EEG-BASED CLASSIFICATION AND ADVANCE WARNING OF EPILEPTIC SEIZURES

THESIS

Lauren E. Clisby, 2nd Lieutenant, USAF

AFIT-ENG-MS-16-M-097

DEPARTMENT OF THE AIR FORCE
AIR UNIVERSITY

AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

AFIT-ENG-MS-16-M-097

EEG-BASED CLASSIFICATION AND ADVANCE WARNING OF EPILEPTIC SEIZURES

THESIS

Presented to the Faculty

Department of Electrical and Computer Engineering

Graduate School of Engineering and Management

Air Force Institute of Technology

Air University

Air Education and Training Command

In Partial Fulfillment of the Requirements for the

Degree of Master of Science in Operations Research

Lauren E. Clisby, BS

2nd Lieutenant, USAF

March 2016

AFIT-ENG-MS-16-M-097

EEG-BASED CLASSIFICATION AND ADVANCE WARNING OF EPILEPTIC SEIZURES

Lauren E. Clisby, BS

2nd Lieutenant, USAF

Committee Membership:

Dr. Brett Borghetti
Chair

Dr. Darryl Ahner
Member

LtCol Ryan Kappedal, PhD
Member

AFIT-ENG-MS-16-M-097

Abstract

Epilepsy is the second most common neurological disease after stroke. Epileptics may suffer hundreds of seizures per day, yet one is enough to put a person in constant fear of the next. The sudden and unexpected onset of seizures has debilitating and sometimes fatal consequences. The development of a real-time seizure prediction and alerting device would greatly improve epileptics' quality of life. Major challenges for such a device include determining predictive features and discovering the maximum prediction window.

Using the novel approach of random forest classification on EEG data, this research investigates the predictive features among the common EEG frequency bands for one patient with partial complex and partial with secondarily generalized seizures. The impact on classifier performance of labeling the transitional brain states is also investigated, using a time-series accuracy graph.

Predictive features are found as far as 40 minutes in advance of two seizures, specifically in the beta frequencies of one brain node. The random forest classifier does not perform well, but shows promise for improved performance with minor adjustments in training. The time-series accuracy graphs prove a useful tool for visualization and insight into classifier performance that is lacking in other evaluation methods.

Acknowledgments

I would like to thank my advisor for his endless patience and understanding, as well as his critical eye and insightful advice.

Thank you to my mentor for thought-provoking ideas and engaging conversation.

Of course, much appreciation goes to my roommate, for keeping my head on straight.

Lauren E. Clisby

Table of Contents

Page

List of Figures

Page

List of Tables

Page

EEG-BASED CLASSIFICATION AND

ADVANCE WARNING OF EPILEPTIC SEIZURES

I. Introduction

This research aims to improve the lives of epileptics by identifying important features to improve seizure prediction methods. Epilepsy is a neurological disease characterized by recurrent, unprovoked seizures (Fisher et al., 2005). Though few people are aware of the effects and impact of epilepsy, it is the second most common neurological disease, after stroke (Mormann, Andrzejak, Elger, & Lehnertz, 2007). Approximately sixty-five million people worldwide live with epilepsy, with at least two hundred thousand new diagnoses per year (C.U.R.E., 2015).

Within the broad umbrella of epilepsy, there is a wide span of experiences. Seizures vary in effect: from staring blankly for a few seconds (absence seizures), to collapsing in convulsions for several minutes (tonic-clonic seizures) (Mayo Clinic, 2015). Seizures may go unnoticed, even by the person experiencing them, or be extremely painful and traumatizing. Some epileptics may suffer from a single seizure in their lifetime, while others have hundreds per day.

While seizures can be fatal, a more stressful issue for many epileptics is simply never knowing when the next seizure will occur (Fisher et al., 2000). Even minor seizures can have devastating consequences if they occur at dangerous times, such as

when the person is swimming or driving a car. The unpredictability of seizures causes many epileptics to lose their driver's license and job. The life-crippling effects of epilepsy have driven neurologists to research the possible causes, develop and improve diagnosis and treatment methods, understand the physiological processes at work, and attempt to predict seizures before they occur.

The complexity and individuality of brains make seizure prediction a daunting task. An algorithm trained to one patient may be completely ineffective when applied to another patient. Researchers have found that features even vary between seizures within a single patient.

This issue with individuality has driven researchers to personalize prediction algorithms, theoretically (yet to be clinically implemented) monitoring patients for a training period before tailoring an algorithm that can monitor and predict seizures in real-time. Such a device would drastically improve the lives of epileptics, either by simply warning them of the impending seizure so they can prepare, or by preventing or reducing the seizure by administering a drug or electrical treatment (details in Appendix A).

While some algorithms have achieved high prediction accuracies, there are still many questions to address before an algorithm will be reliable enough for a clinical trial. The questions this research seeks to answer are the following:

1. What are the key spectral power features in EEG for predicting epileptic seizures?
2. How does predictive performance change when varying the length of the prediction window, which greatly influences the early warning time for the patient?
3. How does a random forest classifier compare to those used in other prediction research, particularly SVMs?

1.1 Approach

This research used electroencephalogram (EEG) data transformed into the frequency domain, using a Fast Fourier Transform (FFT). Then the signals were averaged into bins by the common EEG frequency bands: delta, theta, alpha, beta, and gamma, as well as two high frequency bands. These bands were used as features for a binary classification problem, with 1 indicating a preictal (pre-seizure) sample and 0 indicating an interictal (regular functioning) sample.

After a preliminary feature analysis through correlation, the data were used to train unique random forest classifiers for four different-length early warning windows, as well as two definitions of interictal data. The parameters for each model were tuned with the training data, using Leave-One-Out Cross Validation (LOOCV). Once tuned, classifier performance was evaluated on separate holdout seizures from the same patient. The effects of defining the preictal and interictal windows were analyzed using time-series accuracy graphs. The classifiers with the highest prediction accuracies were used for key feature analysis, by comparing the feature importance as indicated by the random forest. The classifiers with the highest accuracy were also used to determine time under false warning, for comparison to other prediction research, as well as a random predictor.

1.2 Assumptions and Limitations

The present research focuses specifically on epileptics, so findings may not be applicable to people suffering from non-epileptic seizures. Also, this analysis is limited to patients with partial complex seizures and partial seizures with secondary generalization, monitored through intracranial EEG.

One major challenge of this research was manipulating the available data to suit this study. The data were not collected for the purpose of this research, but rather recorded from patients for clinical purposes. This means all patients have individually tailored electrode configurations on their brains, rather than a standardized configuration.

Another challenge of this research was dealing with the unknown length of the preictal horizon. The times of the seizures were marked in the EEG records by epileptologists, but the length of the preictal window (the time before the seizure during which prediction is possible) is unknown. This issue was dealt with through examining four time windows, and by including time-series accuracy graphs in the evaluation.

1.3 Implications

Through identifying predictive features, testing the impact of the preictal window length, and introducing a random forest classifier to the field of research, the present study aims to both improve prediction classifiers and contribute to the basis of knowledge used to study epileptogenesis. The motivation for this research is to improve the lives of epileptics in any way possible.

The following contributions have implications for an early warning prediction and intervention device that would greatly improve the lives of epileptics.

- The results of this research show that predictive features appear as far as 40 minutes in advance of the seizure.
- The random forest classifier was shown to be viable, though requiring further investigation in a context of optimizing seizure prediction accuracy for comparison to other methods.

II. Literature Review

This chapter provides a brief overview of epileptic seizure prediction research. First, it covers the different types of EEG monitoring, then highlights important discoveries from research into the physiological factors and causes of epilepsy. Seizure detection research is discussed solely for the implications of predictive features, and then a performance comparison of current prediction algorithms builds a platform for this research. Finally, it highlights the areas available for improvement in the field that this research aims to address.

2.1 History

Researchers have been attempting to predict seizures as early as the 1970's. However, a survey paper published in 2007 showed much of the previous work in this area was falsely optimistic (Mormann et al., 2007). Several statistical validation issues caused these problems with alleged results.

Most predictors reported classification accuracies for data used in the training set, rather than a separate holdout testing set, resulting in high prediction accuracies. False positive rates were inaccurately low, because the preictal windows (when false positives are impossible) were included in the total time. Researchers also failed to compare their classifiers to a naïve or random prediction scheme. Later comparison showed several of them failed to statistically outperform the random methods. One final problem was that the data were specifically selected and segmented to represent defined interictal and preictal periods, ignoring the many varied brain states that a predictor would have to interpret during long-term, continuous monitoring.

The authors of the study proposed a set of guidelines for future prediction research to establish validity. Briefly, these rules are as follows, with the complete verbiage in Appendix B (Mormann et al., 2007).

- Algorithms should use long-term, continuous data to account for all mental states of the patient.
- Metrics should include sensitivity and specificity with respect to prediction horizon. Time under false warning should also be reported. False prediction rates only apply to the interictal period.
- Results should be tested to prove above-chance prediction rates.
- Algorithms using training data must report results using an independent test set.

2.2 Electroencephalography

EEG monitoring has been the gold standard for epilepsy research since its development by Hans Berger in 1924. EEG data are collected by placing electrodes either on the scalp (extracranial) or directly on the surface of the brain (intracranial). The low-impedance (<5 kΩ for intracranial) electrodes detect the voltages of the neurons in the brain and record them for future analysis (Adeli, Zhou, & Dadmehr, 2003).

Monitoring methods have improved vastly over the years, moving from paper print-outs requiring visual interpretation by expert neurologists, to high-resolution digital storage which can be analyzed by advance computer algorithms. Extracranial EEG is still widely used, but now surgically invasive intracranial EEG—also known as iEEG or electrocorticography (ECoG)—has been invented to monitor patients before brain surgery.

Extracranial EEG is a routine monitoring practice, often used for diagnosis and trying to determine the epileptic focus. Unfortunately for researchers, scalp EEG

includes many artifacts, caused by electrical noise, movement, or line noise (Park, Luo, Parhi, & Netoff, 2011), making it difficult to process.

ECoG requires placing electrodes directly on the surface of the brain. This method is used for monitoring epileptics before brain surgery. ECoG provides cleaner data which can be used to determine what part of the brain should be removed to try to reduce or stop seizures. The invasive procedure consequently makes this data scarce, and has only recently been performed for the purpose of data collection (Ihle et al., 2012).

When using EEG for research, there is a balance between useable data and the applicability of findings. Invasive EEG data may be more useful to researchers than scalp EEG, but the applicability of invasive EEG is debatable. A major goal of prediction research is to develop a real-time monitoring system that epileptics could wear constantly to be warned of seizures. While scalp EEG may seem like the better candidate for this type of device, intracranial monitoring may be feasible and more permanent, as the device could be surgically implanted for long-term monitoring. Such a device may even use a small electrical charge to "reset" the brain and prevent the seizure, somewhat like a pacemaker treats arrhythmias.

Epilepsy research using EEG data can be grouped into two overarching categories: application and theory. The theory category includes research into the characteristics of EEG data and possible physiological and chemical processes involved with epilepsy. Application involves constructing detection algorithms that determine when a patient is suffering from an epileptic seizure and prediction algorithms that warn of impending seizures in advance.

2.3 Theory

There has been a broad range of studies into the physiology of the brain during epileptic seizures. These studies range from observing the behavior of the gamma wave in rats during drug-induced status epilepticus (long-term continuous seizures) (Phelan, Shwe, Abramowitz, Birnbaumer, & Zheng, 2014), to theoretically modeling the epileptic brain as a dynamic system (Lopes da Silva et al., 2003; Moghim & Corne, 2014). Findings have been used to improve treatment methods.

2.3.1 Cause (Epileptogenesis) and Activity.

According to the International League Against Epilepsy (ILAE) and the International Bureau for epilepsy (IBE), an epileptic seizure is defined as "a transient occurrence of signs and/or symptoms due to abnormal excessive or synchronous neuronal activity in the brain" (Fisher et al., 2005). This definition may seem vague, but "synchronous neuronal activity" is common throughout the literature (Adeli et al., 2003; Lopes da Silva et al., 2003; Perez-Velazquez, Valiante, & Carlen, 1994), because the true cause of epilepsy remains unknown.

What researchers have theorized is that there is a physical difference between an epileptic brain and a non-epileptic brain that allows the former to produce a seizure in conditions that the latter would function unperturbed (Lopes da Silva et al., 2003). The difference may be "specific inflammatory pathways" that are routinely utilized during epileptogenesis and remain in epileptic tissue in humans with temporal lobe epilepsy (Ravizza et al., 2008).

The cerebral cortex, brainstem, and thalamocortical interactive systems are all involved in epileptogenesis (Fisher et al., 2005). The hippocampus has also been linked to the formation of seizures by (Perez-Velazquez et al., 1994), who found electrotonic coupling influenced synchronous neuronal firing, by (Vezzani et al., 2002) who induced seizures and status epilepticus in rats through stimulating the hippocampus in a variety of ways, and by (Ernfors, Bengzon, Kokaia, Persson, & Lindvall, 1991) who found an increase in trophic factors (molecules that allow a neuron to develop and maintain connections with its neighbors) in the hippocampus could promote hyperexcitability.

Other EEG signal analysis has revealed brain activity that may be used for predictive purposes. One area of study is signal synchronization between different parts of the brain, such as synchronization of Lyapunov exponents between electrode pairs and a novel Spike Synchronization Measure (L D Iasemidis, Shiau, Sackellares, Pardalos, & Prasad, 2004; Krishnan et al., 2015; L. D. Iasemidis, J. C. Principe, 2000). These studies found an increase in their respective measures of synchronization in the preictal period, and a desynchronization in the postictal period. An analysis of the power of features to distinguish between preictal and interictal using receiver operating characteristic (ROC) curves made a similar discovery that the best bivariate measures were those for phase synchronization based on the Hilbert transform, as well as maximum cross-correlation (Mormann et al., 2005).

Though these studies have found synchronization in the preictal period and desynchronization in the postictal period, the claim of their predictive quality is under debate. Previous Lyapunov studies have been invalidated due to neglect of the interictal

period during statistical validation (Mormann et al., 2007). Also, a study of modeled temporal lobe epilepsy in mice found that there is a reduction in synchronization associated with epileptic mice, but the synchronization is throughout the signal and does not correspond to epileptic events (Arabadzisz, Antal, Parpan, Emri, & Fritschy, 2005).

Identifying the root of seizures is difficult because their manifestations are so varied. Seizures can originate and occur locally to one part of the brain (partial seizures), originate locally and spread to the rest of the brain (secondary generalized seizures), or originate across the entire brain (primary generalized seizures). Suffering from multiple types of epilepsy, as well as a variety of other factors (such as brain maturity a medications) can make every epileptic event unique (Fisher et al., 2005).

2.4 Application

Beyond the studies of epileptic physiology and brain dynamics lies the application of findings that enable seizure detection and prediction. This section reviews previous research in both of these areas.

There is an important distinction between seizure detection and seizure prediction. Seizure detection involves determining if a patient is currently experiencing a seizure, early in the ictal period (distinguishing between *interictal* and *ictal*). Seizure prediction is the much more difficult task of determining if a patient will experience a seizure in the near future (within a prediction window). These methods must predict the seizure in the preictal phase, before the notable change in brain activity that is characteristic of seizure onset (distinguishing between *interictal* and *preictal*). Both tasks are being investigated

using machine-learning algorithms, for their ability to learn from complex data and make assessments in real-time.

2.4.1 Detection.

Detection algorithms are useful for the accurate diagnosis of epilepsy, which is otherwise identified by a neurologist visually inspecting EEG data for a 3-Hz spike and wave complex defined by Weir in 1965 (Adeli et al., 2003; Weir, 1965). Detection algorithms may also be used by clinics to alert staff of an epileptic event, so they can administer treatments more promptly.

Since detection algorithms have achieved 100% accuracy (Chen, 2014; Subasi & Gursoy, 2010; A. T. Tzallas, Tsipouras, & Fotiadis, 2007; A. T. Tzallas, Tsipouras, & Fotiadis, 2009). The features the algorithms use to recognize a seizure is occurring are briefly discussed here. The features used for detection may provide insight into which features are useful for prediction.

Detection researchers have used the gamut of classification techniques, including decision trees (Polat & Güneş, 2007; A. T. Tzallas et al., 2009), several types of neural nets (Ghosh-dastidar, Adeli, & Dadmehr, 2010; Srinivasan, Eswaran, & Sriraam, 2007; A. T. Tzallas et al., 2007; A. T. Tzallas et al., 2009), support vector machines (Subasi & Gursoy, 2010), logistic regression, K-nearest neighbors, and Naïve Bayes (A. T. Tzallas et al., 2009). To train these classifiers, most have used features derived from wavelet and Fourier transforms of the raw EEG data, binning by the common EEG frequency bands and calculating additional signal features such as approximate entropy (Chen, 2014; Ghosh-dastidar et al., 2010; Ocak, 2009; Polat & Güneş, 2007; Subasi & Gursoy, 2010; a

T. Tzallas et al., 2007; A. T. Tzallas et al., 2009). Some have unique approaches to the features, not transforming to the frequency domain and deriving features from the raw EEG signal in the voltage domain, rather than using a transformation (Srinivasan et al., 2007; Venkataraman et al., 2014).

These studies achieved high detection results with their features, those that used feature selection did not report which features were selected as the most valuable. This means we can only make assumptions about the values of the features by comparing results between studies. In general, the studies using Fourier and wavelet transformed features performed better than those using raw approximate entropy and maximum Lyapunov exponents; several of the former achieved 100% detection accuracy.

However, detection algorithms are not enough to give epileptics peace of mind and some of their freedom back. By the time detection algorithms send a warning, the seizure has already begun, making preparation (such as parking the car) impossible. Early detection may work with automatic stimulation to calm seizures before they take full effect and preliminary studies have shown promising results (Fountas et al., 2005; Kossoff et al., 2004; Mormann et al., 2007; Osorio et al., 2005). However, even with perfect detection, stimulation and drug regimens fail to abort seizures for some patients. These patients require a device that warns them before the seizure begins.

2.4.2 Prediction.

The first challenge of predicting seizures is proving that prediction is possible. There are two ways seizures could form: spontaneously or through a gradual build-up. If they are spontaneous, there would not be any preceding indicators to use for prediction.

However, if seizures are the passing of a threshold after a build-up period, there would be preictal features employable for prediction.

The brain may be a dynamic system, with seizures acting as a stable state in an epileptic brain. To transition to the seizure state, a number of state parameters would have to transition. The parameter transitions would be reset or ineffective in a non-epileptic brain, but in the epileptic brain a seizure occurs (Lopes da Silva et al., 2003). A detectable preseizure state may be more likely for partial epilepsy, while primary generalized may be virtually spontaneous (Mormann et al., 2007).

Several studies found preictal physiological changes that would indicate a preseizure brain state, including "an increase in cerebral blood flow (Baumgartner et al., 1998; Weinand et al., 1997), oxygen availability (Adelson et al., 1999), and blood oxygen-level-dependent signal (Federico, Abbott, Briellmann, Harvey, & Jackson, 2005) as well as changes in heart rate"(Delamont, Julu, & Jamal, 1999; Kerem & Geva, 2005; Novak, Reeves, Novak, Low, & Sharbrough, 1999; Mormann et al., 2007). The existence of a pre-seizure state, or preictal window, means seizure prediction is possible.

Since 2007, there have been several patient-specific classification algorithms used for prediction that claim promising results. Most of these studies focus on maximizing the prediction accuracy by providing a large set of features to an algorithm that performs automatic feature selection and prediction. The present research aims instead to discover which specific features are important for prediction. This focus may shed light on the processes underlying seizures and improve prediction algorithms.

Support Vector Machines (SVMs) are the most common classifiers used for prediction in recent research (Bandarabadi, Teixeira, Rasekhi, & Dourado, 2014; Moghim & Corne, 2014; Netoff, Park, & Parhi, 2009; Park et al., 2011; Williamson, Bliss, & Browne, 2011). While these classifiers are known for classifying complex data by mapping it to a higher dimension before choosing the decision boundary, this mapping makes the prediction features difficult to interpret in terms of the original data.

For the sole purpose of high prediction accuracy (disregarding feature analysis), SVMs have performed well, with one study achieving 97.5% sensitivity and a false positive rate of 0.27 per hour for a 30 minute preictal window (Park et al., 2011). This study used 5-fold cross validation to tune a patient-specific, cost-sensitive SVM. For features, this study used the common EEG bands (splitting the gamma band into four bands) processed with the Hjorth mobility parameter (a time-differential method for normalizing data). Using Kernel Fisher Discriminant analysis, they discovered the gamma features often had the top discriminability for seizure prediction. They concluded that changes in gamma band power in respect to total power may be used for seizure prediction.

Another study achieved an S1-score (the harmonic mean between sensitivity and specificity) of 96% for a 13 minute preictal window (Moghim & Corne, 2014). The features selected using the Matlab function ReliefF were not reported. In fact, most research that used preliminary feature selection did not report the feature selection results, only the results of the classifier. These studies were focused specifically on producing accurate prediction algorithms, so the features the algorithm automatically

selected and used were not pertinent to their results. In contrast, the focus of the present research is to identify important prediction features. For this case, the classifier serves only to gauge the predictive power of the features.

All of these studies used different preictal windows, including 5 minutes (Netoff et al., 2009), 30 minutes (Park et al., 2011; Williamson et al., 2011), and ranges of times (Leon D Iasemidis et al., 2003; Mirowski, Madhavan, LeCun, & Kuzniecky, 2009; Moghim & Corne, 2014; Mormann et al., 2005), with a maximum window of 4 hours. Most preictal windows tested were within one hour. Two studies that tested adjustable windows had an average window of approximately one hour (Leon D Iasemidis et al., 2003; Mirowski et al., 2009). This range of windows is due to the unknown length of the preictal window and its significant effect on prediction results.

The maximum length of this preictal window is still under debate, and likely varies between patients and seizures. Preictal indicators have been found as early as 240 minutes in advance, though other research claimed a more modest 22 minute window (Ren et al., 2011). An attempt to determine the optimal preictal window found the optimal window is unique within patients, as well as seizures, and recommended that algorithms should be trained accordingly (Bandarabadi, Rasekhi, Teixeira, & Dourado, 2014).

Due to this debate, prediction algorithms are evaluated in respect to a predetermined preictal window. A long preictal window increases overall sensitivity and specificity, but does not provide the patient with as much certainty for when the seizure will occur. Time under false warning is an important metric, because expecting a seizure

is stressful and time-consuming. For instance, if a long preictal window is chosen, the algorithm may have high accuracy, but patients will spend a large portion of their time awaiting seizures that may or may not occur. However, if the preictal window is too short, the prediction does not give the patient enough time to prepare, rendering the algorithm useless. The present research tested four prediction window times (5, 10, 20, and 40 minutes) to explore the effect on prediction accuracy.

There are divisions in the research as to what type of data and classifiers to use for prediction: univariate or multivariate features, and linear or nonlinear classification methods. A survey of past research concluded that univariate measure have shown significant performance only in node-wise, seizure-wise analysis (Mormann et al., 2007). It should be noted, however, that the research achieving 97.5% sensitivity using univariate features was published more recently than the survey paper (Park et al., 2011).

The same survey paper found that linear methods perform similarly to nonlinear methods. These findings were reinforced more recently by a study in which an SVM predicted the seizures of only 11 out of 21 patients, while a logistic regression predicted for 14 patients (Mirowski et al., 2009). However, a convoluted neural net in the same study predicted the seizures of 20 of the 21 patients. These results suggest nonlinear methods may outperform linear methods after all.

2.5 Uncharted Territory

Using a random forest classifier as a novel seizure prediction method, the present research aims to critically examine the selected predictive features. As part of the process, this research will also examine:

- The effect the length of the preictal window has on classifier performance.
- The performance of a random forest classifier for seizure prediction in comparison to classifiers common in the literature.

While SVMs are favorable in their performance on this type of data, they have the notable disadvantage of mapping the data to a higher dimensional space before selecting the decision boundary. This means the decision boundary is in terms of the higher dimensional features, rather than the original features. The disconnect makes it difficult to determine which original features were valuable for prediction, thus losing the possible inferences about epileptogenesis and applications to feature saliency.

Decision tree classifiers are a completely novel approach to seizure prediction, despite their suitability for the problem. Like SVMs, decision trees work well with high-dimensional data. They also solve quickly, which is important for a real-time monitoring and prediction device.

Decision trees have a tendency to over-fit data. Overfitting is when a classifier virtually memorizes the training data, achieving perfect training accuracy, but sacrifices generalization and prediction accuracy on the holdout testing set. Decision trees have a tendency to do this, because if the minimum size of a leaf (the resulting nodes of a split in the tree) is the default of 1, then every observation can potentially have its own leaf, unless some other limitation is placed (such as a maximum tree depth).

Overfitting can be resolved by using an ensemble classifier of decision trees, called a random forest. A random forest classifier trains multiple trees on subsets of the training data (in-bag samples), then estimate their error on the rest of the samples (out-of-bag samples). Matlab's treebagger function uses approximately 2/3 of the data as in-bag

samples. Each tree chooses a random subset, then bootstraps that sample to equal the size of the full training set. On top of the randomness of the subsets, the trees also branch randomly. This creates a forest of random trees that vote on new samples based on their out-of-bag errors. By combining multiple random trees into a voting system, random forests have all of the same advantages as decision trees, without overfitting.

III. Methodology

This chapter thoroughly explains the process by which the research questions were investigated. First, the research questions and the formulations for answering them are described. Next, the data selected and the steps necessary to process the raw iEEG data into a usable format are covered. Moving from preprocessing to preliminary feature analysis, correlation is used to examine whether any individual features have notable predictive value. Then the random forest classification method is described. The chapter ends with the evaluation metrics and expected results in terms of these metrics.

3.1 Research Objectives and Modeling

This research investigates three questions pertaining to seizure prediction in epileptics:

1. What are the key spectral power features of iEEG for predicting epileptic seizures?
2. How does predictive performance change when varying the length of the preictal window, which greatly influences the impact for the patient?
3. How does a random forest classifier compare to classifiers in other prediction research, particularly SVMs?

To answer these questions, a binary classification problem was formulated that a machine learning algorithm could use to make predictions. For every observation of data (frequency bins of EEG readings for a 5 second sample), the classifier predicts whether the patient is in the interictal (0) or preictal (1) state. If the patient is in the preictal state, that means the patient will have a seizure within the specified time period of the preictal window.

3.2 Data Description

EEG uses cranial sensors to record the electrical activity of the brain. The challenge of EEG is the lack of data, since epileptics have seizures randomly and have to be monitored for days in order to possibly record a seizure. Collecting useful and "clean" data is even more difficult.

The data for this research were obtained through an online database hosted on the IEEG-Portal (Mayo Clinic, University, & Stroke, 2014). The data are free for researchers to use to study seizures and epilepsy. The website also included a Matlab toolbox with commands for interfacing with the website and manipulating data.

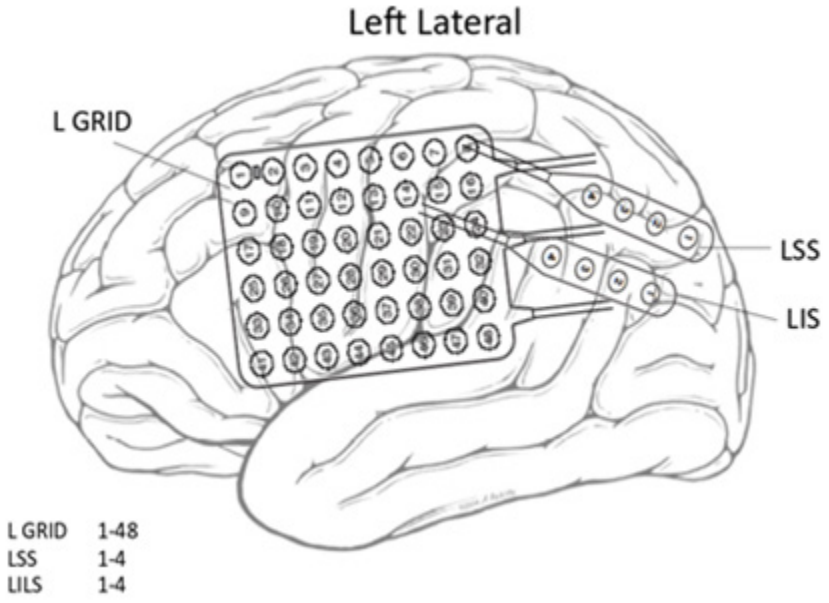


Figure 1. EEG Node Configuration. An example of the intracranial node configuration diagram provided within the documentation for a patient.

The website has over 100 patient datasets, but only 12 were suitable for this research (two of which monitored the same patient). The 11 patients were monitored for long-term, continuous, intracranial EEG. All patients were recorded at the Mayo Clinic, sampling each at either 499.907 or 500 Hz. The shortest recording is 2 days and 5 hours, and the longest is 13 days and 16 hours. The total recording time is 1,951 hours, with 172 total seizures. The number of electrodes ranges from 16 to 116 between patients. The configuration of nodes is unique to each patient and is included in their documentation pages (Figure 1). Of the 11 suitable patients, patient 017 was used for this analysis.

The documentation for each patient also details the patient's experience during monitoring, including injections and actions taken by staff, times of clinical seizure onsets (though not for all seizures), discussion of the EEG, and clinical interpretations. Note that clinical seizure onsets are the (sometimes) visible symptoms of seizure onset, while the onsets labeled in the data are the onsets determined by the EEG signal, which may occur at a different time. The following information is also available for each patient: age at onset, handedness, gender, ethnicity, seizure history, precipitants, age at admission, and contact group of electrodes. Only the raw EEG data was pertinent to the present research, so all of this extraneous information was excluded from the analysis.

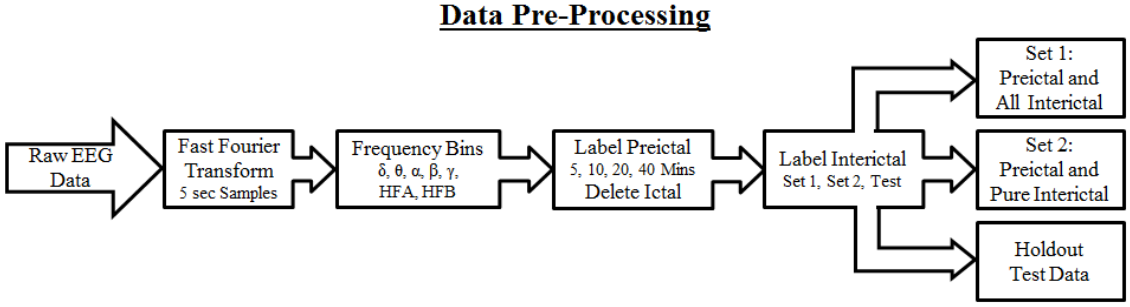


Figure 2. Pre-Processing Flowchart. The steps to process the raw EEG signals into useable datasets.

3.3 Data Pre-Processing

Before performing any analysis on the data, considerable pre-processing was necessary. The following steps (outlined in Figure 2) were necessary to process the raw EEG signal into useable datasets:

1. Transforming the data from voltages to power in the frequency domain using an FFT
2. Averaging the FFT signal into bins by the common EEG frequency bands
3. Labeling the preictal windows based on the start of the ictal labels, then deleting the ictal periods and artifacts
4. Labeling the interictal data for Set 1 and Set 2 and separating the test set

3.3.1 Fast Fourier Transform.

Raw EEG signals are time-series discharges from the patient's brain, recorded as voltages. In order to use the data for spectral power analysis, it was converted to the frequency domain using Matlab's FFT algorithm. The algorithm computes the discrete Fourier transform (DFT) of each sample. To maintain enough detail in the signal to identify important features, a time window of 5 seconds without overlap was used. This

window length is consistent with those used by others performing similar analyses (Bandarabadi, Teixeira, et al., 2014; Mirowski et al., 2009; Moghim & Corne, 2014).

For a vector *X* of length *n*, the DFT is defined as follows:

$$Y(k) = \sum_{j=1}^{n} X(j) W_n^{(j-1)(k-1)} \quad (1)$$

where

$$W_n = e^{(-2\pi i)/n}$$

is one of *n* roots of unity (Mathworks, 2015).

Since the FFT function produces an output mirrored about 0, the negative values were discarded. The amplitudes of the remaining frequencies were doubled (except the Nyquist frequency which appears only once) to preserve the total power of the signal. Each observation was then averaged into bins by the common EEG bands: delta (<4 Hz), theta (4-7 Hz), alpha (8-15 Hz), beta (16-31 Hz), and gamma (32-100 Hz), as well as two high frequency bands: high frequency A (HFA) (100-150 Hz) and HFB (150-200 Hz).

3.3.2 Labeling and Segmentation.

The data had to be segmented and labeled according to the four possible brain states in relation to seizures:

1. Interictal - normal brain functioning between seizures
2. Preictal - the period of unknown duration preceding a seizure during which the seizure is predictable (four possible windows were tested)
3. Ictal - during the seizure
4. Postictal - the period of unknown duration of altered brain function following a seizure – this period was included in the interictal data of the following seizure for one of the datasets (Figure 4)

Of these four states, only the interictal and ictal periods were annotated in the original data (preictal and postictal were also annotated as interictal), since they are directly observable in the EEG recordings (Figure 3). The timing of the transitions between the brain states are unknown, so the impact on classification accuracy of mislabeling samples was investigated by graphing the sliding window average accuracy across the monitoring period (see section 3.6 for full description).

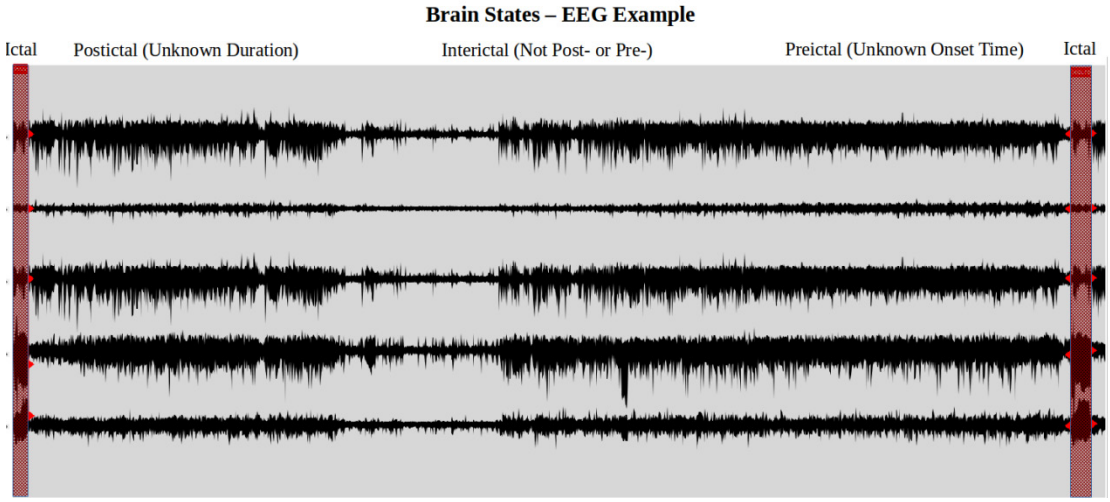


Figure 3. Brain States in EEG. An example of the four states of an epileptic brain. Each signal (5 total) represents a node on the patient's brain and the x axis is the recording time.

Since the onset of the preictal window is unknown, four window lengths concurrent with the literature were tested: 5, 10, 20, and 40 minutes (Mormann et al., 2005; Netoff et al., 2009; Park et al., 2011; Williamson et al., 2011). The 5 minute window was chosen as the common minimum warning time for an alert to be useful. The 40 minute window was chosen as both within the 1 hour windows tested in other literature, and beyond the 30 minute windows already tested. The 10 and 20 minute windows were chosen as reasonable intermediate values. Ideally, every time window

from 5 minutes to 4 hours could be tested. This would either reveal the precise beginning of the preictal window for each seizure, or more likely produce a graph of increasing prediction accuracy as the preictal window approached the seizure and the preictal features presumably strengthened. However, such a study was computationally impossible for this research, so the four selected windows were tested for a similar trend.

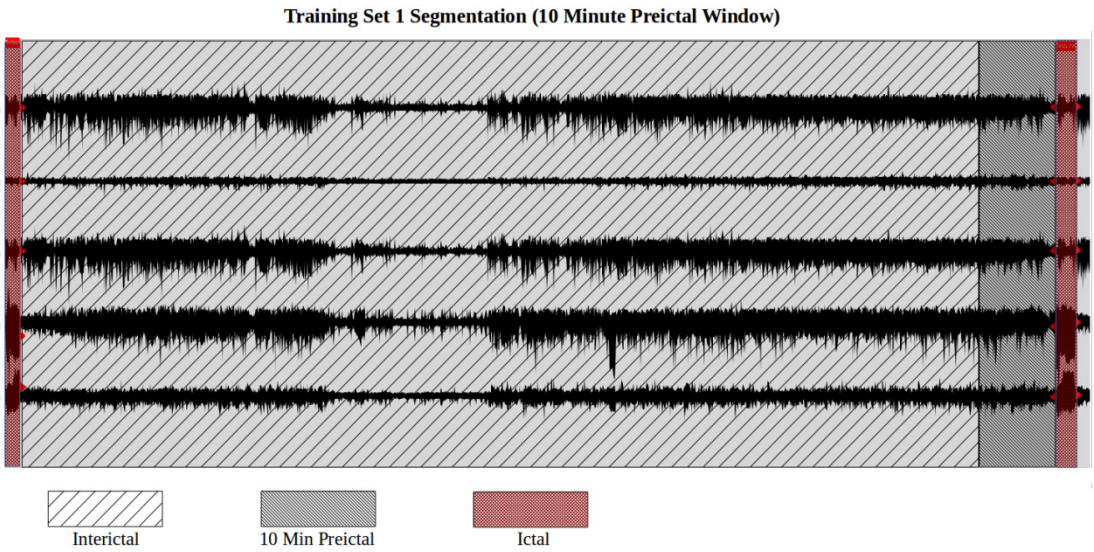


Figure 4. Training Set 1 Labels. The labels for training Set 1, using a 10 minute preictal window. Note that each preictal window labels the remainder of the data as interictal, so some observations will be labeled preictal for one window but interictal for another. Each signal is recorded from a node on the patient's brain and the x axis is the recording time.

Next, the interictal training segments were labeled. Two definitions of the interictal training data were tested (henceforth Set 1 and Set 2). Set 1 interictal data included all of the data that was not preictal (Figure 4). However, this data was expected to cause problems during machine learning training because of the unknown length of the preictal (and postictal) windows. If the preictal window was grossly ill-defined, the mislabeled samples may poorly train the classifier. Also, if the postictal data included in

the interictal class resembles the preictal class, the classifier may not be able to distinguish the classes.

To try to combat this potential issue, Set 2 interictal data was defined as the observations farthest between the end of the last seizure and the beginning of the 40 minute preictal window, sampling an equal number to the preictal observations. The observations would be considered exemplary interictal samples, because they would be as far between the preictal and postictal windows as possible (Figure 5).

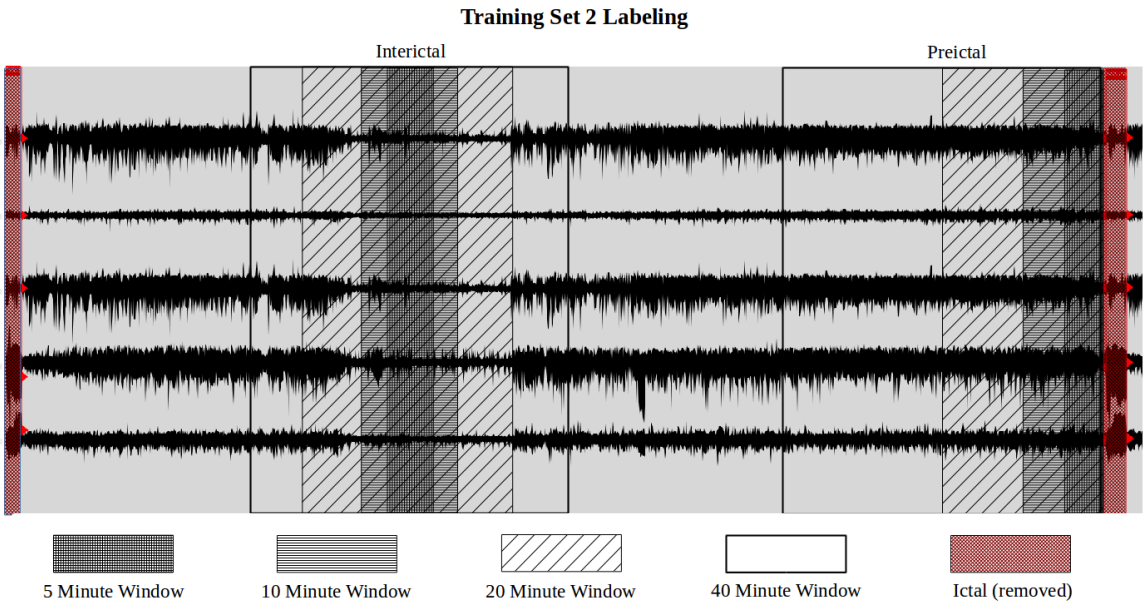


Figure 5. Training Set 2 Labels. The labels for training Set 2, for all windows (overlapped). The interictal samples are spaced as far from the end of the previous seizure and the start of the 40 minute preictal window as possible. Each signal is recorded from a node on the patient's brain and the x axis is the recording time.

The testing interictal set included all samples except ictal and preictal (the same method as Set 1, but separate seizures). This is necessary for testing, since the algorithm must handle the many varying brain states a patient may experience. To clarify, samples

within larger preictal windows were included in the interictal test data of smaller preictal windows.

Once preictal and interictal observations were binarily labeled using the ictal annotations as reference points, the ictal periods were deleted from the data to maintain a binary classification problem: 1 for preictal or 0 for interictal. (This removal was also considered valid because any classifications made during a seizure—whether positive or negative—would not be helpful to the patient during the seizure.) There were also several periods of EEG silence (all variable values of 0) which were removed.

All of this processing yielded eight separate training sets—Set 1 and Set 2 for each of the four preictal windows. For both sets, bootstrapping was used to mitigate different training issues.

Set 1 had highly unbalanced class sizes, because the interictal period was much longer than the preictal. The unequal balance could cause a problem if the classifier learns to ignore the preictal class, due to lack of training data. Uniform random sampling with replacement (bootstrapping) was used increase the preictal class size to an equal number of observations as the interictal class.

For Set 2, bootstrapping was necessary to increase the number of observations for training. Limiting the data to only twice the preictal window, multiplied by the number of training seizures, produced a negligible training set. For this data, bootstrapping was used to create a training set of 10,000 observations. The classes were approximately equal (less than 2% difference), since they were equal before the uniform sampling; this is because the resampling function in Matlab does not consider the original class ratio.

3.4 Preliminary Feature Analysis

The common EEG bands across all nodes for a patient were chosen as original features. Correlation with the binary classes was used as a preliminary feature analysis. Particular attention was paid to the high and low frequencies, because a previous study found a decrease in delta band power (with an increase in other bands) indicative of preictal EEG (Mormann et al., 2005), and several other studies have found indications among high-frequency bands (Alvarado-Rojas et al., 2014; Bandarabadi, Rasekhi, et al., 2014; Worrell et al., 2004).

If any feature had clear differences in readings between interictal and preictal samples, that feature would have high correlation (near 1) with seizure prediction. This information was not used for feature selection, but merely checking for clearly indicative features and determining possible feature saliency.

3.5 Classification

This research framed seizure prediction as a binary classification problem: preictal or interictal. Given an observation consisting of EEG spectral power readings at each node across the sample window, a classifier determined whether the patient was in a preictal state or any other (interictal or postictal) state. If a seizure then occurred within the specified prediction window, the prediction was considered a true positive. Defining the problem in this way focuses training on distinguishing between preictal and interictal, regardless of the duration of the ictal state.

A random forest classifier was chosen for nonlinear classification. This algorithm was chosen above other nonlinear methods for its ability to handle highly dimensional

data, resistance to over-fitting, speed, scalability, and most importantly the interpretability of results in terms of the input variables. Furthermore, using a random forest classifier is a novel approach to this classification problem.

Random forest is an ensemble classifier based on the decision tree classifier (Breiman, 2001). It involves building a forest of decision trees that branch nodes by random, independent sampling, rather than through a metric. The Matlab function Treebagger, is a random forest algorithm that utilizes bagging (or bootstrap aggregating) of trees to improve stability and accuracy (Mathworks, 2015). This algorithm is known for being fast and simple to implement, and robust to noise.

Treebagger has two important parameters that were adjusted to improve classification. The first parameter was the number of trees the algorithm grows before the trees vote on each observation to classify it. The training data for Set 2 was run with 10, 50, 100, 500, and 1000 trees. The Out-Of-Bag Error plot stabilized near 100 trees and maintained a comparable error for up to 1000 trees, so 100 trees were used for all further analysis.

The second parameter was the minimum number of observations per end leaf (minleaf). This parameter dictates how many observations have to be grouped together to form an end node when the tree branches. Though important when tuning a decision tree to prevent overfitting, this value is less critical for random forests, because they are already robust to overfitting. The parameter was tuned using Leave-One-Out Cross Validation (LOOCV) with the training data. Settings of 1, 2, 4, and 8 were tested, with

each model selecting the setting that yielded highest total accuracy across all LOO trials. The setting varied between models, without a trend.

Once the second parameter was tuned, the model was retrained using all of the training data. Then the model was given the test data for classification, and the tree votes for every observation were saved. By saving the votes, the Matlab function perfcurve was able to plot ROC curves. The area under each curve (a built-in metric of perfcurve) was calculated as well. The model with the simultaneous highest true positive rate and lowest false positive rate (the point on the ROC curve closest to the top left corner) was used for the evaluation metrics.

To determine which predictive features each model was using, the Matlab OOBPermutedVarDeltaError (henceforth PVDE) values for each model were stored. This Matlab output is a measure of importance for each predictor variable, calculated by permuting the values of the variable across the out-of-bag observations and measuring the increase in prediction error. After computing this measure for every tree, the values are averaged for the ensemble and reported as the PVDE output (Mathworks, 2015).

3.6 Evaluation Metrics

To evaluate the performance of the classification models, and thereby determine predictive EEG features, the classification accuracy for the holdout test set was used. Accuracy ranges from 0 to 1, with 1 indicating perfect accuracy. In a confusion matrix—with the truth classes as rows and the predicted classes as columns—perfect classification manifests as all samples falling into the diagonal.

This evaluation method differs from other prediction studies that focus on true positive rates (sensitivity), false positive rates (FPR), and the amount of time a patient spends awaiting seizures after false positive warnings (time under false warning). The present research chose accuracy instead of these other measure because of the implications for feature importance.

Studies using sensitivity, FPR, and time under false warning aim to minimize false prediction, while having at least one true prediction per seizure. That research focus completely disregards false negative classifications (misclassifying a preictal sample), as long as at least one preictal sample indicates the impending seizure. For optimizing predictions and the quality of life for the patient, these metrics make sense. However, the focus of the present research is not to develop a superior prediction method. The classification algorithms are merely used to evaluate how useful the EEG features are for distinguishing between all interictal and all preictal samples. That is why this research uses classification accuracy of a holdout test set, rather than focusing on minimizing false positives and disregarding false negatives.

The primary method of evaluation for the models was time-series accuracy graphs. These graphs show a sliding window of the prediction accuracy for the preceding five minutes of observations. (Every five seconds, it shows the accuracy of the past five minutes.) A five minute window was chosen, because it smoothed the accuracy to a point that the graph was interpretable, without smoothing away too much performance detail. Ideally, these graphs would show 100% accuracy at all points in time. However, this was infeasible, so the goals were to minimize misclassifications in the interictal span

to avoid false warnings, and maximize positive classifications during the preictal, to show the seizure was predicted. These graphs were particularly important for analysis, because a classifier could potentially show high total accuracy by classifying all samples as interictal, due to the extreme difference in class sizes. With these graphs, such an algorithm would clearly show low accuracy for the preictal window. In short, these graphs visualized the changes in classification accuracy in relation to the brain states.

Since the preictal and interictal training sets were sampled from as far away from each other as possible in Set 2, the classification accuracy was expected to diminish in the brain state transitional periods. If the assumption were correct, the time-series accuracy graphs would show peaks in the purely interictal and preictal areas, and decreased accuracy in the transitional areas. These graphs may also provide insight into the true length of the preictal window, as the accuracy near and within the preictal window would show the true state of the EEG. For instance, if the window is defined too short, there would be many misclassified interictal samples preceding the window, since they would be "misclassified" as preictal (because that is what they should actually be labeled).

In order to compare this research to other prediction algorithms, approximate time-under-warning graphs were created to show how much time a patient would spend awaiting seizures that may or may not occur. To create these graphs, it was assumed that once a warning is issued, another warning cannot be issued for the length of the preictal window. For comparison, the number of true positive classifications within a time span necessary for a warning was increased, since this reduces the time under false warning.

Ideally, though there may be false positives, the patient would never be alerted of a seizure unless one was going to occur within the specified preictal window.

3.7 Summary

Raw EEG data were transformed to the frequency domain using an FFT and averaged into common bands by frequency. The data were segmented into two sets of training data with different interictal definitions, and a testing set. Each set included four separate preictal window definitions: 5, 10, 20, and 40 minutes. Testing each of the four selected preictal windows and two interictal sets individually, a unique random forest classification model was trained to the patient. LOOCV was used to train the minleaf parameter, which reduces the chance of overfitting. Then the tuned classifiers were used to predict two holdout test seizures from the same patient. Performance was analyzed with the prediction accuracy of the test set, ROC curves, and time-series accuracy graphs.

IV. Analysis and Results

These results are for Patient 017, who suffers from two types of epileptic seizures: partial complex and partial with secondary generalization. The patient was monitored with 16 nodes for 7 days and 17 hours. The patient had nine seizures during that time, seven of which were used for training (including LOOCV) and two were used as the holdout test set.

4.1 Preliminary Feature Analysis

The correlation coefficients for the features using the 5 minute preictal window of Set 1 are shown in Table 1, with the rest of the windows included in Appendix C. Considering correlation coefficients range between -1 and 1—with -1 indicating a perfect inverse linear relationship and 1 indicating a perfect direct linear relationship—the negligible magnitudes of all of the coefficients imply that none of the features are individually linearly connected with the classes.

To compare across training set and preictal windows, the graphs were conditionally formatted according to coefficient magnitude, with the largest magnitudes (whether positive or negative) shading to dark grey, while values close to 0 were white. Note that for this discussion, the numbers in the tables may be illegible, but they are not as relevant as the shades. Shading the tables shows that limiting the samples to the "pure" data of Set 2 slightly increases the correlation coefficient magnitudes (Figure 6).

Shading the tables also shows that the somewhat predictive features are consistent across the varying seizure windows for Set 2, and the strength of the correlation increases

with the wider preictal windows. The increase in correlation is reassuring that the 40 minute preictal window is not too long. Though the coefficients are small, the shades show that gamma, HFA, and HFB have the largest magnitude coefficients. This is consistent with the previously discussed literature, which found high frequency bands to be more predictive than low frequency bands. Note also, that some of the least indicative features (white) are found within the same bands, but on different nodes.

Overall, none of the coefficients are large enough to have real meaning. This preliminary analysis has shown that there is not a substantial linear relationship between any singular feature and the classes that could be used for prediction.

Table 1. Correlation Coefficients

Set 1 – 5 Min Window - Correlation Coefficients

Node	Delta	Theta	Alpha	Beta	Gamma	HFA	HFB
1	-0.00110	-0.00006	0.00069	0.00342	0.00351	0.00116	0.00045
2	-0.00096	-0.00179	-0.00281	-0.00352	-0.00248	-0.00028	-0.00068
3	-0.00016	-0.00087	-0.00074	-0.00105	-0.00147	-0.00157	-0.00169
4	-0.00013	-0.00086	-0.00086	-0.00113	-0.00183	-0.00176	-0.00173
5	-0.00042	-0.00092	-0.00097	-0.00132	-0.00210	-0.00164	-0.00167
6	-0.00092	-0.00119	-0.00130	-0.00166	-0.00237	-0.00260	-0.00222
7	-0.00085	-0.00121	-0.00142	-0.00159	-0.00188	-0.00172	-0.00158
8	-0.00101	-0.00024	0.00075	0.00313	0.00350	0.00112	0.00035
9	0.00073	0.00098	0.00381	0.00498	0.00463	0.00224	0.00086
10	-0.00065	-0.00072	0.00156	0.00490	0.00282	0.00390	0.00224
11	-0.00078	-0.00065	0.00152	0.00526	0.00551	0.00383	0.00290
12	-0.00062	-0.00061	-0.00003	0.00095	0.00052	0.00127	0.00057
13	-0.00074	0.00006	0.00078	0.00283	0.00287	0.00064	-0.00069
14	-0.00039	-0.00078	-0.00045	0.00009	-0.00078	-0.00184	-0.00203
15	-0.00021	-0.00054	-0.00042	-0.00073	-0.00281	-0.00239	-0.00225
16	-0.00029	-0.00076	-0.00064	-0.00082	-0.00344	-0.00281	-0.00234

Set 1 – 5 M in Window - Correlation Coefficients

Node	Delta	Theta	Alpha	Beta	Gamma	HFA	HFB
1	-0.00110	-0.00006	0.00069	0.00342	0.00351	0.00116	0.00045
2	-0.00096	-0.00179	-0.00281	-0.00352	-0.00248	-0.00028	-0.00068
3	-0.00016	-0.00087	-0.00074	-0.00105	-0.00147	-0.00157	-0.00169
4	-0.00013	-0.00086	-0.00086	-0.00113	-0.00183	-0.00176	-0.00173
5	-0.00042	-0.00092	-0.00097	-0.00132	-0.00210	-0.00164	-0.00167
6	-0.00092	-0.00119	-0.00130	-0.00166	-0.00237	-0.00260	-0.00222
7	-0.00085	-0.00121	-0.00142	-0.00159	-0.00188	-0.00172	-0.00158
8	-0.00101	-0.00024	0.00075	0.00313	0.00350	0.00112	0.00035
9	0.00073	0.00098	0.00381	0.00498	0.00463	0.00224	0.00086
10	-0.00065	-0.00072	0.00156	0.00490	0.00282	0.00390	0.00224
11	-0.00078	-0.00065	0.00152	0.00526	0.00551	0.00383	0.00290
12	-0.00062	-0.00061	-0.00003	0.00095	0.00052	0.00127	0.00057
13	-0.00074	0.00006	0.00078	0.00283	0.00267	0.00064	-0.00069
14	-0.00039	-0.00078	-0.00045	0.00009	-0.00078	-0.00184	-0.00203
15	-0.00021	-0.00054	-0.00042	-0.00073	-0.00281	-0.00239	-0.00225
16	-0.00029	-0.00076	-0.00064	-0.00082	-0.00344	-0.00281	-0.00234

Set 2 – 5 M in Window - Correlation Coefficients

Node	Delta	Theta	Alpha	Beta	Gamma	HFA	HFB
1	-0.09325	-0.08534	-0.10194	-0.10877	-0.09902	-0.08175	-0.11390
2	-0.09419	-0.10077	-0.11539	-0.10156	-0.11163	-0.11292	-0.10664
3	-0.08854	-0.09316	-0.10428	-0.10817	-0.11948	-0.11805	-0.10097
4	-0.08851	-0.08856	-0.09758	-0.10339	-0.09868	-0.10264	-0.08588
5	-0.08762	-0.08322	-0.09497	-0.10380	-0.06253	-0.08804	-0.08395
6	-0.08957	-0.08047	-0.09422	-0.10235	-0.09104	-0.08579	-0.09203
7	-0.08801	-0.07738	-0.08588	-0.09633	-0.08503	-0.07852	-0.08669
8	-0.09241	-0.08530	-0.10209	-0.10894	-0.09908	-0.08208	-0.11439
9	-0.05660	-0.07319	-0.05423	-0.08685	-0.06523	-0.04375	-0.04100
10	-0.08940	-0.07634	-0.06680	-0.05618	-0.03215	-0.00220	-0.00552
11	-0.08871	-0.05146	-0.02233	-0.03616	-0.02686	0.00182	-0.06659
12	-0.08910	-0.08618	-0.08878	-0.07937	-0.07821	-0.02691	-0.05058
13	-0.08872	-0.08954	-0.08929	-0.09037	-0.07953	-0.07185	-0.10081
14	-0.08865	-0.08488	-0.10543	-0.11196	-0.11642	-0.13320	-0.14236
15	-0.08806	-0.08314	-0.09244	-0.10060	-0.08335	-0.10886	-0.10568
16	-0.08373	-0.07875	-0.08962	-0.09537	-0.06567	-0.09602	-0.09789

Set 1 – 10 M in Window - Correlation Coefficients

Node	Delta	Theta	Alpha	Beta	Gamma	HFA	HFB
1	-0.00142	0.00031	0.00141	0.00630	0.00704	0.00246	0.00170
2	-0.00124	-0.00240	-0.00408	-0.00515	-0.00374	-0.00033	-0.00112
3	-0.00003	-0.00093	-0.00082	-0.00086	-0.00216	-0.00238	-0.00268
4	-0.00011	-0.00107	-0.00117	-0.00144	-0.00303	-0.00279	-0.00288
5	-0.00054	-0.00121	-0.00137	-0.00181	-0.00365	-0.00272	-0.00279
6	-0.00143	-0.00166	-0.00186	-0.00238	-0.00390	-0.00411	-0.00362
7	-0.00130	-0.00170	-0.00205	-0.00226	-0.00308	-0.00271	-0.00257
8	-0.00139	0.00004	0.00157	0.00579	0.00702	0.00237	0.00150
9	0.00035	0.00106	0.00483	0.00740	0.00776	0.00405	0.00130
10	-0.00108	-0.00124	0.00182	0.00692	0.00394	0.00522	0.00315
11	-0.00118	-0.00066	0.00233	0.00790	0.01029	0.00550	0.00529
12	-0.00098	-0.00096	-0.00021	0.00128	0.00049	0.00069	-0.00015
13	-0.00118	0.00022	0.00110	0.00390	0.00339	-0.00024	-0.00197
14	-0.00067	-0.00101	-0.00064	0.00013	-0.00145	-0.00297	-0.00329
15	-0.00041	-0.00083	-0.00063	-0.00104	-0.00411	-0.00370	-0.00363
16	-0.00075	-0.00118	-0.00091	-0.00117	-0.00478	-0.00443	-0.00382

Set 2 – 10 M in Window - Correlation Coefficients

Node	Delta	Theta	Alpha	Beta	Gamma	HFA	HFB
1	-0.08651	-0.07207	-0.07903	-0.06344	-0.05639	-0.05918	-0.07676
2	-0.08766	-0.09485	-0.10975	-0.10718	-0.12186	-0.09927	-0.10902
3	-0.08115	-0.08303	-0.09080	-0.09253	-0.12823	-0.12229	-0.11669
4	-0.08265	-0.08149	-0.08926	-0.09553	-0.12477	-0.11984	-0.11321
5	-0.08239	-0.07767	-0.08683	-0.09808	-0.10015	-0.10947	-0.10921
6	-0.08739	-0.07657	-0.08586	-0.09383	-0.11711	-0.10337	-0.10135
7	-0.08564	-0.07387	-0.08108	-0.08885	-0.10899	-0.09560	-0.10535
8	-0.08725	-0.07337	-0.07999	-0.06454	-0.05798	-0.06228	-0.08032
9	-0.06567	-0.07421	-0.04665	-0.06226	-0.02319	-0.01608	-0.03469
10	-0.08567	-0.07892	-0.05855	-0.03877	-0.01742	0.00747	0.00139
11	-0.08544	-0.04361	-0.00725	-0.02463	-0.00602	0.00169	-0.03881
12	-0.08655	-0.08284	-0.07686	-0.06965	-0.07933	-0.05647	-0.08156
13	-0.08661	-0.07753	-0.08211	-0.08535	-0.09147	-0.10529	-0.12986
14	-0.08614	-0.08183	-0.09843	-0.10486	-0.12618	-0.13529	-0.14620
15	-0.08456	-0.07962	-0.08404	-0.09096	-0.10203	-0.11599	-0.12011
16	-0.08165	-0.07545	-0.07969	-0.08457	-0.08759	-0.10940	-0.11289

Set 1 – 20 M in Window - Correlation Coefficients

Node	Delta	Theta	Alpha	Beta	Gamma	HFA	HFB
1	-0.00218	0.00109	0.00288	0.01062	0.01315	0.00536	0.00416
2	-0.00197	-0.00331	-0.00577	-0.00615	-0.00382	0.00058	-0.00084
3	-0.00027	-0.00140	-0.00149	-0.00111	-0.00341	-0.00363	-0.00439
4	-0.00028	-0.00172	-0.00207	-0.00216	-0.00518	-0.00446	-0.00486
5	-0.00087	-0.00186	-0.00232	-0.00277	-0.00639	-0.00444	-0.00470
6	-0.00220	-0.00245	-0.00285	-0.00359	-0.00661	-0.00628	-0.00559
7	-0.00194	-0.00250	-0.00305	-0.00337	-0.00494	-0.00418	-0.00405
8	-0.00207	0.00073	0.00329	0.00982	0.01320	0.00533	0.00390
9	-0.00034	0.00070	0.00498	0.00998	0.01112	0.00552	0.00132
10	-0.00182	-0.00221	0.00139	0.00871	0.00494	0.00796	0.00510
11	-0.00179	-0.00104	0.00306	0.01111	0.01663	0.00822	0.00897
12	-0.00150	-0.00165	-0.00108	0.00122	0.00073	0.00087	-0.00048
13	-0.00179	0.00012	0.00138	0.00533	0.00579	-0.00004	-0.00261
14	-0.00104	-0.00158	-0.00113	-0.00007	-0.00233	-0.00446	-0.00520
15	-0.00068	-0.00144	-0.00128	-0.00167	-0.00658	-0.00565	-0.00587
16	-0.00128	-0.00188	-0.00150	-0.00178	-0.00776	-0.00680	-0.00613

Set 2 – 20 M in Window - Correlation Coefficients

Node	Delta	Theta	Alpha	Beta	Gamma	HFA	HFB
1	-0.06979	-0.02765	-0.01866	-0.00009	0.01622	-0.00324	-0.02071
2	-0.07086	-0.08042	-0.10766	-0.07350	-0.06984	-0.04319	-0.05649
3	-0.06221	-0.06262	-0.08187	-0.06545	-0.12563	-0.10081	-0.10632
4	-0.06328	-0.06398	-0.08146	-0.07552	-0.14533	-0.11191	-0.11223
5	-0.06344	-0.05941	-0.07868	-0.08179	-0.14309	-0.10954	-0.10795
6	-0.07050	-0.06274	-0.07804	-0.08228	-0.15794	-0.11398	-0.09993
7	-0.06926	-0.06089	-0.07384	-0.07882	-0.14204	-0.10182	-0.10514
8	-0.06981	-0.02842	-0.01915	-0.00055	0.01547	-0.00472	-0.02219
9	-0.05911	-0.07715	-0.06704	-0.03005	0.00986	0.00388	-0.01637
10	-0.07064	-0.07361	-0.07063	-0.01784	0.00483	0.01569	0.01041
11	-0.06879	-0.03036	0.00625	-0.00168	0.01886	0.01033	0.00331
12	-0.06986	-0.07516	-0.08209	-0.06201	-0.06623	-0.05409	-0.07805
13	-0.06995	-0.06666	-0.07513	-0.06495	-0.07181	-0.09570	-0.11734
14	-0.06955	-0.07000	-0.09070	-0.09332	-0.13326	-0.12491	-0.14106
15	-0.06686	-0.06750	-0.07556	-0.07242	-0.12819	-0.10761	-0.11769
16	-0.06607	-0.06217	-0.06428	-0.06081	-0.11747	-0.11249	-0.11399

Set 1 – 40 M in Window - Correlation Coefficients

Node	Delta	Theta	Alpha	Beta	Gamma	HFA	HFB
1	-0.00337	0.00075	0.00299	0.01219	0.01763	0.00813	0.00546
2	-0.00313	-0.00497	-0.00847	-0.00894	-0.00522	0.00134	0.00010
3	-0.00085	-0.00310	-0.00320	-0.00323	-0.00587	-0.00580	-0.00707
4	-0.00084	-0.00346	-0.00401	-0.00436	-0.00877	-0.00694	-0.00770
5	-0.00165	-0.00359	-0.00430	-0.00525	-0.01069	-0.00680	-0.00738
6	-0.00331	-0.00383	-0.00448	-0.00572	-0.00976	-0.00902	-0.00829
7	-0.00284	-0.00380	-0.00467	-0.00524	-0.00730	-0.00611	-0.00598
8	-0.00315	0.00024	0.00334	0.01122	0.01770	0.00813	0.00511
9	-0.00091	0.00073	0.00732	0.01042	0.01238	0.00514	0.00044
10	-0.00281	-0.00347	0.00163	0.01029	0.00666	0.01303	0.01082
11	-0.00258	-0.00198	0.00375	0.01541	0.02810	0.01703	0.01869
12	-0.00232	-0.00271	-0.00203	0.00086	0.00102	0.00625	0.00417
13	-0.00280	-0.00050	0.00145	0.00573	0.00870	0.00217	-0.00187
14	-0.00159	-0.00266	-0.00221	-0.00161	-0.00428	-0.00645	-0.00776
15	-0.00123	-0.00281	-0.00300	-0.00391	-0.01102	-0.00864	-0.00915
16	-0.00235	-0.00334	-0.00312	-0.00416	-0.01302	-0.01018	-0.00946

Set 2 – 40 M in Window - Correlation Coefficients

Node	Delta	Theta	Alpha	Beta	Gamma	HFA	HFB
1	-0.05793	-0.03326	-0.03039	-0.02197	0.01364	0.00950	-0.01563
2	-0.05995	-0.08682	-0.11636	-0.08047	-0.05569	-0.02368	-0.01938
3	-0.05062	-0.07066	-0.08987	-0.07669	-0.12977	-0.09801	-0.10905
4	-0.05034	-0.06736	-0.08826	-0.08359	-0.16207	-0.11201	-0.11566
5	-0.05171	-0.06303	-0.08357	-0.08770	-0.16001	-0.10661	-0.10972
6	-0.05656	-0.05900	-0.07533	-0.07902	-0.15153	-0.10099	-0.09264
7	-0.05527	-0.05440	-0.07065	-0.07478	-0.14309	-0.09559	-0.09932
8	-0.05736	-0.03360	-0.03071	-0.02221	0.01318	0.00856	-0.01668
9	-0.04803	-0.06917	-0.03701	-0.05593	-0.02084	-0.03140	-0.03485
10	-0.05796	-0.07196	-0.05308	-0.02795	0.02101	0.03802	0.04820
11	-0.05384	-0.03860	-0.00231	-0.00176	0.03707	0.04884	0.04457
12	-0.05616	-0.07366	-0.07316	-0.05972	-0.04671	0.00922	0.00152
13	-0.05835	-0.07890	-0.08232	-0.08633	-0.06409	-0.04458	-0.06429
14	-0.05458	-0.06553	-0.08881	-0.10315	-0.14068	-0.11176	-0.13015
15	-0.05269	-0.07102	-0.08148	-0.08040	-0.13623	-0.10616	-0.11948
16	-0.05426	-0.06358	-0.06995	-0.07021	-0.12233	-0.10586	-0.11200

Figure 6. Shaded Correlation Tables. Shading the correlation tables shows the increase in correlation when the data are restricted to "pure" samples. The second column also shows the same features darkened across the four windows. Though the values are small, this trend may indicate predictive features. The two columns correspond to Set 1 and Set 2; the rows are the preictal windows—5, 10, 20, and 40—from top to bottom.

4.2 Classification

The results of the random forest classifier will be discussed first by the sets individually, discussing the effects of the lengths of the preictal window. Then Set 1 and Set 2 will be compared to determine the effect of training on different definitions of interictal samples.

4.1.1 Training Set 1 – All Interictal.

Recall that the data for Set 1 were sampled from the entire interictal period, regardless of the unknown transition times between postictal, interictal, and preictal states. Due to the drastic unbalancing of the data, the preictal observations were bootstrapped to equal the number of interictal observations.

The ROC curves in Figure 7—produced by varying the preictal vote threshold on the training set—show that both the true and false positive rates are stunted for the shorter preictal windows. Not until the 20 and 40 minute windows do the graphs begin to show variation depending on the threshold for preictal classifications. For the shorter preictal windows, even setting the threshold as low as requiring only 10% of the trees to classify as preictal, the vast majority of samples are classified as interictal.

The reason for this lack of preictal presence to the classifier may stem from two possible sources. The first issue may be the lack of unique preictal observations. Though the data were bootstrapped to equal the number of interictal observations, bootstrapping does not create any new information. The case may be that the classifier has not seen enough unique preictal data to recognize new samples, necessary for generalizing on the left-out training sample.

A more likely cause of this lack of preictal classifications is the incorrect definition of the preictal window. If the observations in the 20 and 40 minute preictal windows resemble those in the five minute window, labeling the larger windows as interictal for the smaller window datasets would contradict what the classifier was learning as preictal samples. This explanation is supported by the earlier discussed trend in feature importance observed in the correlation tables.

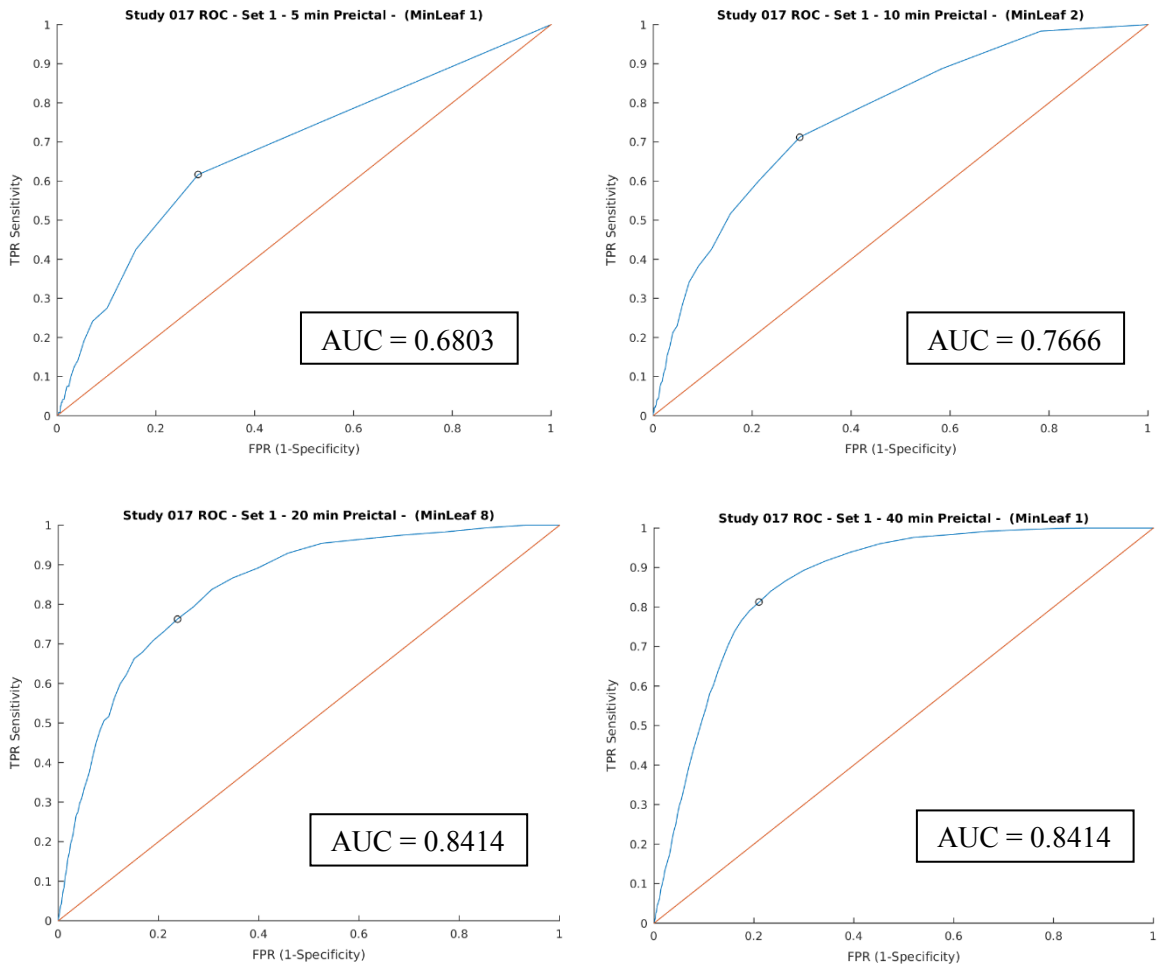


Figure 7. Set 1 ROC Curves. From left to right and top to bottom, the ROC curves for the 5, 10, 20, and 40 minute preictal windows. The area under each curve (AUC) is also shown.

Figure 8 shows the time-series accuracy graphs for both test seizures, with a 5 minute preictal window. The graphs show only the last 4 hours of data before the seizure. (All of the full time-series accuracy graphs are included in Appendix D and Appendix E, but for most analysis the 4 hour intervals will be examined.) Note the grey box highlighting the defined preictal window.

Figure 8 illustrates the fundamental issue with reporting solely the classification accuracy. The classifier performs relatively well (particularly well on Seizure 3), but completely fails at predicting the seizure. The line sloping immediately to 0 upon reaching the preictal window shows that the classifier was classifying the majority of samples as interictal. Although the line shows a steep decline upon reaching the window, it must be noted that the graph is a sliding average, which means the slope is only due to averaging the accuracy from previous observations (including the correct interictal).

This classifier has not trained to recognize the preictal class. The results are concurrent with the ROC curve, which implied all samples are classified as interictal, most likely due to a poor definition of the preictal window.

The 10 minute preictal window has similar results (Figure 9). Compared to the 5 minute window, the 10 minute model classifies more samples as preictal; however, it still fails to recognize the true preictal samples. The same plummet in accuracy occurs at the start of the preictal window. In general, the graph is less extreme than the 5 minute graphs: the accuracy does not start as high or plummet as low. The classifier is paying more attention to the importance of the preictal class, however, it still does cannot distinguish the two.

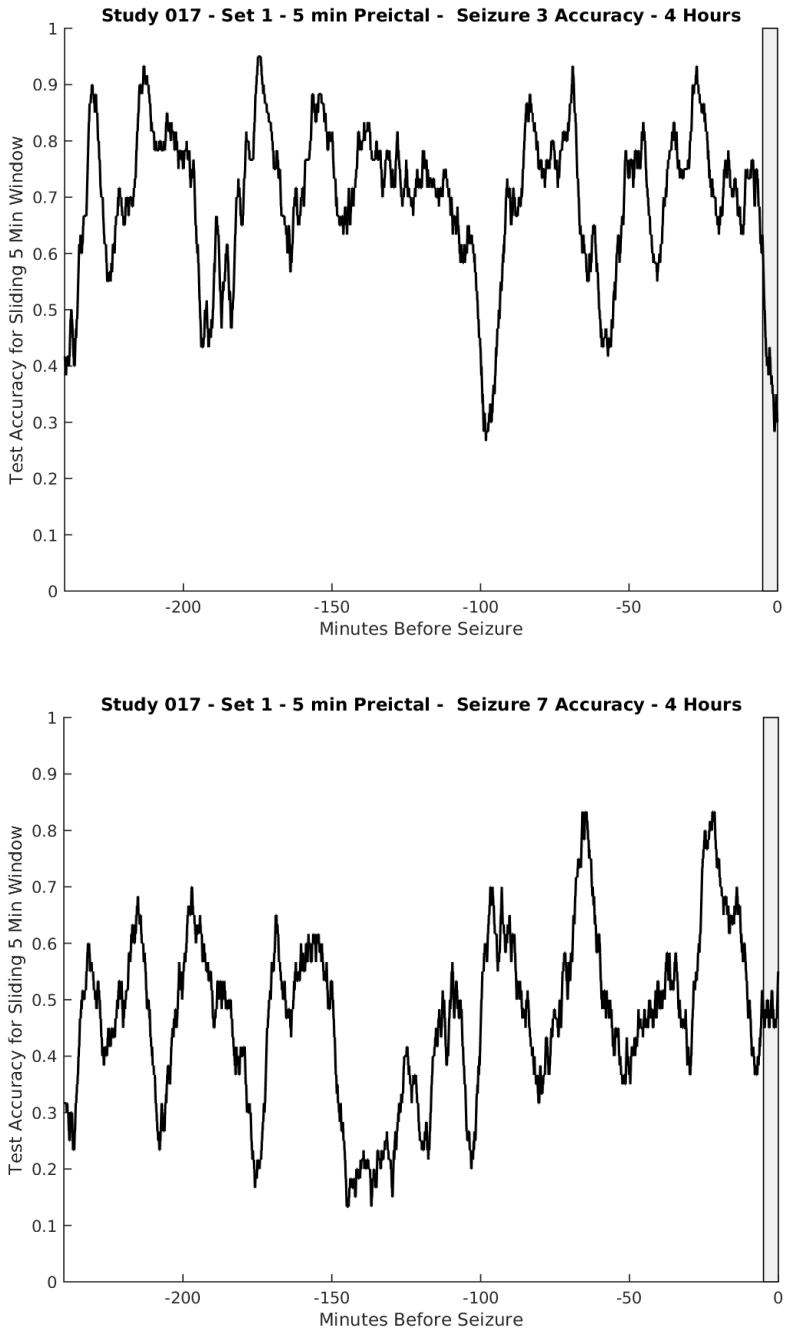


Figure 8. Set 1 - 5 Minute Accuracy. The 5-minute moving window accuracy for the last 4 hours before both test seizures. The grey boxes on the right mark the preictal window.

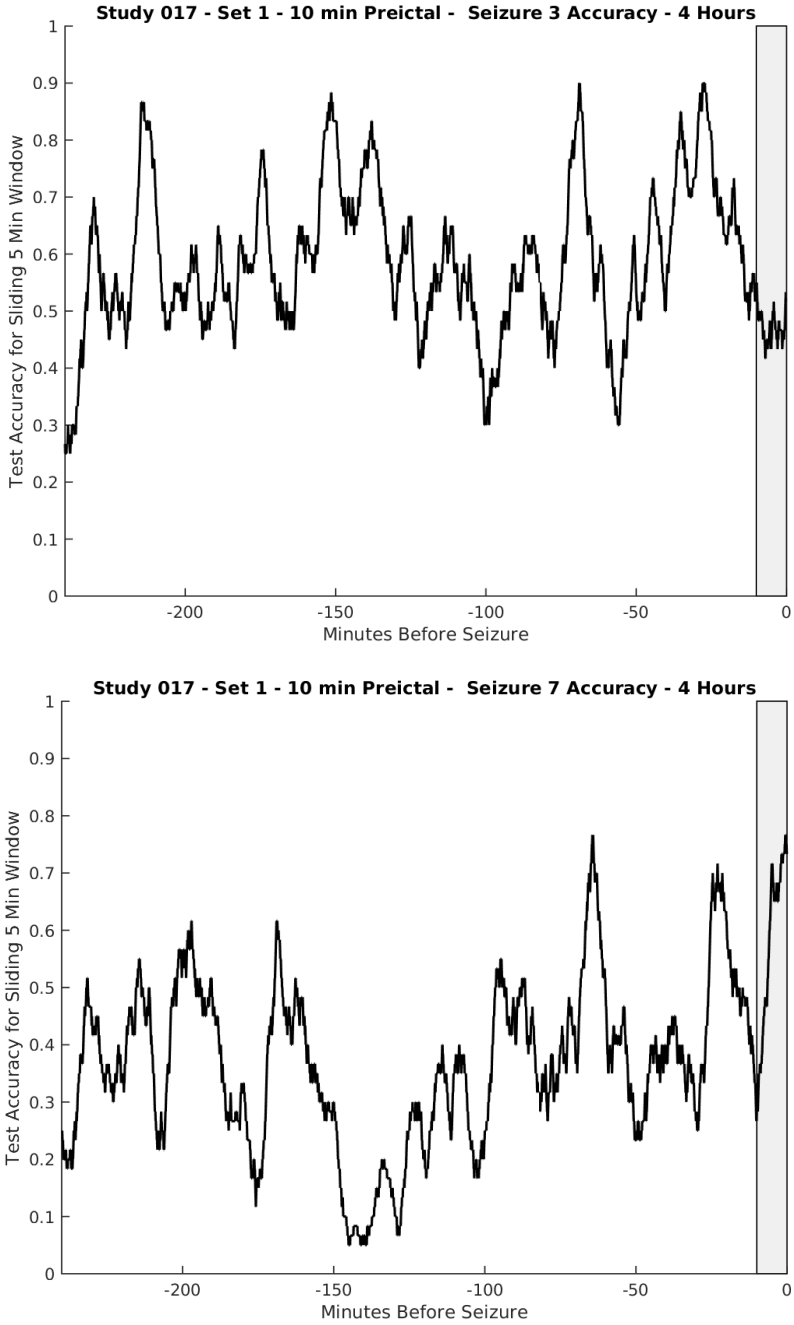


Figure 9. Set 1 - 10 Minute Accuracy. The 5-minute moving window accuracy for the last 4 hours before both test seizures. The grey boxes on the right mark the preictal window.

When trained using a 20 minute preictal window, the classifier begins favoring preictal samples as the dominant class (despite the truth being the opposite). Figure 10 shows that for one seizure the model predicts near randomly, with a slight preference for preictal, as shown by the below average interictal classification and increase in accuracy at the start of the preictal window. For the second seizure, the model has completely changed strategy from the shorter preictal windows. For the 5 and 10 minute windows, the model voted for interictal the majority of the time, but for the 20 minute window it predicts preictal the majority of the time. While the classifier predicts the seizure, it would be a poor early warning device, because the patient would constantly be awaiting seizures which would not occur.

The 40 minute model pushes even farther to the preictal class for the 4 hour span (Figure 11). This shift for the 20 and 40 minute windows is likely due to the preictal windows being too long. Labeling a number of interictal samples as preictal teaches the classifier that both classes resemble preictal data.

Although there are issues in the 4 hours preceding the seizures, the 20 and 40 minute preictal windows perform considerably well over the entire timespan for Seizure 3 (Figure 12). The graphs exhibit the same trends in accuracy, including two large dips in the second half of the timespan. These dips may be artifacts that resemble preictal data or correspond to a particular function the patient is performing during that time (such as sleeping). Excluding the major dip between -2000 and -1000 minutes, the performance nearly resembles what was expected for a time-series accuracy graph: lower accuracy at the beginning of the recording, due to the postictal period, then relatively high accuracy

for the clearly interictal period, and finishing with an area of questionable accuracy caused by the definition of the preictal window. Note the critical difference between these relatively high accuracy graphs and the high accuracy of the 5 and 10 minute windows is that the larger windows have high accuracy during the preictal window.

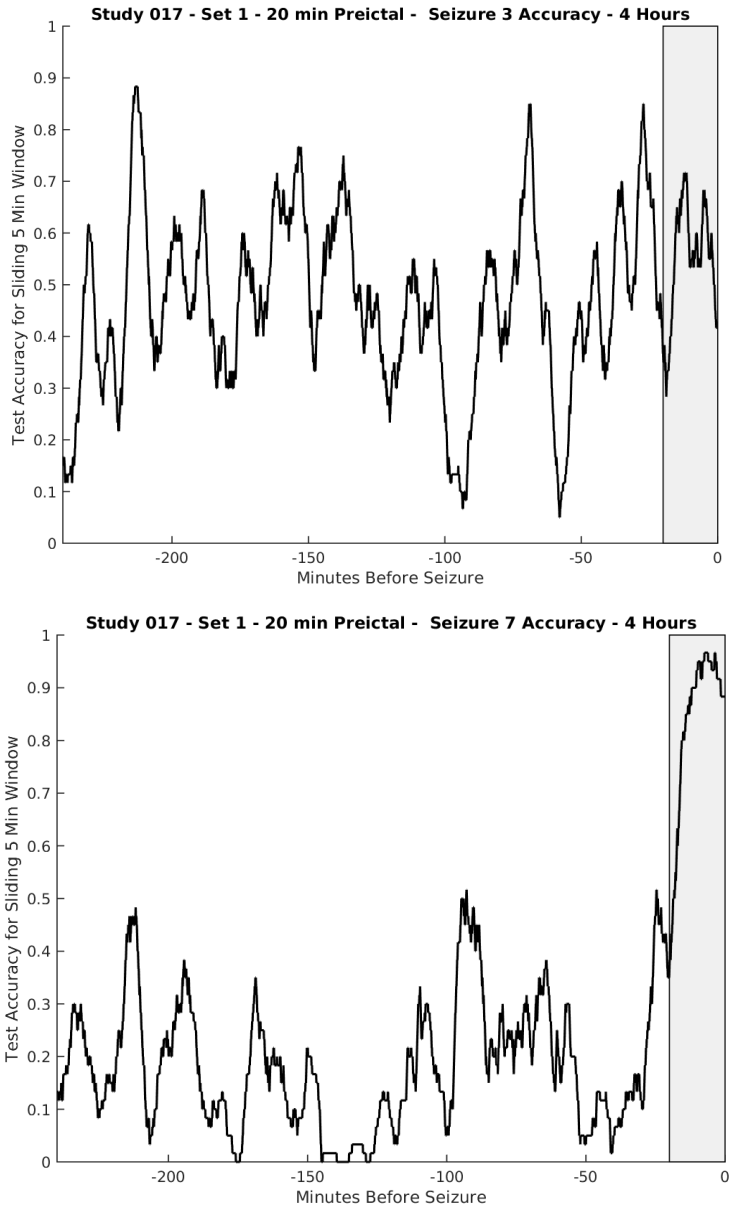


Figure 10. Set 1 - 20 Minute Accuracy. The 5-minute moving window accuracy for the last 4 hours before both test seizures. The grey boxes on the right mark the preictal window.

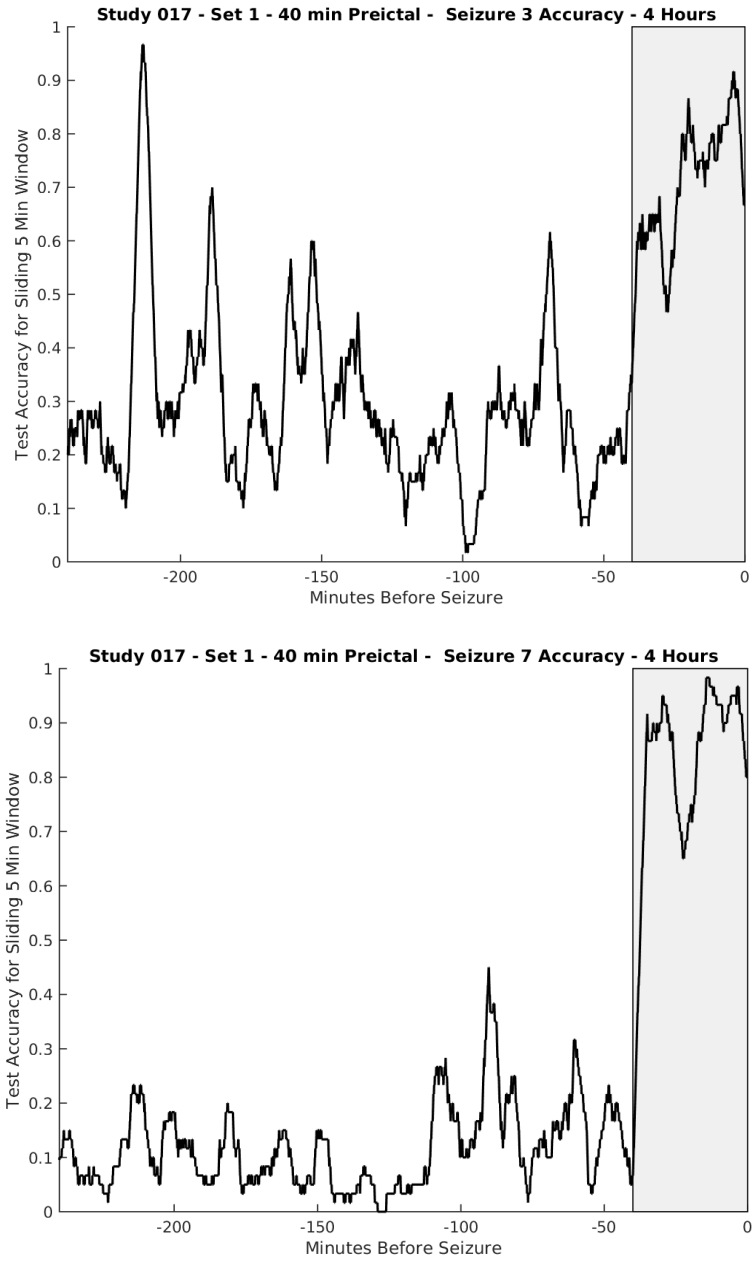


Figure 11. Set 1 - 40 Minute Accuracy. The 5-minute moving window accuracy for the last 4 hours before both test seizures. The grey boxes on the right mark the preictal window.

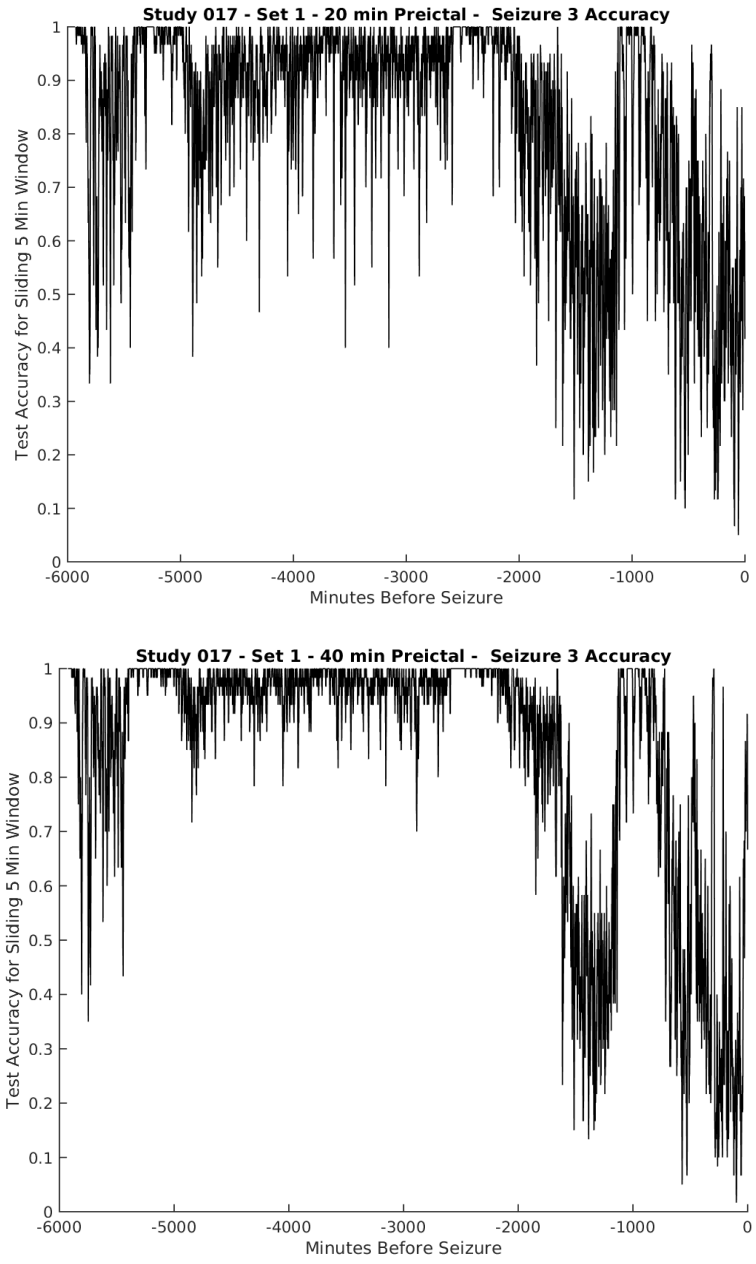


Figure 12. Set 1 – 20 and 40 Minute Accuracy. The 5-minute moving window accuracy for the full timespan before Seizure 3. The two preictal windows yield near identical results.

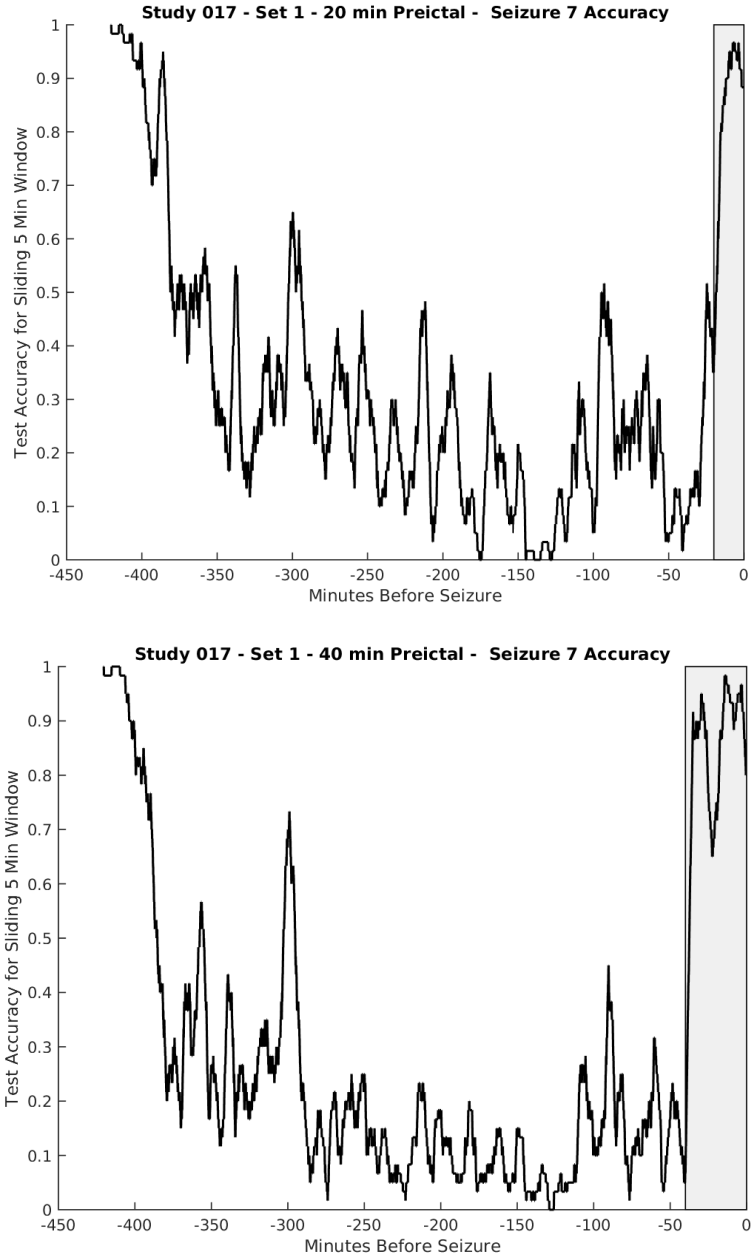


Figure 13. Set 1 – 20 and 40 Minute Accuracy. The 5-minute moving window accuracy for the full timespan before Seizure 7. The classifiers do not exhibit the same performance as for Seizure 3.

Unfortunately, these favorable results are not consistent with Seizure 7 (Figure 13). The difference in performance may be due to the features selected by the classifier working well for Seizure 3, but not 7. More likely, the difference is caused by Seizure 7 having a much shorter timespan following Seizure 6, which perhaps does not allow for the brain to return to a purely interictal state. This explanation is purely conjecture, as the time to return to normal brain functioning is unknown.

4.1.2 Training Set 2 – Pure Interictal.

In an attempt to reduce the mislabeling of data due to undefined brain state transitions, Set 2 used "pure" interictal samples, as defined in section 3.3.2. Reducing the vast interictal periods to a span of exemplars caused a shortage of data, so both the preictal and interictal classes were bootstrapped to a total of 10,000 observations.

The ROC curves for Set 2 indicate there is no longer a problem with ignoring the preictal class (Figure 14). However, the graphs also indicate the classifier would not out-perform a random guess. All of the ROC curves for the Set 2 data are near-perfect diagonals, showing that the classifier did not learn from the data.

Figures 15, 16, 17, and 18 show that the classifier predicts randomly for the 4 hours preceding the seizure. This lack of training must be attributed to the limitation of data. With so few examples of the highly variable interictal data, the model must classify observations that are unlike anything from the training set.

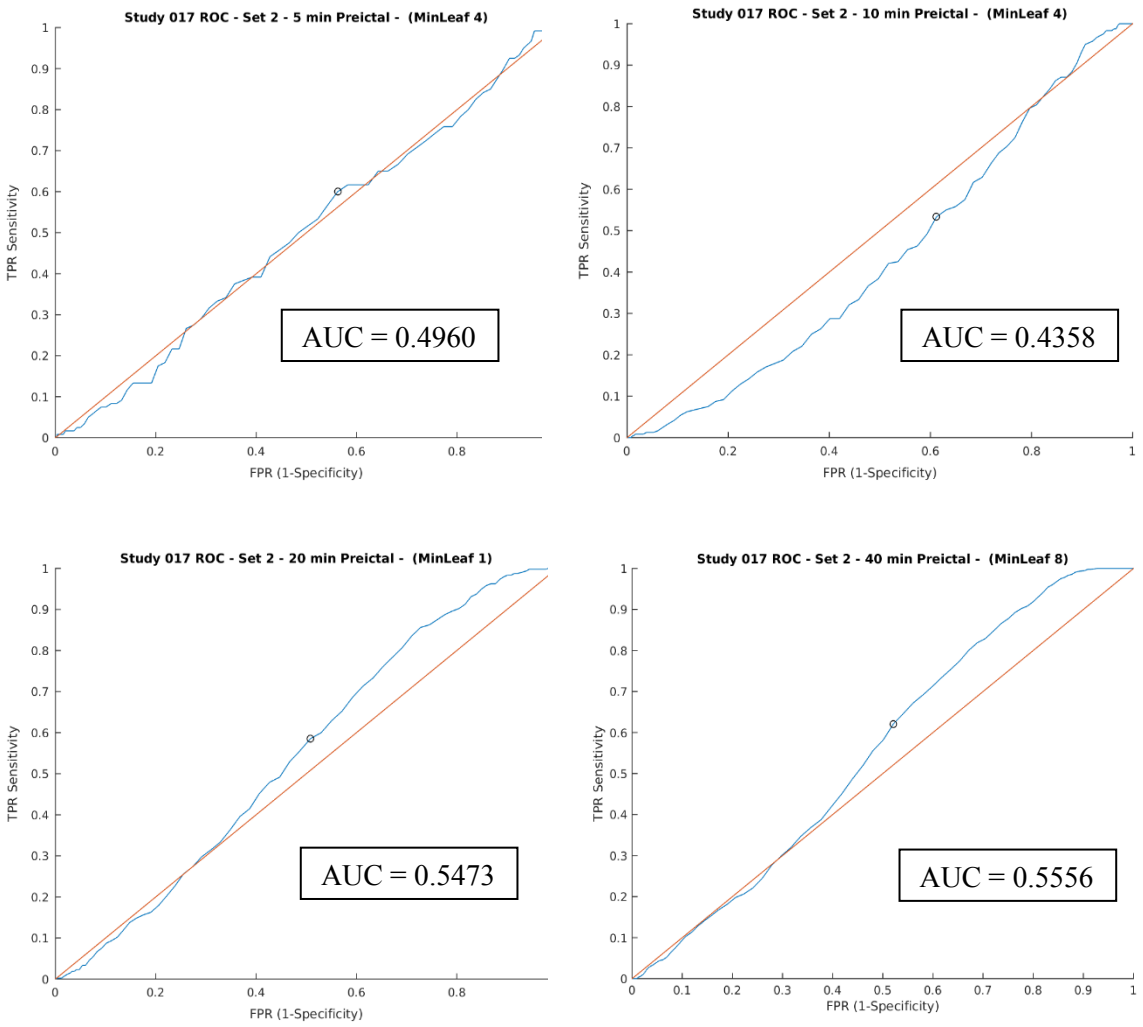


Figure 14. Set 2 ROC Curves. From left to right and top to bottom, the ROC curves for the 5, 10, 20, and 40 minute preictal windows. The area under the curve (AUC) is also shown.

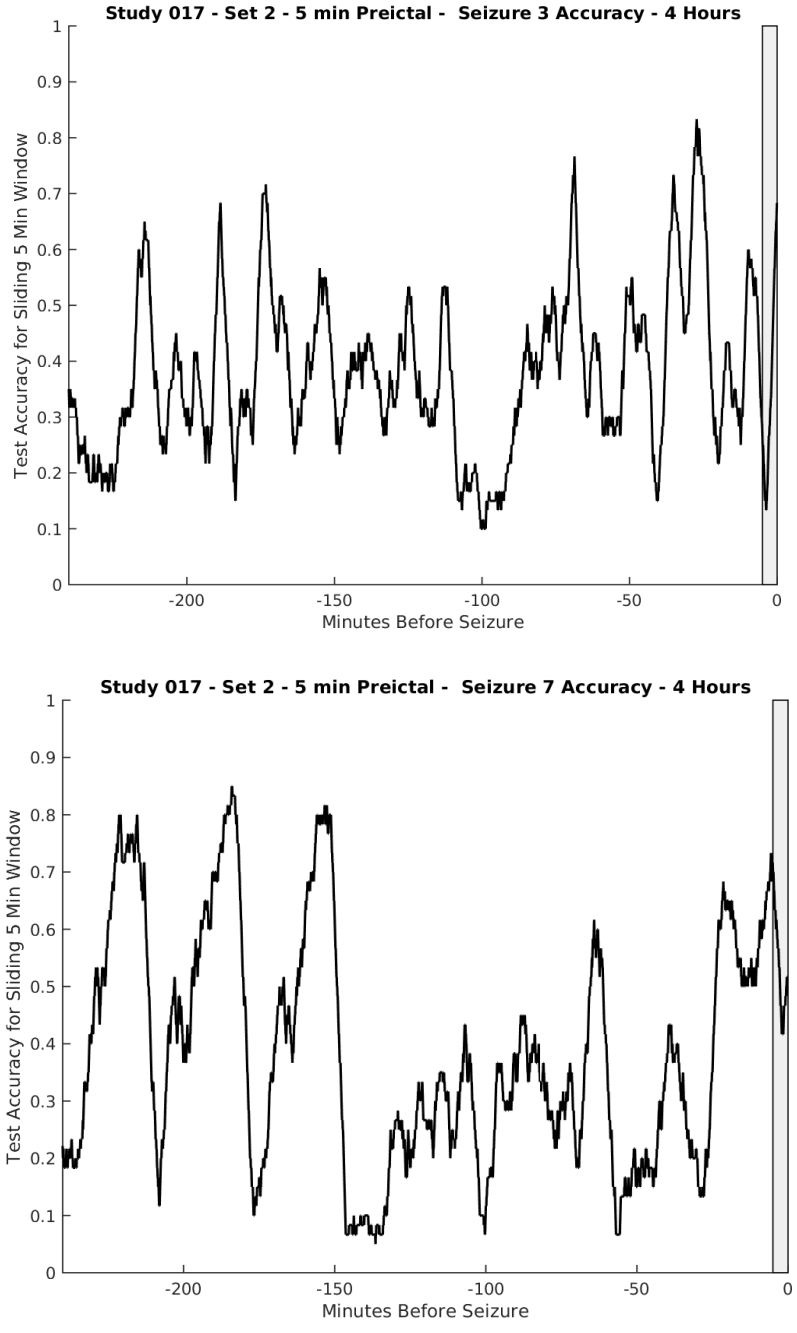


Figure 15. Set 2 -5 Minute Accuracy. The 5-minute moving window accuracy for the last 4 hours before both test seizures. The grey boxes on the right mark the preictal window.

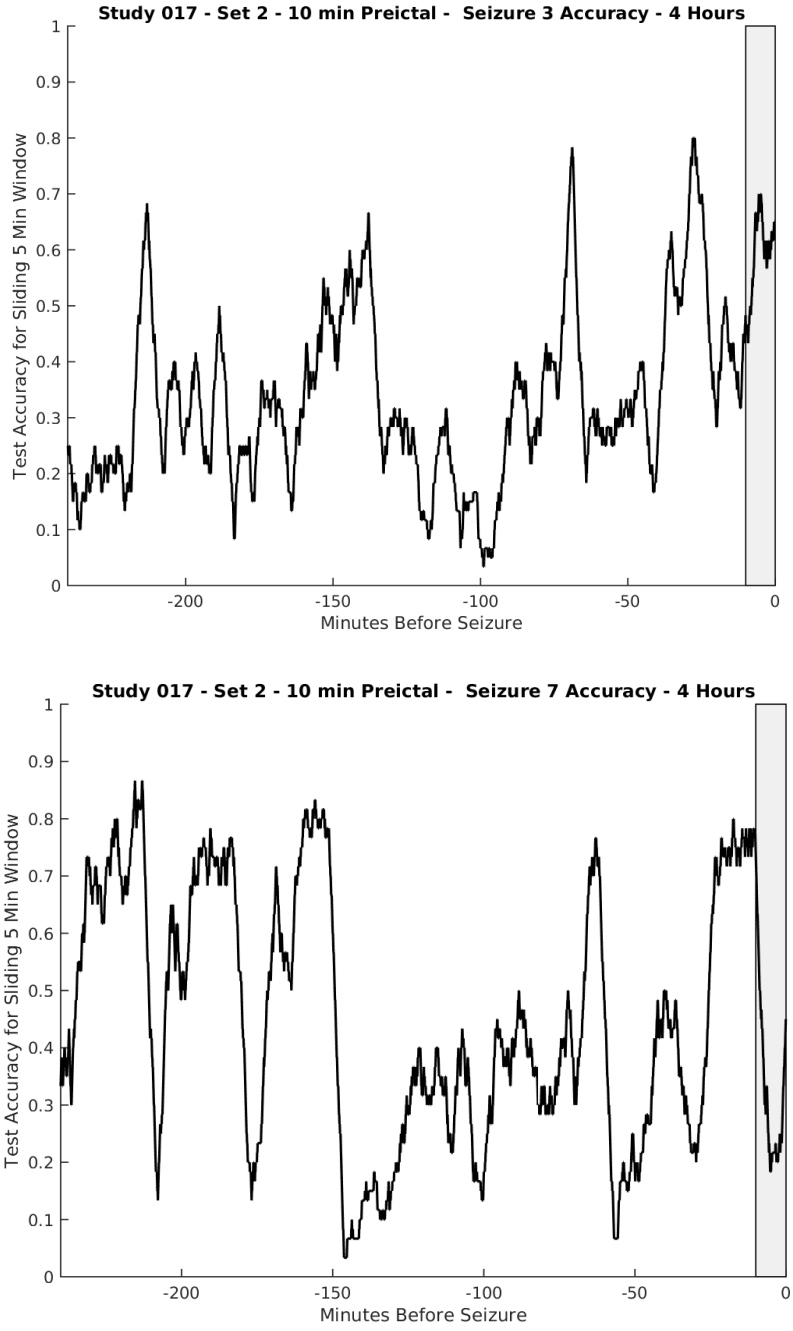


Figure 16. Set 2 - 10 Minute Accuracy. The 5-minute moving window accuracy for the last 4 hours before both test seizures. The grey boxes on the right mark the preictal window.

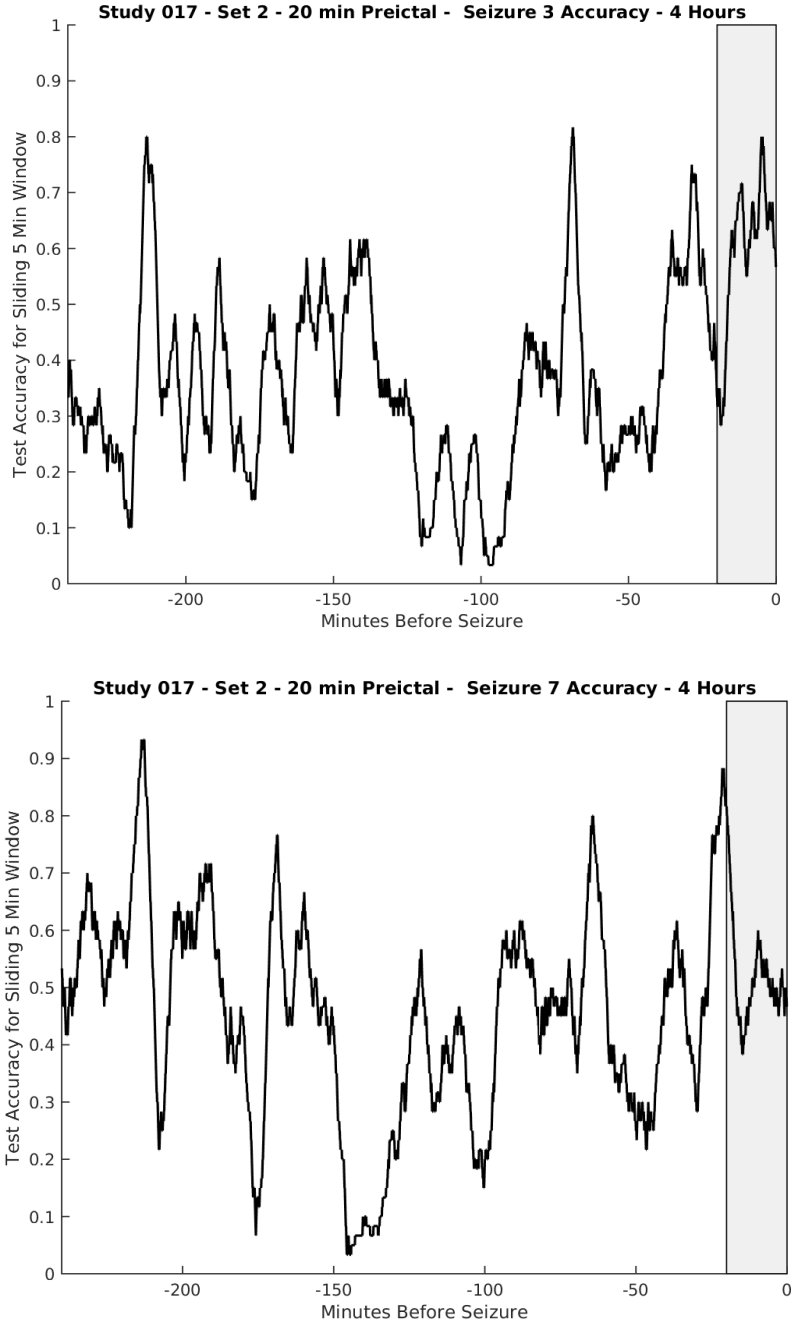


Figure 17. Set 2 - 20 Minute Accuracy. The 5-minute moving window accuracy for the last 4 hours before both test seizures. The grey boxes on the right mark the preictal window.

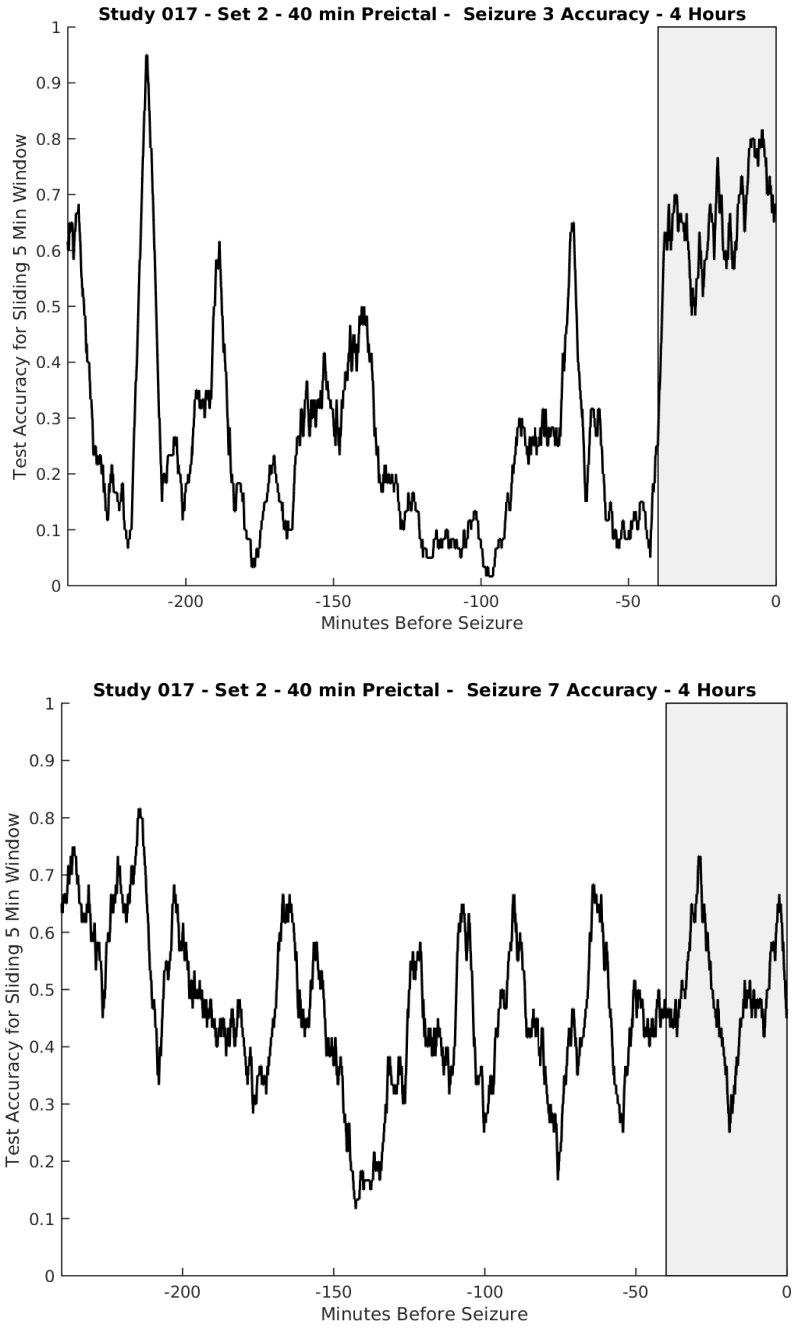


Figure 18. Set 2 - 40 Minute Accuracy. The 5-minute moving window accuracy for the last 4 hours before both test seizures. The grey boxes on the right mark the preictal window.

4.1.3 Training Set Comparison.

Limiting Set 2 to a subset samples did not have the desired outcome of learning the data through exemplars. Instead, the data limitation reduced the classifier's performance to random for most windows.

Training on the full data showed promising performance for the larger preictal windows, but had the expected decreases in accuracy in the transitional areas, as well as some unexplained dips that may have been caused by artifacts or patient activity. However, these transitional areas will presumably always pose an issue until the lengths of all states can be determined. This fact once again highlights the utility of the time-series accuracy graph as an evaluation metric for seizure predictors.

4.1.4 Feature Importance.

As described in section 3.5.1, the PVDE values were stored for each classifier. In the same manner as the correlation matrices, the matrices were conditionally formatted to shade by value for visual inspection of prominent features, keeping in mind the best performance shown by Set 1 with 20 and 40 minute windows (Figure 19).

Darker shades indicate higher values, though each table is shaded in comparison to only itself. Set 1 has higher overall values, which follows from the classifiers finding stronger relationships between that data and output. Set 2 has low values, because the classifier did not learn strong connections between the features and classes.

Across all windows, Set 1 emphasizes the beta band of node 12, particularly for the 20 and 40 minute windows. The fact that all four independently trained classifiers returned similar feature selections is indicative that the features are predictive. The poor

performance of the classifiers in the shorter windows may partially be explained by those tables emphasizing many factors that are not indicative in the larger windows. The shorter windows may be identifying features as indicative that do not have true predictive value. The longer windows have higher contrast, by emphasizing only a few key features, specifically the beta band of node 12.

In the documentation included with the patient mentions a beta frequency discharge in node 12 (as well as 11 and 13) during seizures. The node was monitoring the right side of the patient's brain, where the seizures would originate before sometimes spreading. While this information may indicate the feature identified is predictive, the beta bands in nodes 11 and 13 have low feature importance in the chart, so the connection is inconclusive.

Though the beta band of node 12 may be an indicative feature, the classifier performance was not considered high enough to perform more thorough analysis, such as comparison to a random predictor or determining the time under false warning. These analyses would be auxiliary to the primary task of identifying features, in order to compare the predictive value of the features to features used in other studies.

Training a new classifier with a compromise between the two interictal definitions would determine if the beta band of node 12 was truly indicative. If the new classifier had higher performance and showed the same feature as indicative, further investigation into the position and significance of the node in respect to monitoring would be warranted.

Set 1 – 5 Min Window – OOBE

Node	Delta	Theta	Alpha	Beta	Gamma	HFA	HFB
1	1.0212	1.4184	1.4318	1.3435	1.2098	1.4622	2.0856
2	1.0054	1.6570	1.9746	2.3460	1.5737	1.9475	2.0054
3	0.8673	1.5710	1.2860	1.2342	1.4757	1.7557	2.2734
4	0.9421	1.7302	1.5037	1.3462	2.3711	1.2292	1.9654
5	0.7985	1.5623	1.3916	2.0480	2.5219	1.5897	1.6630
6	0.8234	1.9997	1.6930	1.9122	1.5411	1.6789	1.4237
7	1.7478	1.9154	2.2174	2.2477	1.8962	2.0304	2.1279
8	0.8595	1.2457	1.4655	1.0902	1.0388	1.2938	1.7948
9	1.8542	1.8861	1.7423	2.5291	2.8413	1.6575	2.2710
10	1.1851	1.8851	2.1406	1.7050	1.6286	2.1421	1.3917
11	1.3127	2.1200	1.6349	1.6516	1.7795	2.4496	2.1820
12	1.5028	1.6771	2.0677	3.1263	2.9774	2.5164	2.6272
13	0.8285	1.7531	1.8191	1.7399	1.2668	1.9917	2.0443
14	1.2200	1.6629	1.3919	1.2583	1.3797	2.4155	2.0330
15	1.0589	1.3962	1.2405	1.6548	2.5367	1.4118	1.9358
16	1.3373	1.7946	1.8282	1.5848	2.3244	1.3914	2.1557

Set 2 – 5 Min Window – OOBE

Node	Delta	Theta	Alpha	Beta	Gamma	HFA	HFB
1	0.6503	0.6386	0.4671	0.5398	0.6645	0.5202	0.4059
2	0.7284	0.3957	0.6949	0.7035	0.7763	0.6972	0.6676
3	0.7252	0.6208	0.3760	0.4366	0.5725	0.6238	0.8192
4	0.6398	0.6363	0.3641	0.5218	0.5693	0.5722	0.7525
5	0.6823	0.7943	0.4733	0.4821	0.7279	0.5367	0.6145
6	0.7067	0.8428	0.5541	0.5313	0.7149	0.8277	0.6384
7	0.7182	0.5820	0.5617	0.6269	0.8842	0.9145	0.8292
8	0.6195	0.6269	0.5208	0.6016	0.4286	0.6210	0.5040
9	0.7680	0.6676	0.6217	0.5691	0.6213	0.6761	0.5752
10	0.5718	0.8316	0.6502	0.6311	0.7195	0.4953	0.4951
11	0.9007	0.8589	0.9374	0.6187	0.6959	0.7652	0.4803
12	0.5279	0.4974	0.6306	0.6915	0.7267	0.6506	0.6857
13	0.8095	0.6757	0.6290	0.5401	0.6801	0.7544	0.7546
14	0.7135	0.5598	0.3444	0.4341	0.7075	0.4741	0.7894
15	0.6617	0.6074	0.5680	0.6856	0.6878	0.5797	0.5760
16	0.5500	0.6595	0.6282	0.6988	0.7668	0.8224	0.6633

Set 1 – 10 Min Window – OOBE

Node	Delta	Theta	Alpha	Beta	Gamma	HFA	HFB
1	1.1356	1.0410	1.0326	1.2168	1.4365	1.4407	1.5673
2	1.1595	1.7184	2.3732	2.3185	1.9690	1.7024	1.9821
3	1.0826	1.2141	1.5517	1.5457	1.7677	2.0626	1.9811
4	0.8567	1.5614	1.4559	1.7139	2.2813	1.9440	1.7639
5	0.7088	1.4934	1.5430	1.7668	2.1662	1.4026	1.3421
6	0.7419	1.8214	2.4513	1.7664	1.6921	1.6615	1.8674
7	1.2964	1.9046	1.5063	2.5282	1.7592	1.9974	2.4817
8	0.8777	1.2195	1.2691	1.3552	1.2861	1.4744	2.0570
9	1.4042	2.6062	2.0053	1.7891	1.7575	1.6592	2.1598
10	1.2146	1.7896	1.9158	1.4271	1.7611	1.4742	1.3490
11	1.6066	2.2680	2.0514	1.7219	1.4920	2.3577	2.0151
12	0.8490	1.9138	2.0559	2.9532	1.5314	2.1597	1.9447
13	0.8344	1.5068	1.6288	1.8224	1.3949	1.5673	2.0536
14	1.0022	1.2336	1.8496	1.4154	1.8540	1.9688	1.8949
15	1.2470	1.7739	1.4892	1.7882	1.6588	1.4178	1.9379
16	1.2814	2.2652	1.8183	1.3550	1.4652	1.7963	1.5220

Set 2 – 10 Min Window – OOBE

Node	Delta	Theta	Alpha	Beta	Gamma	HFA	HFB
1	0.7439	0.7047	0.5684	0.6138	0.7939	0.7761	0.8473
2	0.7621	0.4022	0.8297	0.8596	0.8029	0.8471	0.7602
3	1.0100	0.6850	0.5343	0.6147	0.8446	0.8274	0.9524
4	0.6010	0.8588	0.6715	0.5427	0.7500	0.7189	0.6740
5	0.6830	0.8891	0.4192	0.5724	0.9051	0.7476	0.6541
6	0.8488	0.8520	0.9542	0.8029	0.9703	1.3377	0.9763
7	0.7220	0.6322	0.9486	0.7702	0.9130	1.2011	0.9658
8	0.6454	0.8073	0.6117	0.6703	0.5629	0.7918	0.6684
9	0.7107	0.9966	0.6238	0.8788	0.7831	0.8038	0.8216
10	0.5591	0.7631	0.8620	0.7270	0.7340	0.5456	0.6941
11	0.6566	0.6257	0.8869	0.6024	0.7436	0.8322	0.6010
12	0.7956	0.6650	0.8549	0.6137	0.8005	0.5812	0.6963
13	0.8240	0.6206	0.5878	0.6528	0.5170	0.7525	0.9083
14	0.7479	0.5046	0.4803	0.5546	0.7368	0.8846	0.6924
15	0.7120	0.8936	0.7806	0.5964	0.8666	0.6054	0.8406
16	0.5079	0.7032	0.8466	0.9059	0.9868	0.7259	0.6616

Set 1 – 20 Min Window – OOBE

Node	Delta	Theta	Alpha	Beta	Gamma	HFA	HFB
1	0.6632	0.9545	0.7121	0.7132	0.9546	1.3676	1.7177
2	0.5795	1.3626	1.7064	1.7083	1.1638	1.6207	1.5335
3	0.6362	1.2007	0.9851	1.1504	1.2376	1.8528	1.4396
4	0.7184	1.5223	1.3551	1.2060	1.8549	1.5825	1.5761
5	0.6706	1.5395	1.0013	1.6031	1.7143	1.7299	1.6053
6	0.6062	1.9046	1.6330	1.5866	1.3031	2.1087	1.9506
7	1.1612	1.3936	1.3404	1.1901	1.4383	1.9647	1.8473
8	0.5214	0.9894	1.0255	0.8642	0.7664	1.2569	1.5100
9	1.3007	1.3754	1.9229	1.6675	1.3461	1.1566	1.8754
10	0.9041	1.3172	1.5661	1.3217	1.1163	1.2849	1.3742
11	0.6318	1.3893	1.7784	1.0263	0.9551	2.3808	1.6467
12	0.8270	1.4193	1.8270	3.0612	1.9059	1.5290	1.7285
13	0.7156	1.1284	1.3529	1.3481	1.0938	1.2735	1.4612
14	0.9407	1.0971	1.1320	1.3744	0.9359	1.9331	1.3709
15	1.3253	1.0148	1.3017	1.1289	1.4157	1.3813	1.7294
16	1.2496	1.5203	1.2500	1.1745	1.2314	1.5792	1.2781

Set 2 – 20 Min Window – OOBE

Node	Delta	Theta	Alpha	Beta	Gamma	HFA	HFB
1	0.7880	0.6552	0.6709	0.7523	0.6724	1.0179	0.6975
2	0.9321	0.9703	1.3551	1.1511	1.0574	1.0590	1.1857
3	1.0263	0.7607	0.5125	0.7352	1.0346	1.1901	0.9435
4	0.7903	0.8483	0.4998	0.8291	0.8404	0.9180	1.0691
5	1.1488	0.8914	0.8723	0.9864	1.1114	0.9122	0.9189
6	0.8461	1.0463	0.9787	0.8393	0.8961	1.6851	1.2025
7	0.6054	1.1491	1.0898	0.9943	1.1298	1.4450	1.1417
8	0.6902	0.6754	0.6762	0.8671	0.7278	0.9601	0.9665
9	0.7483	1.0899	1.1043	1.0372	1.0688	0.7344	1.0209
10	0.7833	1.1484	1.0827	1.1303	0.7922	0.6448	0.7765
11	0.8278	0.7059	0.7745	0.7937	0.7436	1.1157	0.7956
12	0.8305	0.5237	0.8279	1.0822	1.0085	0.9866	0.7713
13	0.8860	0.7872	1.1025	0.6231	0.9412	1.0004	0.9593
14	0.6921	1.1098	0.5470	0.6517	0.8133	0.9282	0.7461
15	0.6675	0.7109	1.0800	0.6309	1.1211	0.7030	0.8878
16	0.7523	1.0269	0.8336	1.1383	1.2059	0.8491	0.7724

Set 1 – 40 Min Window – OOBE

Node	Delta	Theta	Alpha	Beta	Gamma	HFA	HFB
1	0.8989	1.3831	1.3947	1.1089	0.9725	1.8292	2.2059
2	1.3180	2.1247	2.3190	2.1688	1.6890	2.6901	2.0606
3	0.8955	1.2830	1.4560	1.7864	1.6994	2.2262	1.6928
4	0.9297	1.6873	1.7100	1.8571	2.3941	2.0912	1.7832
5	0.8780	1.6393	1.7885	1.6784	1.8572	2.1062	2.4054
6	1.0947	2.3913	2.0026	2.1356	1.5185	2.8034	3.7681
7	1.6665	1.9708	1.8681	2.4198	1.4754	2.6721	3.2127
8	0.9052	1.2101	1.2946	1.6256	1.0183	1.7737	1.8653
9	1.5553	1.7035	2.1971	2.7230	2.3797	1.7801	2.6881
10	0.9700	1.4851	1.8934	2.3930	2.1491	2.0907	1.9404
11	1.5834	1.7496	1.8779	1.2476	1.3672	2.4965	1.5213
12	1.0968	1.9811	2.4081	3.7130	2.0716	1.7376	1.8924
13	1.0496	1.6657	1.7198	2.2447	1.4714	1.8191	1.7278
14	1.2301	1.7775	1.9474	2.0489	1.6325	2.4931	1.9061
15	1.5489	1.8734	2.3147	2.0955	1.6756	2.7103	2.1305
16	1.6957	2.3135	2.3033	3.2590	1.6169	2.3296	2.2012

Set 2 – 40 Min Window – OOBE

Node	Delta	Theta	Alpha	Beta	Gamma	HFA	HFB
1	0.6912	0.6369	0.5631	0.6634	0.5547	0.6650	0.5289
2	0.7227	0.8154	0.9620	0.8848	0.9467	0.6529	0.9138
3	0.6391	0.5315	0.5634	0.4880	0.6054	0.8401	0.7466
4	0.7083	0.7186	0.7448	0.8314	0.7559	0.7048	0.7056
5	0.6020	0.7392	0.7330	0.7697	0.9419	1.1002	0.7865
6	0.6082	0.8761	0.9073	0.9451	0.7463	1.5266	1.2640
7	0.5606	0.8795	0.9267	0.9404	0.9525	1.0708	0.9795
8	0.5710	0.5974	0.6020	0.7061	0.6417	0.7689	0.6950
9	0.7138	0.9322	1.0764	0.7413	0.6575	0.9210	1.0910
10	0.7987	0.9966	0.9164	0.7505	1.1577	0.5598	0.9551
11	0.7395	0.7775	0.7207	0.6410	0.6895	1.1825	0.6292
12	0.6727	0.9349	0.5976	0.9900	1.2342	0.7882	0.8823
13	0.7047	0.6785	0.8445	0.8307	0.6258	0.8649	0.9258
14	0.5322	0.6749	0.6817	0.6738	0.6565	0.6409	0.8041
15	0.7566	0.7189	0.7197	0.7895	0.7639	0.7861	0.7825
16	0.6756	0.5893	0.8317	0.7820	0.5653	0.7139	0.8037

Figure 19. Shaded PVDE Tables. The shading is normalized within each OOBPermutedVarDeltaError (PVDE) table, with dark grey indicating maximum values. The columns correspond to Set 1 and Set 2, with the rows from top to bottom indicating windows 5 to 40. The columns within the tables are the EEG frequency bins, while the rows are the nodes on the brain.

V. Conclusions and Recommendations

5.1 Conclusions of Research

This research used machine learning methods to investigate three research questions:

1. What are the key spectral power features in EEG for predicting epileptic seizures?
2. How does predictive performance change when varying the length of the preictal window, which greatly influences the impact for the patient?
3. How does a random forest classifier compare to those used in other prediction research, particularly SVMs?

Though inconclusive, results showed the answer to the first question may lie in the beta frequency bins of specific nodes. Further investigation would be necessary to confirm this result. Adjusting the alert threshold to reduce the number of false warnings would be necessary for comparison to other research. The results are also limited to a single patient, due to the collection process of the data. Investigating the features of patients with similar epilepsy may show the beta band of a particular region of the brain is indicative.

Investigation of the second question was impacted by the definitions of the interictal window. For Set 1, using a longer preictal window provided better classification results, however, including the possibly preictal samples in the interictal data of the shorter windows may have been the key factor behind this difference, rather than the more accurate definition of the preictal window. Still, a predictive feature was found as far as 40 minutes in advance and maintained validity into the shorter windows, indicating a 40 minute preictal window is viable.

The answer to the third question is still open to debate. While the classifier had inconsistent accuracy across the full timespans, much of this error may have been due to the segmentations of the data windows. Also, the performance of the classifier cannot be fairly compared to other research that focused on seizure prediction, as the focus of this research was classification. The modest performance with Set 1 for the 20 and 40 minute windows would imply that higher prediction accuracy is achievable. This question requires further research and analysis to reach a proper conclusion.

5.2 Significance of Research

The identification of a predictive feature as far as 40 minutes prior to seizure onset is encouraging for the prospect of real-time monitoring systems. If predictive features are found at 40 minutes, the algorithm could possibly give a preliminary warning, then keep the patient updated on the status of whether or not the algorithm still predicts a seizure as the preictal window shortens.

This research also emphasized the importance of identifying the transitional states, and their major impacts on classifier performance. The use of a time-series accuracy graph was a novel approach to investigating classifier performance, and would enhance the analysis of future classifiers.

Though inconclusive, the application of a random forest classifier to this field of research showed promise of improved performance. The speed and interpretability of these algorithms may have positive implications for future research, now that they have been introduced to the field as a viable method.

5.3 Recommendations for Future Research

The process of this study revealed several issues that require significant investigation, as well as elucidated possible methodologies for such investigations. This section emphasizes those findings with the hope that future works will address them.

Testing a third definition of an interictal training sample may yield better results. A definition that excludes a postictal period and leaves a gap before the preictal period, but without eliminating as much data as Set 2, may perform well.

This analysis focused specifically on classification in order to determine predictive features, rather than focusing on the prediction algorithm. One disadvantage to this methodology is the lack of a comparison measure to other prediction research. Purely for this comparison, a basic thresholding analysis to increase the number of positive classifications necessary to alert the patient, as well as a time-under-false-warning graph would improve this research.

To improve upon this research, a wavelet transform preprocessing method, to localize by time and frequency (rather than just frequency with FFT), would maintain more information in the signal. Changing this preprocessing step, but maintaining the rest of the methodology, may provide significant improvements.

Another area of improvement would be comparing the nonlinear random forest classifier to a linear classifier, such as logistic regression. Binary classifiers are either linear or nonlinear, depending on how they form the decision boundary between classes. Though trained on the same data, classifier performance depends on the relationship between the observations and the outcome. In this application, if there is a linear

relationship between the features and class labels, then a linear classifier such as logistic regression should perform well. However when a nonlinear method is used on a linearly separable classification problem, it may over-train to noise in the signal, complication the decision boundary and preforming poorly on the test set. Such a comparison would contribute to the debate of linear versus nonlinear classifiers. Also, linear methods are preferable to more complicated and time-consuming nonlinear methods when the linear method is suitable for the problem.

Assuming predictive features can be determined in individual patients, another possible improvement would be to use a dataset with standardized node placement across all patients. Such a dataset may be unattainable due to the intrusiveness of intracranial EEG. Generalizing features between patients was virtually impossible for this research, since each patient had a unique number and configuration of nodes, even if they suffered from similar types of epilepsy.

The undefined duration of the postictal period may have substantial implications when using any prediction algorithm. This is especially the case for patients that suffer from frequently recurring seizures, with a short interictal period. Determining a method for handling the postictal period that is robust to frequent seizures is a topic for future work.

In the realm of optimizing seizure prediction, the next investigation should include various synchronization measures and their predictive values. As discussed in section 2.3.1 of the Literature Review, synchronization measures may be suitable features

to use for seizure prediction. Such features could then be classified using random forests and evaluated using the same measures as our research.

Finally, the present study was not able to capture the time-series aspect of the data. Each observation was treated as an independent data point, without considering its relationship with previous observations. Using a time-series method such as a Hidden Markov Model (HMM) may address the loss of time-series information, which would address several issues associated with the time windows. Hidden Markov Models are useful when the states are not directly observable, but a signal from the states is observable. A Hidden Markov Model may be suitable for this application if the states were defined as interictal, preictal, ictal, and postictal. The EEG signal is always visible, but the preictal and postictal states are hidden.

Using an HMM may account for the unknown length of the preictal and postictal windows, the dilemma of setting the alert threshold, as well as the likelihood that postictal data resembles preictal data. By accounting for the known states (interictal and ictal) the probability of transitioning to the hidden states would incorporate the fact that a person in an interictal state cannot move to a postictal state. This may greatly improve prediction accuracy for closely spaced seizures with brief (or nonexistent) interictal periods, which are a current obstacle for prediction algorithms. The issue with using an HMM may be the increased computational complexity. The algorithm would need to run quickly to be useful in an early warning device.

Appendix A – Treatment Methods

There are currently several options to treat epilepsy. The most common treatments are drug regimens, electrical stimulation, and brain surgery. Drug administration and electrical stimulation during a seizure are common to mitigate effects. In fact, the threats of permanent brain damage and death during status epilepticus have compelled doctors to recommend erring on the side of excessive medication, rather than risk under-dosing (Roth & Blum, 2014). Brain surgery to attempt to remove the epileptic focus of the brain is a last resort when other methods fail. However, for some people even surgery fails and they continue to have seizures.

Treatment is a fragile process, as the medication to treat one form of epilepsy may exacerbate a different form. Epileptologists have to be careful with drug prescriptions, especially with patients suffering from multiple types of epilepsy.

There is some debate about the effectiveness of open-loop versus closed-loop treatments. Open-loop treatments are on a schedule, independent of any feedback from the patient. Treatments can be either drug dosages or electrical neuronal stimulations. Closed-loop treatments consider biofeedback and are designed to administer only when they are needed, such as imminently before or shortly after the start of a seizure. A significant amount of research has concentrated on implementing closed-loop systems, despite closed-loop treatment having yet to prove superior to open-loop as far as efficacy and tolerability (Mormann et al., 2007). The future of treatment may consist of "fully-automated closed-loop seizure prevention systems" (Mormann et al., 2007), which would predict and treat seizures before their onset.

Appendix B – Guidelines for Seizure Prediction Algorithms

(Mormann et al., 2007)

- Prediction algorithms should be tested on unselected continuous long-term recordings covering several days of EEG in order to comprise the full spectrum of physiological and pathophysiological states for an individual patient.

- Studies should assess both sensitivity and specificity and should report these quantities with respect to the applied prediction horizon. Rather than false prediction rates, the portion of time under false warning should be reported. If false prediction rates are reported, they should be reported only for the seizure-free interval.

- Results should be tested using statistical validation methods based on Monte Carlo simulations or naïve prediction schemes to prove that a given prediction algorithm performs indeed above chance level. This is particularly important for studies that contain in-sample optimization such as retrospective adjustment of parameters or selection of EEG channels.

- If prediction algorithms are optimized using training data (in-sample), they should be tested on independent testing data (out-of-sample). If part of the data from an individual patient are used for patient-specific parameter adjustment or EEG channel selection, these data must be excluded when evaluating the performance out-of-sample. Performance of an algorithm should always be reported separately for the testing data.

Appendix C – Correlation Tables

Set 1 – 5 Min Window - Correlation Coefficients

Node	Delta	Theta	Alpha	Beta	Gamma	HFA	HFB
1	-0.00110	-0.00006	0.00069	0.00342	0.00351	0.00116	0.00045
2	-0.00096	-0.00179	-0.00281	-0.00352	-0.00248	-0.00028	-0.00068
3	-0.00016	-0.00087	-0.00074	-0.00105	-0.00147	-0.00157	-0.00169
4	-0.00013	-0.00086	-0.00086	-0.00113	-0.00183	-0.00176	-0.00173
5	-0.00042	-0.00092	-0.00097	-0.00132	-0.00210	-0.00164	-0.00167
6	-0.00092	-0.00119	-0.00130	-0.00166	-0.00237	-0.00260	-0.00222
7	-0.00085	-0.00121	-0.00142	-0.00159	-0.00188	-0.00172	-0.00158
8	-0.00101	-0.00024	0.00075	0.00313	0.00350	0.00112	0.00035
9	0.00073	0.00098	0.00381	0.00498	0.00463	0.00224	0.00086
10	-0.00065	-0.00072	0.00156	0.00490	0.00282	0.00390	0.00224
11	-0.00078	-0.00065	0.00152	0.00526	0.00551	0.00383	0.00290
12	-0.00062	-0.00061	-0.00003	0.00095	0.00052	0.00127	0.00057
13	-0.00074	0.00006	0.00078	0.00283	0.00287	0.00064	-0.00069
14	-0.00039	-0.00078	-0.00045	0.00009	-0.00078	-0.00184	-0.00203
15	-0.00021	-0.00054	-0.00042	-0.00073	-0.00281	-0.00239	-0.00225
16	-0.00029	-0.00076	-0.00064	-0.00082	-0.00344	-0.00281	-0.00234

Set 1 – 10 Min Window - Correlation Coefficients

Node	Delta	Theta	Alpha	Beta	Gamma	HFA	HFB
1	-0.00142	0.00031	0.00141	0.00630	0.00704	0.00246	0.00170
2	-0.00124	-0.00240	-0.00408	-0.00515	-0.00374	-0.00033	-0.00112
3	-0.00003	-0.00093	-0.00082	-0.00086	-0.00216	-0.00238	-0.00268
4	-0.00011	-0.00107	-0.00117	-0.00144	-0.00303	-0.00279	-0.00288
5	-0.00054	-0.00121	-0.00137	-0.00181	-0.00365	-0.00272	-0.00279
6	-0.00143	-0.00166	-0.00186	-0.00238	-0.00390	-0.00411	-0.00352
7	-0.00130	-0.00170	-0.00205	-0.00226	-0.00308	-0.00271	-0.00257
8	-0.00139	0.00004	0.00157	0.00579	0.00702	0.00237	0.00150
9	0.00035	0.00106	0.00483	0.00740	0.00776	0.00405	0.00130
10	-0.00108	-0.00124	0.00182	0.00692	0.00394	0.00522	0.00315
11	-0.00118	-0.00066	0.00233	0.00790	0.01029	0.00550	0.00529
12	-0.00098	-0.00096	-0.00021	0.00128	0.00049	0.00069	-0.00015
13	-0.00118	0.00022	0.00110	0.00390	0.00339	-0.00024	-0.00197
14	-0.00067	-0.00101	-0.00064	0.00013	-0.00145	-0.00297	-0.00329
15	-0.00041	-0.00083	-0.00063	-0.00104	-0.00411	-0.00370	-0.00363
16	-0.00075	-0.00118	-0.00091	-0.00117	-0.00478	-0.00443	-0.00382

Set 1 – 20 Min Window - Correlation Coefficients

Node	Delta	Theta	Alpha	Beta	Gamma	HFA	HFB
1	-0.00218	0.00109	0.00288	0.01062	0.01315	0.00536	0.00416
2	-0.00197	-0.00331	-0.00577	-0.00615	-0.00382	0.00058	-0.00084
3	-0.00027	-0.00140	-0.00149	-0.00111	-0.00341	-0.00363	-0.00439
4	-0.00028	-0.00172	-0.00207	-0.00216	-0.00518	-0.00446	-0.00486
5	-0.00087	-0.00186	-0.00232	-0.00277	-0.00639	-0.00444	-0.00470
6	-0.00220	-0.00245	-0.00285	-0.00359	-0.00661	-0.00628	-0.00559
7	-0.00194	-0.00250	-0.00305	-0.00337	-0.00494	-0.00418	-0.00405
8	-0.00207	0.00073	0.00329	0.00982	0.01320	0.00533	0.00390
9	-0.00034	0.00070	0.00498	0.00998	0.01112	0.00552	0.00132
10	-0.00182	-0.00221	0.00139	0.00871	0.00494	0.00796	0.00510
11	-0.00179	-0.00104	0.00306	0.01111	0.01663	0.00822	0.00897
12	-0.00150	-0.00165	-0.00108	0.00122	0.00073	0.00087	-0.00048
13	-0.00179	0.00012	0.00138	0.00533	0.00579	-0.00004	-0.00261
14	-0.00104	-0.00158	-0.00113	-0.00007	-0.00233	-0.00446	-0.00520
15	-0.00068	-0.00144	-0.00128	-0.00167	-0.00658	-0.00565	-0.00587
16	-0.00128	-0.00188	-0.00150	-0.00178	-0.00776	-0.00680	-0.00613

Set 1 – 40 Min Window - Correlation Coefficients

Node	Delta	Theta	Alpha	Beta	Gamma	HFA	HFB
1	-0.00337	0.00075	0.00299	0.01219	0.01763	0.00813	0.00546
2	-0.00313	-0.00497	-0.00847	-0.00894	-0.00522	0.00134	0.00010
3	-0.00085	-0.00310	-0.00320	-0.00323	-0.00587	-0.00580	-0.00707
4	-0.00084	-0.00346	-0.00401	-0.00436	-0.00877	-0.00694	-0.00770
5	-0.00165	-0.00359	-0.00430	-0.00525	-0.01069	-0.00680	-0.00738
6	-0.00331	-0.00383	-0.00448	-0.00572	-0.00976	-0.00902	-0.00829
7	-0.00284	-0.00380	-0.00467	-0.00524	-0.00730	-0.00611	-0.00598
8	-0.00315	0.00024	0.00334	0.01122	0.01770	0.00813	0.00511
9	-0.00091	0.00073	0.00732	0.01042	0.01238	0.00514	0.00044
10	-0.00281	-0.00347	0.00163	0.01029	0.00666	0.01303	0.01082
11	-0.00258	-0.00198	0.00375	0.01541	0.02810	0.01703	0.01869
12	-0.00232	-0.00271	-0.00203	0.00086	0.00102	0.00625	0.00417
13	-0.00280	-0.00050	0.00145	0.00573	0.00870	0.00217	-0.00187
14	-0.00159	-0.00266	-0.00221	-0.00161	-0.00428	-0.00645	-0.00776
15	-0.00123	-0.00281	-0.00300	-0.00391	-0.01102	-0.00864	-0.00915
16	-0.00235	-0.00334	-0.00312	-0.00416	-0.01302	-0.01018	-0.00946

Set 2 – 5 Min Window - Correlation Coefficients

Node	Delta	Theta	Alpha	Beta	Gamma	HFA	HFB
1	-0.09325	-0.08534	-0.10194	-0.10877	-0.09902	-0.08175	-0.11390
2	-0.09419	-0.10077	-0.11539	-0.10156	-0.11163	-0.11292	-0.10664
3	-0.08854	-0.09316	-0.10428	-0.10817	-0.11948	-0.11805	-0.10097
4	-0.08851	-0.08856	-0.09758	-0.10339	-0.09868	-0.10264	-0.08588
5	-0.08762	-0.08322	-0.09497	-0.10380	-0.06253	-0.08804	-0.08395
6	-0.08957	-0.08047	-0.09422	-0.10235	-0.09104	-0.08579	-0.09203
7	-0.08801	-0.07738	-0.08588	-0.09633	-0.08503	-0.07852	-0.08669
8	-0.09241	-0.08530	-0.10209	-0.10894	-0.09908	-0.08208	-0.11439
9	-0.05660	-0.07319	-0.05423	-0.08685	-0.06523	-0.04375	-0.04100
10	-0.08940	-0.07634	-0.06680	-0.05618	-0.03215	-0.00220	-0.00552
11	-0.08871	-0.05146	-0.02233	-0.03616	-0.02686	0.00182	-0.06659
12	-0.08910	-0.08618	-0.08878	-0.07937	-0.07821	-0.02691	-0.05058
13	-0.08872	-0.08654	-0.08929	-0.09037	-0.07953	-0.07185	-0.10081
14	-0.08865	-0.08488	-0.10543	-0.11196	-0.11642	-0.13320	-0.14236
15	-0.08806	-0.08314	-0.09244	-0.10060	-0.08335	-0.10886	-0.10568
16	-0.08373	-0.07875	-0.08962	-0.09537	-0.06567	-0.09602	-0.09789

Set 2 – 10 Min Window - Correlation Coefficients

Node	Delta	Theta	Alpha	Beta	Gamma	HFA	HFB
1	-0.08651	-0.07207	-0.07903	-0.06344	-0.05639	-0.05918	-0.07676
2	-0.08766	-0.09485	-0.10975	-0.10718	-0.12186	-0.09927	-0.10902
3	-0.08115	-0.08303	-0.09080	-0.09253	-0.12823	-0.12229	-0.11669
4	-0.08265	-0.08149	-0.08926	-0.09553	-0.12477	-0.11984	-0.11321
5	-0.08239	-0.07767	-0.08683	-0.09808	-0.10015	-0.10947	-0.10921
6	-0.08739	-0.07657	-0.08586	-0.09383	-0.11711	-0.10337	-0.10135
7	-0.08564	-0.07387	-0.08108	-0.08885	-0.10899	-0.09560	-0.10535
8	-0.08725	-0.07337	-0.07999	-0.06454	-0.05798	-0.06228	-0.08032
9	-0.06567	-0.07421	-0.04665	-0.06226	-0.02319	-0.01608	-0.03469
10	-0.08567	-0.07892	-0.05855	-0.03877	-0.01742	0.00747	0.00139
11	-0.08544	-0.04361	-0.00725	-0.02463	-0.00602	0.00169	-0.03881
12	-0.08655	-0.08284	-0.07686	-0.06965	-0.07933	-0.05647	-0.08156
13	-0.08661	-0.07753	-0.08211	-0.08535	-0.09147	-0.10529	-0.12986
14	-0.08614	-0.08183	-0.09843	-0.10486	-0.12618	-0.13529	-0.14620
15	-0.08456	-0.07962	-0.08404	-0.09096	-0.10203	-0.11599	-0.12011
16	-0.08165	-0.07545	-0.07969	-0.08457	-0.08759	-0.10940	-0.11289

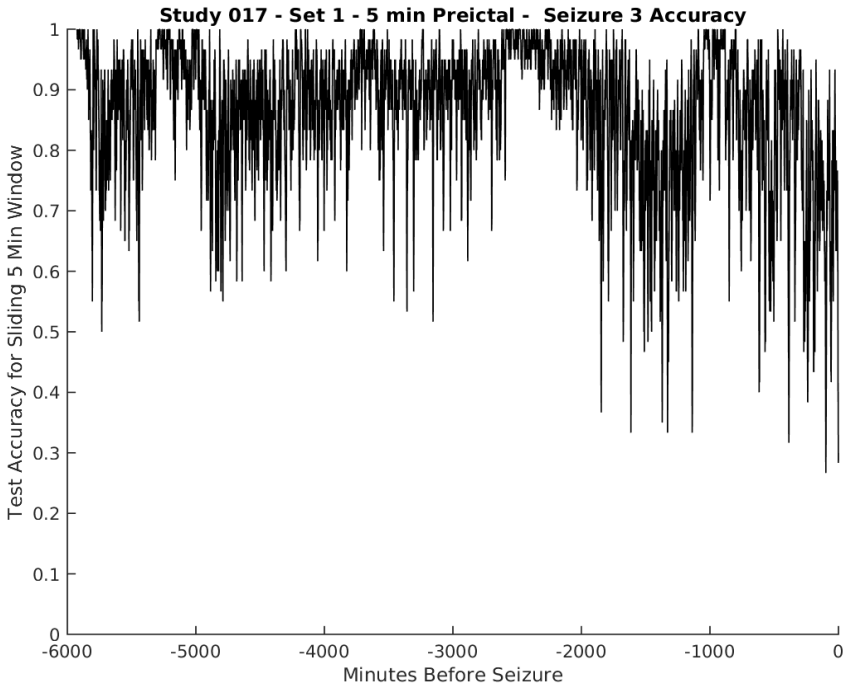
Set 2 – 20 Min Window - Correlation Coefficients

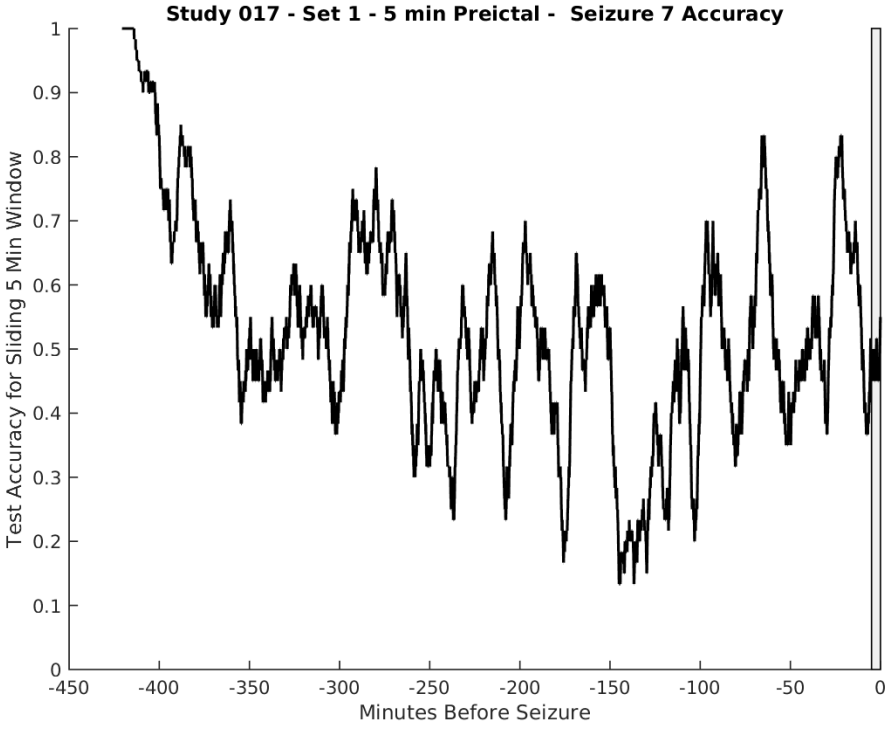
Node	Delta	Theta	Alpha	Beta	Gamma	HFA	HFB
1	-0.06979	-0.02765	-0.01866	-0.00009	0.01622	-0.00324	-0.02071
2	-0.07086	-0.08042	-0.10766	-0.07350	-0.06984	-0.04319	-0.05649
3	-0.06221	-0.06262	-0.08187	-0.06545	-0.12563	-0.10081	-0.10632
4	-0.06328	-0.06398	-0.08146	-0.07552	-0.14533	-0.11191	-0.11223
5	-0.06344	-0.05941	-0.07868	-0.08179	-0.14309	-0.10954	-0.10795
6	-0.07050	-0.06274	-0.07804	-0.08228	-0.15794	-0.11398	-0.09993
7	-0.06926	-0.06089	-0.07384	-0.07882	-0.14204	-0.10182	-0.10514
8	-0.06981	-0.02842	-0.01915	-0.00055	0.01547	-0.00472	-0.02219
9	-0.05911	-0.07715	-0.06704	-0.03005	0.00986	0.00388	-0.01637
10	-0.07064	-0.07361	-0.07063	-0.01784	0.00483	0.01569	0.01041
11	-0.06879	-0.03036	0.00625	-0.00168	0.01886	0.01033	0.00331
12	-0.06986	-0.07516	-0.08209	-0.06201	-0.06623	-0.05409	-0.07805
13	-0.06995	-0.06666	-0.07513	-0.06495	-0.07181	-0.09570	-0.11734
14	-0.06955	-0.07000	-0.09070	-0.09332	-0.13326	-0.12491	-0.14106
15	-0.06686	-0.06750	-0.07556	-0.07242	-0.12819	-0.10761	-0.11769
16	-0.06607	-0.06217	-0.06428	-0.06081	-0.11747	-0.11249	-0.11399

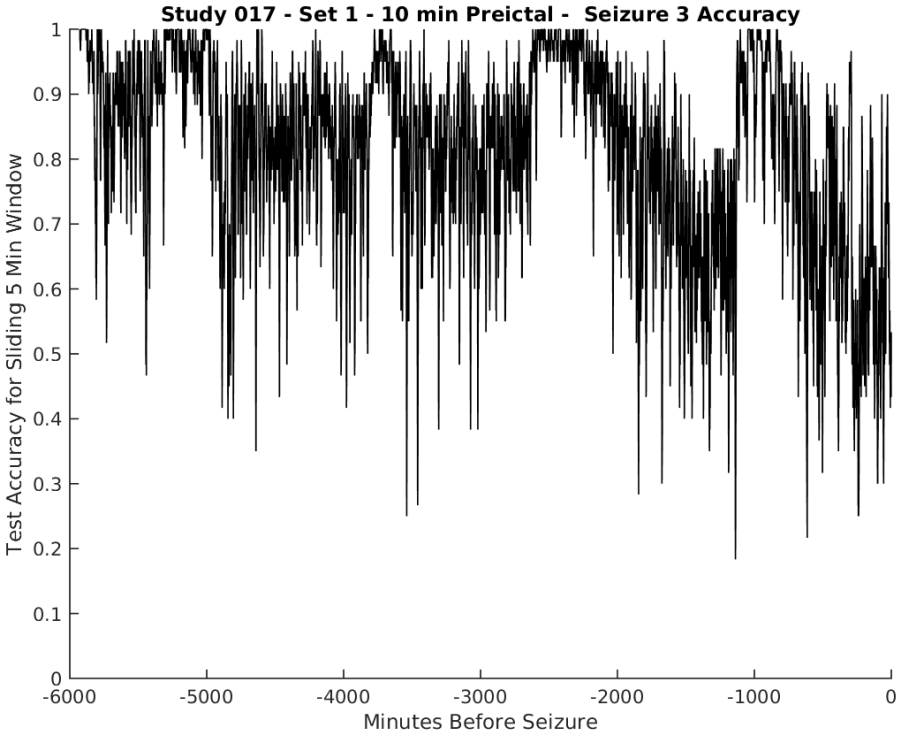
Set 2 – 40 Min Window - Correlation Coefficients

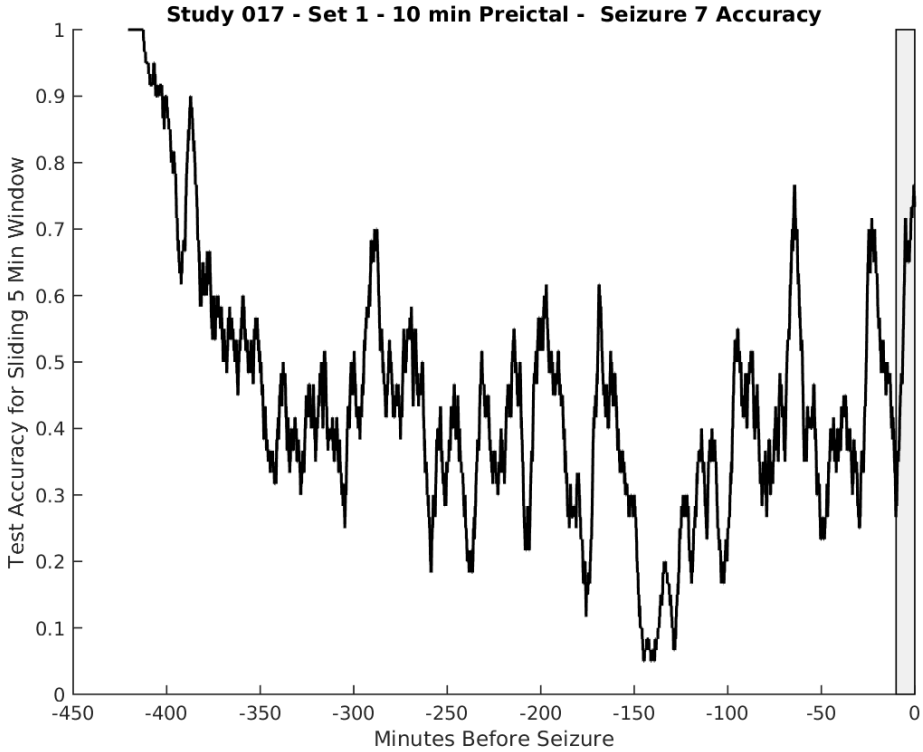
Node	Delta	Theta	Alpha	Beta	Gamma	HFA	HFB
1	-0.05793	-0.03326	-0.03039	-0.02197	0.01364	0.00950	-0.01563
2	-0.05995	-0.08682	-0.11636	-0.08047	-0.05569	-0.02368	-0.01938
3	-0.05062	-0.07066	-0.08987	-0.07669	-0.12977	-0.09801	-0.10905
4	-0.05034	-0.06736	-0.08826	-0.08359	-0.16207	-0.11201	-0.11566
5	-0.05171	-0.06303	-0.08357	-0.08770	-0.16001	-0.10661	-0.10972
6	-0.05656	-0.05900	-0.07533	-0.07902	-0.15153	-0.10099	-0.09264
7	-0.05527	-0.05440	-0.07065	-0.07478	-0.14309	-0.09559	-0.09932
8	-0.05736	-0.03360	-0.03071	-0.02221	0.01318	0.00856	-0.01668
9	-0.04803	-0.06917	-0.03701	-0.05593	-0.02084	-0.03140	-0.03485
10	-0.05796	-0.07196	-0.05308	-0.02795	0.02101	0.03802	0.04820
11	-0.05384	-0.03860	-0.00231	-0.00176	0.03707	0.04884	0.04457
12	-0.05616	-0.07366	-0.07316	-0.05972	-0.04671	0.00922	0.00152
13	-0.05835	-0.07890	-0.08232	-0.08633	-0.06409	-0.04458	-0.06429
14	-0.05458	-0.06553	-0.08881	-0.10315	-0.14068	-0.11176	-0.13015
15	-0.05269	-0.07102	-0.08148	-0.08040	-0.13623	-0.10616	-0.11948
16	-0.05426	-0.06358	-0.06995	-0.07021	-0.12233	-0.10586	-0.11200

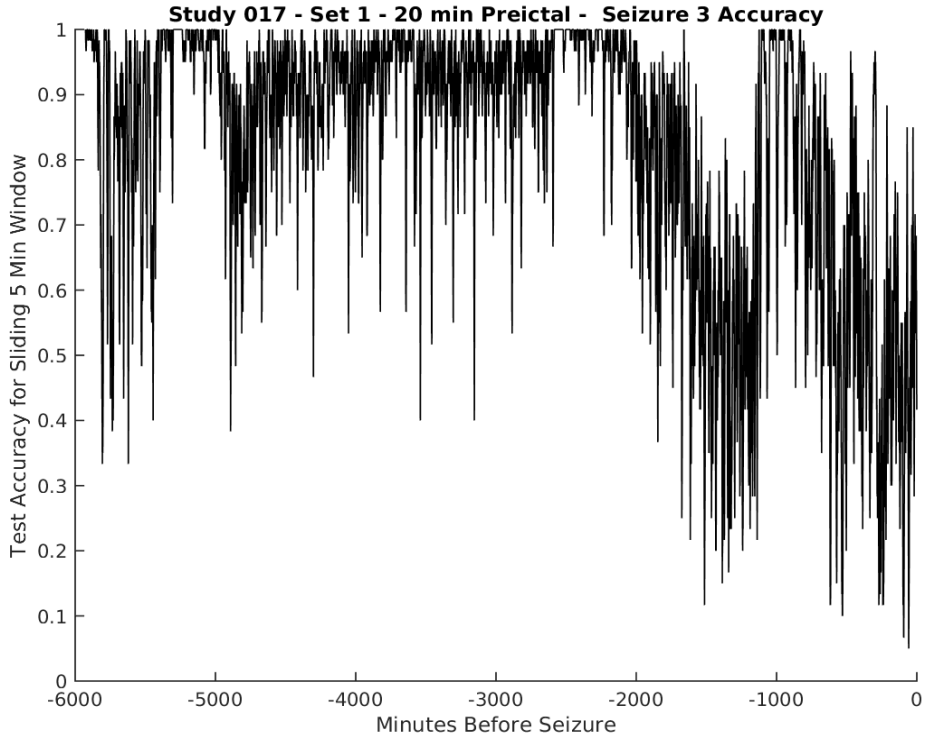
Appendix D – Set 1 - Full Time-Series Accuracy Graphs

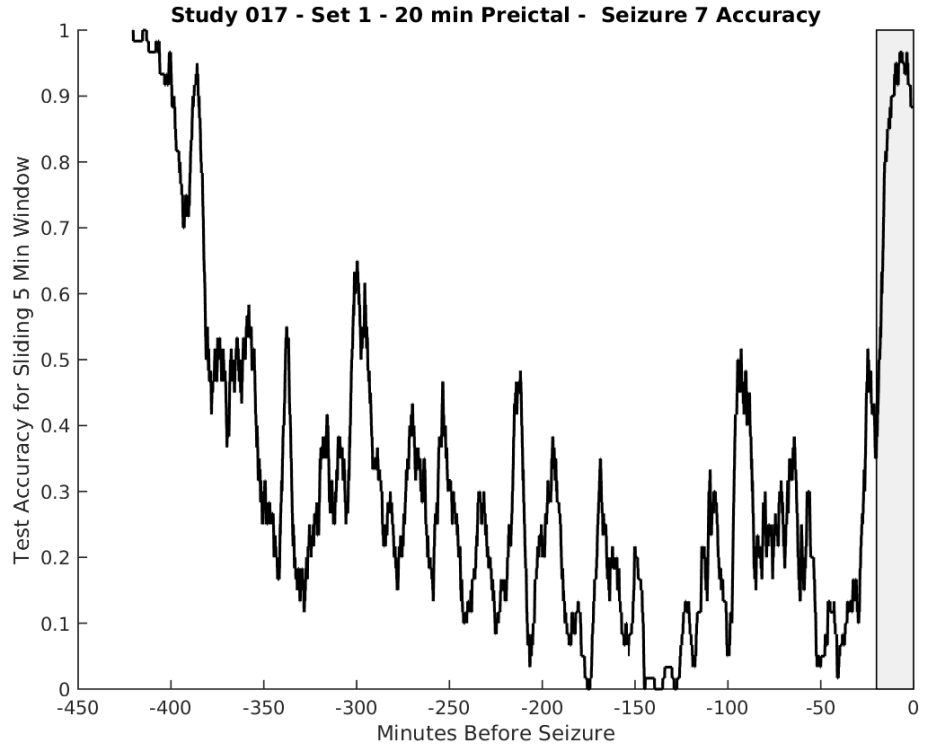


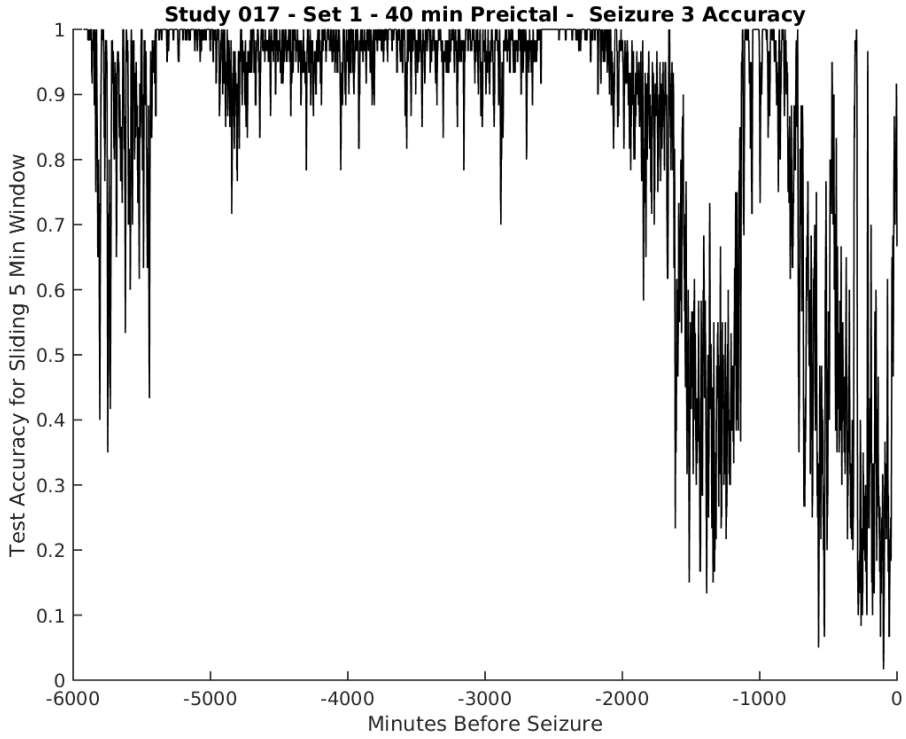


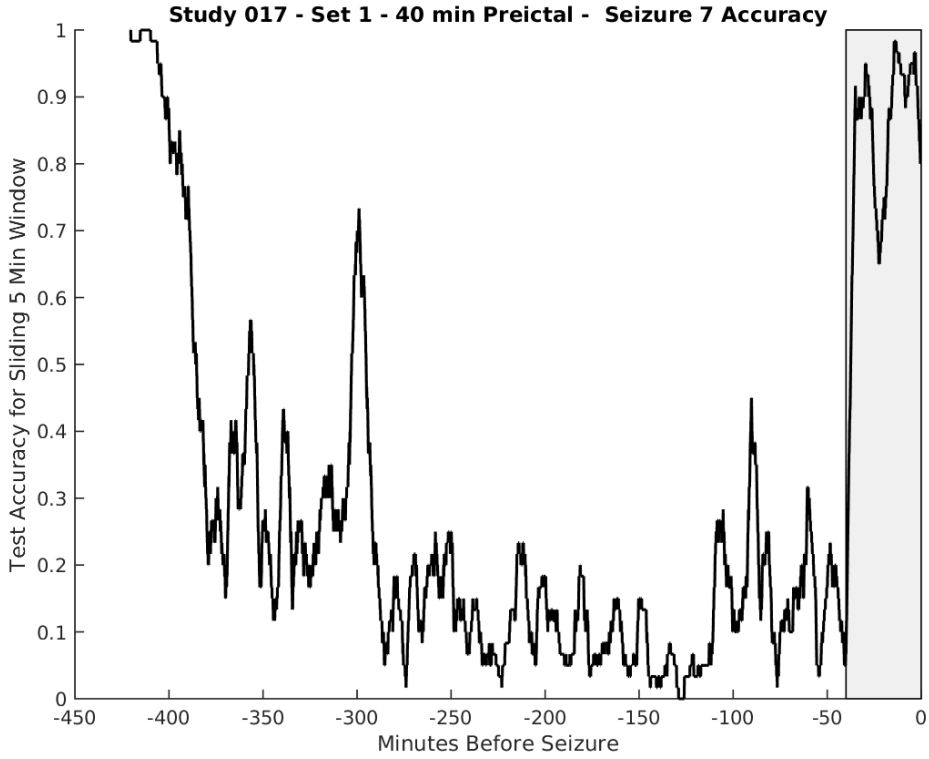

Study 017 - Set 1 - 10 min Preictal - Seizure 3 Accuracy
Test Accuracy for Sliding 5 Min Window
Minutes Before Seizure
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
-6000
-5000
-4000
-3000
-2000
-1000
0


Study 017 - Set 1 - 10 min Preictal - Seizure 7 Accuracy
Test Accuracy for Sliding 5 Min Window
Minutes Before Seizure
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
-450
-400
-350
-300
-250
-200
-150
-100
-50
0

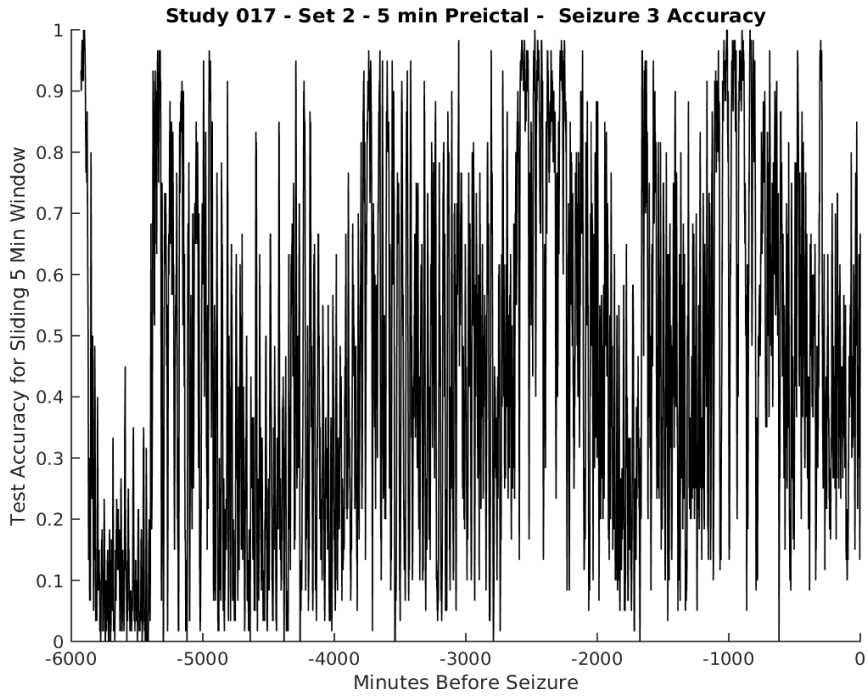

Study 017 - Set 1 - 20 min Preictal - Seizure 3 Accuracy
Test Accuracy for Sliding 5 Min Window
Minutes Before Seizure
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
-6000
-5000
-4000
-3000
-2000
-1000
0

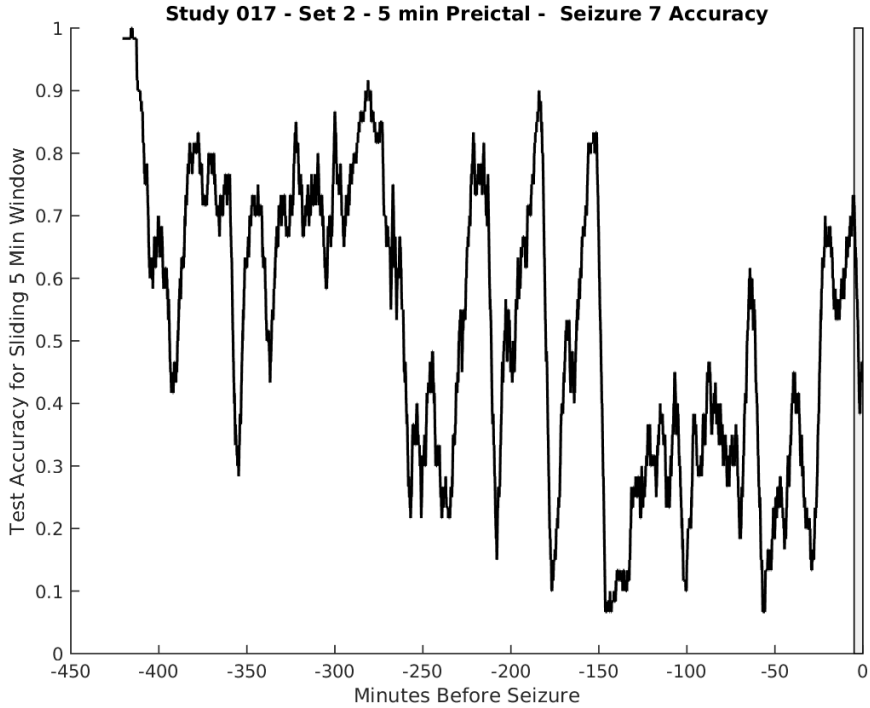

Study 017 - Set 1 - 20 min Preictal - Seizure 7 Accuracy
Test Accuracy for Sliding 5 Min Window
Minutes Before Seizure
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
-450
-400
-350
-300
-250
-200
-150
-100
-50
0

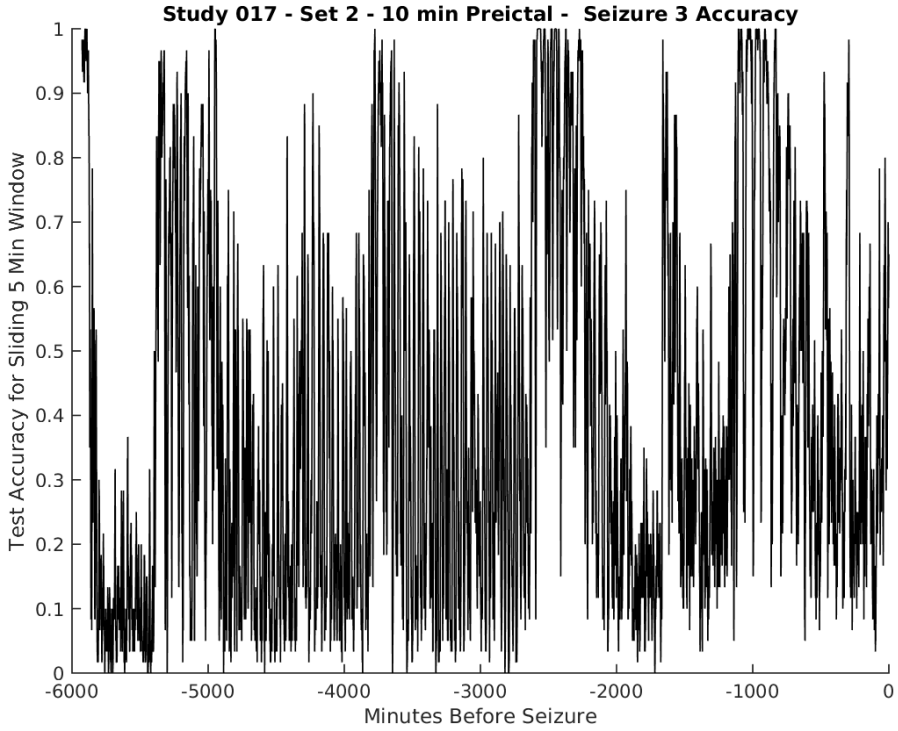

Study 017 - Set 1 - 40 min Preictal - Seizure 3 Accuracy
Test Accuracy for Sliding 5 Min Window
Minutes Before Seizure
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
-6000
-5000
-4000
-3000
-2000
-1000
0

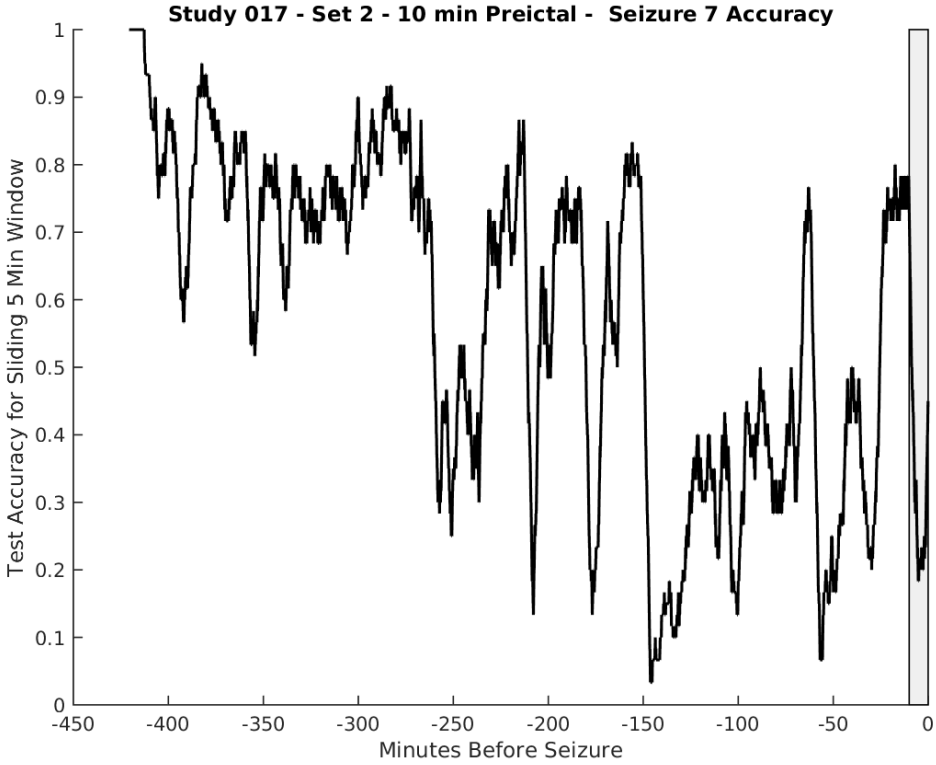

Study 017 - Set 1 - 40 min Preictal - Seizure 7 Accuracy
Test Accuracy for Sliding 5 Min Window
Minutes Before Seizure
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
-450
-400
-350
-300
-250
-200
-150
-100
-50
0

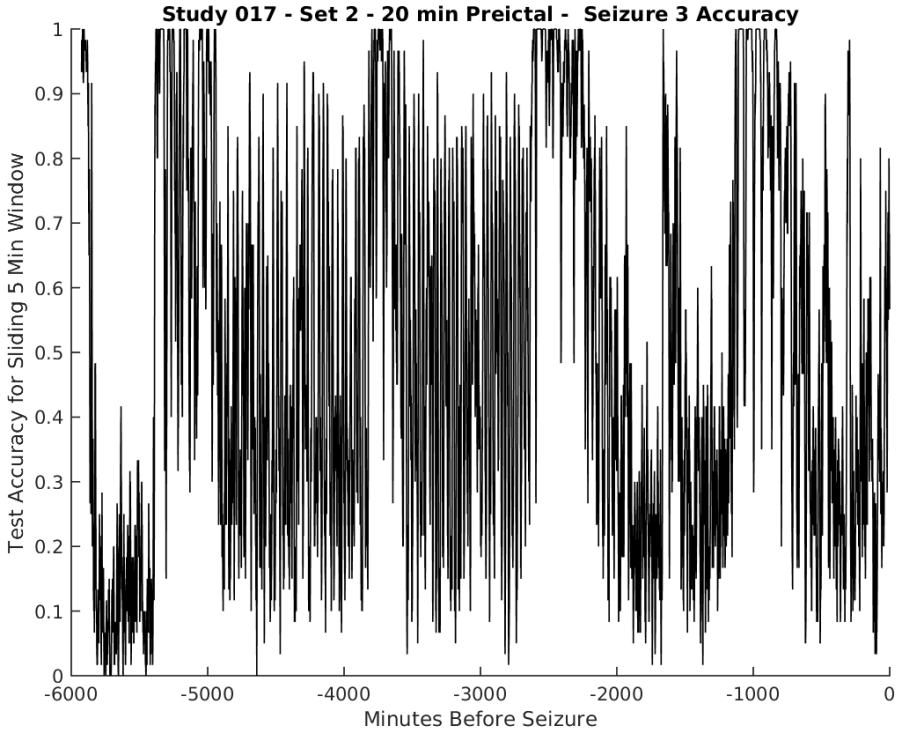
Appendix E – Set 2 – Full Time-Series Accuracy Graphs

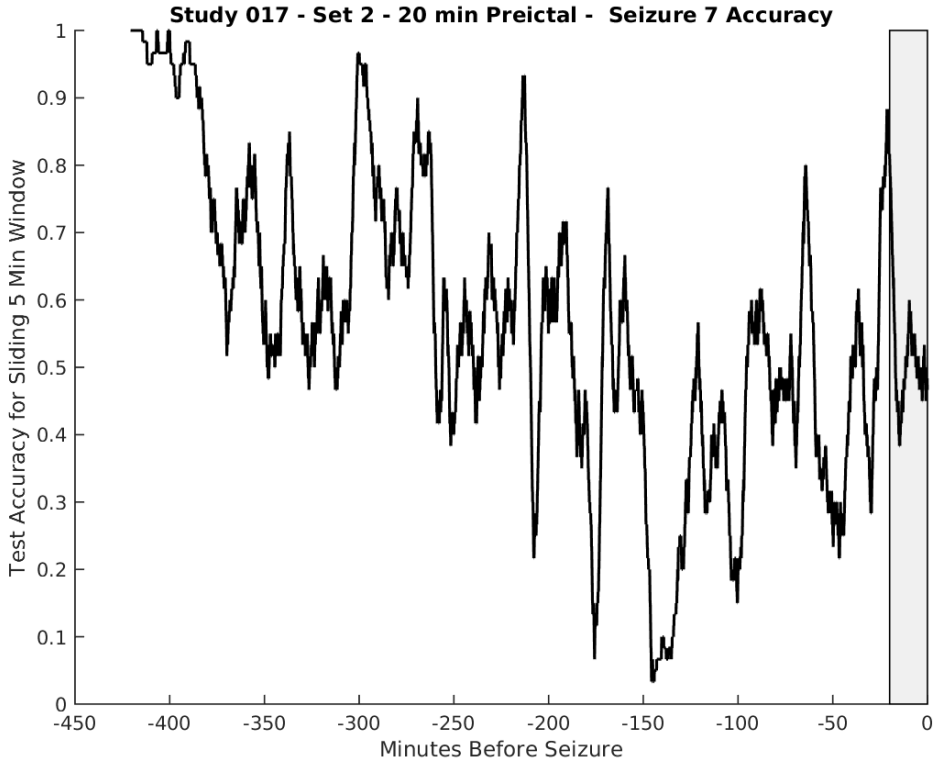


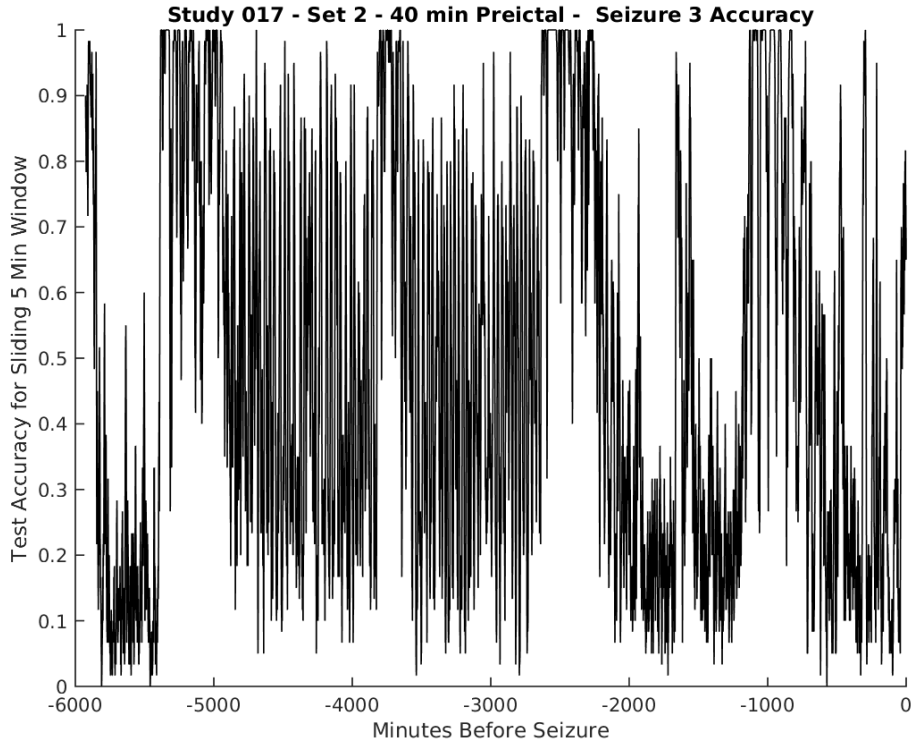


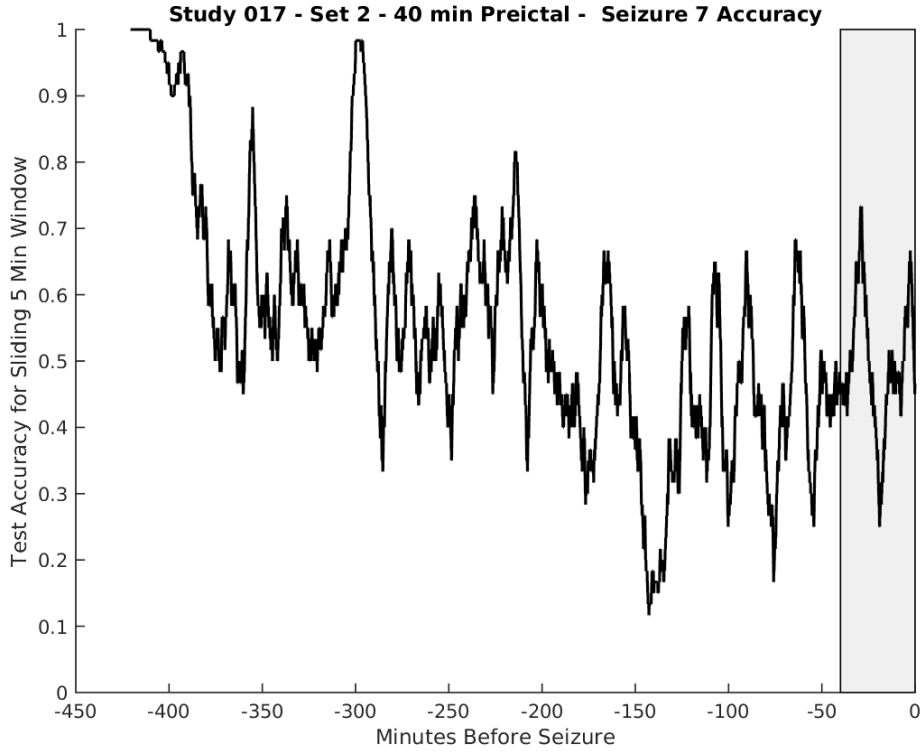

Study 017 - Set 2 - 10 min Preictal - Seizure 3 Accuracy
Test Accuracy for Sliding 5 Min Window
Minutes Before Seizure
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
-6000
-5000
-4000
-3000
-2000
-1000
0


Study 017 - Set 2 - 10 min Preictal - Seizure 7 Accuracy
Test Accuracy for Sliding 5 Min Window
Minutes Before Seizure
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
-450
-400
-350
-300
-250
-200
-150
-100
-50
0


Study 017 - Set 2 - 20 min Preictal - Seizure 3 Accuracy
Test Accuracy for Sliding 5 Min Window
Minutes Before Seizure
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
-6000
-5000
-4000
-3000
-2000
-1000
0


Study 017 - Set 2 - 20 min Preictal - Seizure 7 Accuracy
Test Accuracy for Sliding 5 Min Window
Minutes Before Seizure
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
-450
-400
-350
-300
-250
-200
-150
-100
-50
0


Study 017 - Set 2 - 40 min Preictal - Seizure 3 Accuracy
Test Accuracy for Sliding 5 Min Window
Minutes Before Seizure
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
-6000
-5000
-4000
-3000
-2000
-1000
0


Study 017 - Set 2 - 40 min Preictal - Seizure 7 Accuracy
Test Accuracy for Sliding 5 Min Window
Minutes Before Seizure
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
-450
-400
-350
-300
-250
-200
-150
-100
-50
0

Bibliography

Adeli, H., Zhou, Z., & Dadmehr, N. (2003). Analysis of EEG records in an epileptic patient using wavelet transform, *123*, 69–87.

Adelson, P. D., Nemoto, E., Scheuer, M., Painter, M., Morgan, J., & Yonas, H. (1999). Noninvasive continuous monitoring of cerebral oxygenation periictally using near-infrared spectroscopy: a preliminary report. *Epilepsia*, *40*(1 1), 1484–1489.

Alvarado-Rojas, C., Valderrama, M., Fouad-Ahmed, A., Feldwisch-Drentrup, H., Ihle, M., Teixeira, C. a, … Le Van Quyen, M. (2014). Slow modulations of high-frequency activity (40-140-Hz) discriminate preictal changes in human focal epilepsy. *Scientific Reports*, *4*, 4545. http://doi.org/10.1038/srep04545

Arabadzisz, D., Antal, K., Parpan, F., Emri, Z., & Fritschy, J.-M. (2005). Epileptogenesis and chronic seizures in a mouse model of temporal lobe epilepsy are associated with distinct EEG patterns and selective neurochemical alterations in the contralateral hippocampus. *Experimental Neurology*, *194*(1), 76–90. http://doi.org/10.1016/j.expneurol.2005.01.029

Bandarabadi, M., Rasekhi, J., Teixeira, C. A., & Dourado, A. (2014). *Optimal preictal period in seizure prediction*.

Bandarabadi, M., Teixeira, C. a, Rasekhi, J., & Dourado, A. (2014). Epileptic seizure prediction using relative spectral power features. *Clinical Neurophysiology : Official Journal of the International Federation of Clinical Neurophysiology*, *126*(2), 237–248. http://doi.org/10.1016/j.clinph.2014.05.022

Baumgartner, C., Serles, W., Leutmezer, F., Pataraia, E., Aull, S., Czech, T., … Podreka, I. (1998). Preictal SPECT in temporal lobe epilepsy: regional cerebral blood flow is increased prior to electroencephalography-seizure onset. *Journal of Nuclear Medicine : Official Publication, Society of Nuclear Medicine*, *39*(6), 978–982. http://doi.org/9627329

Breiman, L. (2001). Random forests. *Machine Learning*, 5–32. http://doi.org/10.1023/A:1010933404324

C.U.R.E. (2015). Citizens United Research in Epilepsy. Retrieved September 2, 2015, from www.cureepilepsy.org

Chen, G. (2014). Automatic EEG seizure detection using dual-tree complex wavelet-Fourier features. *Expert Systems with Applications*, *41*(5), 2391–2394. http://doi.org/10.1016/j.eswa.2013.09.037

Delamont, R. S., Julu, P. O., & Jamal, G. a. (1999). Changes in a measure of cardiac vagal activity before and after epileptic seizures. *Epilepsy Research*, *35*(2), 87–94. http://doi.org/10.1016/S0920-1211(98)00100-4

Ernfors, P., Bengzon, J., Kokaia, Z., Persson, H., & Lindvall, O. (1991). Increased levels of messenger RNAs for neurotrophic factors in the brain during kindling epileptogenesis. *Neuron*, *7*(1), 165–176. http://doi.org/10.1016/0896-6273(91)90084-D

Federico, P., Abbott, D. F., Briellmann, R. S., Harvey, a. S., & Jackson, G. D. (2005). Functional MRI of the pre-ictal state. *Brain*, *128*(8), 1811–1817. http://doi.org/10.1093/brain/awh533

Fisher, R. S., Van Emde Boas, W., Blume, W., Elger, C., Genton, P., Lee, P., & Engel, J. (2005). Epileptic seizures and epilepsy: Definitions proposed by the International League Against Epilepsy (ILAE) and the International Bureau for Epilepsy (IBE). *Epilepsia*, *46*(4), 470–472. http://doi.org/10.1111/j.0013-9580.2005.66104.x

Fisher, R. S., Vickrey, B. G., Gibson, P., Hermann, B., Penovich, P., Scherer, A., & Walker, S. (2000). The impact of epilepsy from the patient's perspective I. Descriptions and subjective perceptions. *Epilepsy Research*, *41*(1), 39–51. http://doi.org/10.1016/S0920-1211(00)00126-1

Fountas, K. N., Smith, J. R., Murro, A. M., Politsky, J., Park, Y. D., & Jenkins, P. D. (2005). Implantation of a closed-loop stimulation in the management of medically refractory focal epilepsy: A technical note. *Stereotactic and Functional Neurosurgery*, *83*(4), 153–158. http://doi.org/10.1159/000088656

Ghosh-dastidar, S., Adeli, H., & Dadmehr, N. (2010). Principal Component Analysis-Enhanced Cosine Radial Basis Function Neural Network. *Automated EEG-Based Diagnosis of Neurological Disorders*, *55*(2), 163–182. http://doi.org/doi:10.1201/9781439815328-c9

Iasemidis, L. D., Shiau, D. S., Sackellares, J. C., Pardalos, P. A., & Prasad, A. (2004). Dynamical resetting of the human brain at epileptic seizures: Application of nonlinear dynamics and global optimization techniques. *Ieee Transactions on Biomedical Engineering*, *51*(3), 493–506. http://doi.org/10.1109/TBME.2003.821013

Iasemidis, L. D., Shiau, D.-S., Chaovalitwongse, W., Sackellares, J. C., Pardalos, P. M., Principe, J. C., … Tsakalis, K. (2003). Adaptive epileptic seizure prediction system. *IEEE Transactions on Bio-Medical Engineering*, *50*(5), 616–627. http://doi.org/10.1109/TBME.2003.810689

Ihle, M., Feldwisch-Drentrup, H., Teixeira, C. a, Witon, A., Schelter, B., Timmer, J., & Schulze-Bonhage, A. (2012). EPILEPSIAE - a European epilepsy database. *Computer Methods and Programs in Biomedicine*, *106*(3), 127–38. http://doi.org/10.1016/j.cmpb.2010.08.011

Kerem, D. H., & Geva, a. B. (2005). Forecasting epilepsy from the heart rate signal. *Medical and Biological Engineering and Computing*, *43*(2), 230–239. http://doi.org/10.1007/BF02345960

Kossoff, E. H., Ritzl, E. K., Politsky, J. M., Murro, A. M., Smith, J. R., Duckrow, R. B., … Bergey, G. K. (2004). Effect of an External Responsive Neurostimulator on Seizures and Electrographic Discharges during Subdural Electrode Monitoring, *45*(12), 1560–1567.

Krishnan, B., Vlachos, I., Faith, A., Mullane, S., Williams, K., & Iasemidis, L. (2015). Spatiotemporal dynamics of interictal spikes. *Journal of the Mississippi Academy of Sciences*, *60*(Supplemental), 207–210.

L. D. Iasemidis, J. C. Principe, J. C. S. (2000). Measurement and Quantification of Spatiotemporal Dynamics of Human Epileptic Seizures. *Nonlinear Biomedical Signal Processing, Dynamic Analysis and Modeling*, 294–318. http://doi.org/10.1109/9780470545379.ch12

Lopes da Silva, F., Blanes, W., Kalitzin, S. N., Parra, J., Suffczynski, P., & Velis, D. N. (2003). Epilepsies as dynamical diseases of brain systems: basic models of the transition between normal and epileptic activity. *Epilepsia*, *44 Suppl 1*, 72–83. http://doi.org/12005 [pii]

Mathworks. (2015). MATLAB and Statistics Toolbox. Natick, Massachusetts: The MathWorks, Inc.

Mayo Clinic. (2015). Mayo Clinic. Retrieved September 2, 2015, from www.mayoclinic.org

Mayo Clinic, University, P. S., & Stroke, N. I. of N. D. and. (2014). IEEG-Portal. Retrieved January 1, 2015, from www.ieeg.org

Mirowski, P., Madhavan, D., LeCun, Y., & Kuzniecky, R. (2009). Classification of patterns of EEG synchronization for seizure prediction. *Clinical Neurophysiology*, *120*(11), 1927–1940. http://doi.org/10.1016/j.clinph.2009.09.002

Moghim, N., & Corne, D. W. (2014). Predicting epileptic seizures in advance. *PLoS ONE*, *9*(6). http://doi.org/10.1371/journal.pone.0099334

Mormann, F., Andrzejak, R. G., Elger, C. E., & Lehnertz, K. (2007). Seizure prediction: The long and winding road. *Brain*, *130*(2), 314–333. http://doi.org/10.1093/brain/awl241

Mormann, F., Kreuz, T., Rieke, C., Andrzejak, R. G., Kraskov, A., David, P., … Lehnertz, K. (2005). On the predictability of epileptic seizures. *Clinical Neurophysiology*, *116*(3), 569–587. http://doi.org/10.1016/j.clinph.2004.08.025

Netoff, T., Park, Y., & Parhi, K. (2009). Seizure prediction using cost-sensitive support vector machine. In *Proceedings of the 31st Annual International Conference of the IEEE Engineering in Medicine and Biology Society: Engineering the Future of Biomedicine, EMBC 2009* (pp. 3322–3325). http://doi.org/10.1109/IEMBS.2009.5333711

Novak, V., Reeves, a L., Novak, P., Low, P. a, & Sharbrough, F. W. (1999). Time-frequency mapping of R-R interval during complex partial seizures of temporal lobe origin. *Journal of the Autonomic Nervous System*, *77*(2-3), 195–202. http://doi.org/10.1016/S0165-1838(99)00044-2

Ocak, H. (2009). Automatic detection of epileptic seizures in EEG using discrete wavelet transform and approximate entropy. *Expert Systems with Applications*, *36*(2 PART 1), 2027–2036. http://doi.org/10.1016/j.eswa.2007.12.065

Osorio, I., Frei, M. G., Sunderam, S., Giftakis, J., Bhavaraju, N. C., Schaffner, S. F., & Wilkinson, S. B. (2005). Automated seizure abatement in humans using electrical stimulation. *Annals of Neurology*, *57*(2), 258–268. http://doi.org/10.1002/ana.20377

Park, Y., Luo, L., Parhi, K. K., & Netoff, T. (2011). Seizure prediction with spectral power of EEG using cost-sensitive support vector machines. *Epilepsia*, *52*(10), 1761–1770. http://doi.org/10.1111/j.1528-1167.2011.03138.x

Perez-Velazquez, J. L., Valiante, T. a, & Carlen, P. L. (1994). Modulation of gap junctional mechanisms during calcium-free induced field burst activity: a possible role for electrotonic coupling in epileptogenesis. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *14*(7), 4308–4317.

Phelan, K. D., Shwe, U. T., Abramowitz, J., Birnbaumer, L., & Zheng, F. (2014). Critical role of canonical transient receptor potential channel 7 in initiation of seizures. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(31), 1–6. http://doi.org/10.1073/pnas.1411442111

Polat, K., & Güneş, S. (2007). Classification of epileptiform EEG using a hybrid system based on decision tree classifier and fast Fourier transform. *Applied Mathematics and Computation*, *187*, 1017–1026. http://doi.org/10.1016/j.amc.2006.09.022

Ravizza, T., Gagliardi, B., Noé, F., Boer, K., Aronica, E., & Vezzani, A. (2008). Innate and adaptive immunity during epileptogenesis and spontaneous seizures: Evidence from experimental models and human temporal lobe epilepsy. *Neurobiology of Disease*, *29*(1), 142–160. http://doi.org/10.1016/j.nbd.2007.08.012

Ren, L., Terada, K., Baba, K., Usui, N., Umeoka, S., Usui, K., … Inoue, Y. (2011). Ictal very low frequency oscillation in human epilepsy patients. *Annals of Neurology*, *69*(1), 201–206. http://doi.org/10.1002/ana.22158

Roth, J. L., & Blum, A. S. (2014). Status Epilepticus Treatment & Management. Retrieved July 12, 2015, from http://emedicine.medscape.com

Srinivasan, V., Eswaran, C., & Sriraam, N. (2007). Approximate entropy-based epileptic EEG detection using artificial neural networks. *IEEE Transactions on Information Technology in Biomedicine*, *11*(3), 288–295. http://doi.org/10.1109/TITB.2006.884369

Subasi, A., & Gursoy, M. I. (2010). EEG signal classification using PCA, ICA, LDA and support vector machines. *Expert Systems with Applications*, *37*(12), 8659–8666. http://doi.org/10.1016/j.eswa.2010.06.065

Tzallas, A. T., Tsipouras, M. G., & Fotiadis, D. I. (2007). Automatic seizure detection based on time-frequency analysis and artificial neural networks. *Computational Intelligence and Neuroscience*, *2007*, 80510. http://doi.org/10.1155/2007/80510

Tzallas, A. T., Tsipouras, M. G., & Fotiadis, D. I. (2009). Epileptic seizure detection in EEGs using time-frequency analysis. *IEEE Transactions on Information Technology in Biomedicine : A Publication of the IEEE Engineering in Medicine and Biology Society*, *13*(5), 703–710. http://doi.org/10.1109/TITB.2009.2017939

Venkataraman, V., Vlachos, I., Faith, A., Krishnan, B., Tsakalis, K., Treiman, D., & Iasemidis, L. (2014). Brain Dynamics Based Automated Epileptic Seizure Detection, 946–949.

Vezzani, A., Moneta, D., Richichi, C., Aliprandi, M., Burrows, S. J., Ravizza, T., … Grazia De Simoni, M. (2002). Functional role of inflammatory cytokines and antiinflammatory molecules in seizures and epileptogenesis. *Epilepsia*, *43*(SUPPL. 5), 30–35. http://doi.org/10.1046/j.1528-1157.43.s.5.14.x

Weinand, M. E., Carter, L. P., el-Saadany, W. F., Sioutos, P. J., Labiner, D. M., & Oommen, K. J. (1997). Cerebral blood flow and temporal lobe epileptogenicity. *Journal of Neurosurgery*, *86*(2), 226–32. http://doi.org/10.3171/jns.1997.86.2.0226

Weir, B. (1965). The morphology of the spike-wave complex. *Electroencephalography and Clinical Neurophysiology*, *19*(3), 284–290. http://doi.org/10.1016/0013-4694(65)90208-7

Williamson, J., Bliss, D., & Browne, D. (2011). Epileptic seizure prediction using the spatiotemporal correlation structure of intracranial EEG., 665–668.

Worrell, G. a., Parish, L., Cranstoun, S. D., Jonas, R., Baltuch, G., & Litt, B. (2004). High-frequency oscillations and seizure generation in neocortical epilepsy. *Brain*, *127*(7), 1496–1506. http://doi.org/10.1093/brain/awh149

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 074-0188

1. REPORT DATE *(DD-MM-YYYY)*	**2. REPORT TYPE**	**3. DATES COVERED** *(From – To)*
24-03-2016	Master's Thesis	September 2014 – March 2016

TITLE AND SUBTITLE

EEG-Based Classification and Advanced Warning of Epileptic Seizures

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Clisby, Lauren E., 2nd Lieutenant, USAF

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAMES(S) AND ADDRESS(S)

Air Force Institute of Technology
Graduate School of Engineering and Management (AFIT/EN)
2950 Hobson Way, Building 640
WPAFB OH 45433-8865

8. PERFORMING ORGANIZATION REPORT NUMBER

AFIT-ENG-MS-16-M-097

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Intentionally left blank

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT

DISTRUBTION STATEMENT A. APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

Epilepsy is the second most common neurological disease, after stroke. Epileptics may suffer hundreds of seizures per day, yet one is enough to put a person in constant fear of the next. The sudden and unexpected onset of seizures has debilitating and sometimes fatal consequences. The development of a real-time seizure prediction and alerting device would greatly improve epileptics' quality of life. Major challenges for such a device include determining predictive features and discovering the maximum prediction window. Using the novel approach of random forest classification on EEG data, this research investigates the predictive features among the common EEG frequency bands for one patient with partial complex and partial with secondarily generalized seizures. The impact on classifier performance of labeling the transitional brain states is also investigated, using a time-series accuracy graph. Predictive features are found as far as 40 minutes in advance of two seizures, specifically in the beta frequencies of one brain node. The random forest classifier does not perform well, but shows promise for improved performance with minor adjustments in training. The time-series accuracy graphs prove a useful tool for visualization and insight into classifier performance that is lacking in other evaluation methods.

15. SUBJECT TERMS

EEG, FFT, epilepsy, seizure prediction, random forest

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			Dr. Brett Borghetti, AFIT/ENG
U	**U**	**U**	**UU**	91	**19b. TELEPHONE NUMBER** *(Include area code)* (937) 785-6565, ext 4612 (brett.borghetti@afit.edu)

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39-18