circuits before every seizure, but not so often so as to result in early seizure termination (see Fig. 5-5).

Finally, our lab and others are broadly interested in closed-loop preemption of seizures. Now that we have created predictive models that perform at clinically relevant levels of sensitivity and specificity (AUROC > 0.90), we will use our best-performing algorithms, as well as the state of the art from other groups, to preempt seizures in a closed-loop experimental setup in rodents. We will predict the seizures and intervene prior to initiation, either with high-power electrical or optogenetic stimulation.

Using a novel LED-based method being developed by colleagues Fumeaux & Moraes, we will be able to apply optogenetic approaches to parse out the circuitry involved in mTLE seizure initiation with a specific goal of targeting established excitatory generators. The first-line target to preempt seizures using direct optogenetic inhibition of the mTLE model generator should be hippocampal CA3 pyramidal neurons due to their role in the pathophysiology of mTLE.

These future experiments carried out by colleagues in the lab and collaborator groups will further our ability to predict, understand and ultimately prevent seizure initiation. This thesis provides a foundation that can be built upon via these avenues to continue improving therapies for our patients.
Appendices
A.1. Code Documentation

All software created in the course of this work is available open source here for review; it is also free for use in whole or in part under the MIT License: https://github.com/senane/ADELPHI.

Documentation for all software created in this work can be found at:

The GUI used for data viewing for Chapters 2-4 can be found at:
https://github.com/senane/ADELPHI/tree/master/EEGViewer

Analyses presented in Chapter 5 can be found at:
https://github.com/senane/ADELPHI/tree/master/Stim%20Analysis

The code for this work was written jointly by the ADELPHI Team (Awesome Detectives of ELectrographic Events Leading to PHysiologic Ictus) of the Cash lab. Contributors to ADELPHI Team code on GitHub, in addition to myself, include in alphabetical order: Aafreen Azmi, Nicolas Fumeaux, Adesh Kadambi, Constantin Krempp, Marcio Moraes, Angelique Paulk, and Emma Rogge.
A.2. Supplementary Figures

The following are Supplementary Figures referenced throughout this work.

**Supplementary Figure A-1:** Schematic of rodent epilepsy monitoring unit (EMU). We recreated the EMU found in tertiary care clinical settings for our rodent model of focal epilepsy, in order to capture all seizures by recording 24/7 for months at a time with 99.8% uptime. See KimchiLab.org/opbox for more details and specifications for components.
Supplementary Figure A-2: Hyperparameter optimization for seizure detection. The results of the hyperparameter grid search for the GLM in Chapter 2-2 are shown. Analogous optimizations were done in Chapters 4 and 5.

Supplementary Figure A-3: Diurnal seizure clustering varies between subjects. A preliminary analysis of temporal patterns of seizures in the focal epilepsy model from the main focal epilepsy dataset addressed in Chapter 2.1. Time since injection of kainic acid versus time of day of seizure plotted for five animals. Medoid seizure highlighted with cross.
Supplementary Figure A-4: Autoencoding of predictive features by a three-layer CNN.
A three-layer CNN was trained on 80% of the data sampled from four animals in the focal epilepsy
dataset with >20 seizures, and tested on a holdout 20% from those four animals. (A) Convergence
visualized as loss and accuracy over training epochs for the CNN autoencoder. (B) ROC of the training
and test set (AUROC = 0.759). (C) Convolution functions plotted from the first layer. These represent
one-dimensional filters convoluting 50 samples (plotted on x axis). (D) Representation of dataset in the
space of the first three PCs to show separability by class. An ellipsoid fit of the same data is set out,
wherein the ellipsoid dimensionality is proportional to the standard deviation of each class along each PC.
References


49. Fürbass, F. *et al.* Automatic detection of rhythmic and periodic patterns in critical care EEG based on American Clinical Neurophysiology Society (ACNS) standardized terminology.


158.Salam, M. T., Perez-Velazquez, J. L. & Genov, R. Comparative analysis of seizure control
efficacy of 5Hz and 20Hz responsive deep brain stimulation in rodent models of epilepsy. in *2015 IEEE Biomedical Circuits and Systems Conference (BioCAS)* 1–4 (2015).


188. Killick, R., Fearnhead, P. & Eckley, I. A. Optimal Detection of Changepoints With a Linear


