Automated Epileptic Seizure Onset Detection

by

Arvind Dorai

A thesis

presented to the University of Waterloo

in fulfillment of the

thesis requirement for the degree of

Master of Applied Science

in

Systems Design Engineering

Waterloo, Ontario, Canada, 2009

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

Epilepsy is a serious neurological disorder characterized by recurrent unprovoked seizures due to abnormal or excessive neuronal activity in the brain. An estimated 50 million people around the world suffer from this condition, and it is classified as the second most serious neurological disease known to humanity, after stroke. With early and accurate detection of seizures, doctors can gain valuable time to administer medications and other such anti-seizure countermeasures to help reduce the damaging effects of this crippling disorder.

The time-varying dynamics and high inter-individual variability make early prediction of a seizure state a challenging task. Many studies have shown that EEG signals do have valuable information that, if correctly analyzed, could help in the prediction of seizures in epileptic patients before their occurrence. Several mathematical transforms have been analyzed for its correlation with seizure onset prediction and a series of experiments were done to certify their strengths. New algorithms are presented to help clarify, monitor, and cross-validate the classification of EEG signals to predict the ictal (*i.e.* seizure) states, specifically the preictal, interictal, and postictal states in the brain. These new methods show promising results in detecting the presence of a preictal phase prior to the ictal state.

Keywords: Epilepsy, Seizure, Ictal, EEG, Prediction, Wavelet, Entropy, Synchronization, Chaos, Coherence, Signals.

Acknowledgements

First and foremost, it is difficult to overstate my gratitude to my supervisor, Dr. Kumaraswamy Ponnambalam. Throughout my thesis-writing period he provided encouragement, sound advice, good teaching, good company, and lots of good ideas. I would have been lost without him. I would also like to thank Tiuley, for reminding me that I am still a musician, and Canmanie, for finding new and interesting ways to keep me alert. Also, a special thanks to Drs. Taufik Valiante and Richard Wennberg of Toronto Western General Hospital for initiating this study.

During my life, I have made several friends that have influenced and shaped me into the individual I am today. I would like to thank Aaron Troisi and Stephanie Chapman for their support and constant supply of great ideas and thoughts. I would also like to thank Tim Benke, who provided good company, great conversation, and the opportunity to cook indian meals completely from scratch. Lastly, I would like to thank Ingrid Elise for her love and determination; she made sure I never felt alone even while feeling the full force of my insanity.

I would like to make a special acknowledgement to all the police officers on I-87 for their totally obvious concern about my driving skills.

I wish to thank my entire extended family for providing a loving and encouraging ennvironment for me. Lastly, and most importantly, I would like to thank both my parents for extending their medical knowledge and critical inquiry to bridge the gaps between the medical and computational aspects of this research. They also raised me, by the way.

Contents

List of Tables

List of Figures

List of Acronyms

Chapter 1

Introduction

Thousands of years ago, ancient civilizations were mystified by the epileptic condition. Ancient Greeks thought that one got epilepsy by offending the moon goddess Selene. Ancient Romans believed that epilepsy came from the demons. Ancient Babylonians thought similarly, but that different spirits caused the different types of seizures. All came up with their own explanations as to who or what the cause was because none of them could comprehend it [12].

1.1   Historical Background

In 400 BC, Hippocrates, the "Father of Medicine", wrote *On the Sacred Disease*, where he refuted the idea that epilepsy was a curse or prophetic power:

> "*It is thus with regard to the disease called Sacred: It appears to me to be nowise more divine nor more sacred than other diseases, but has a natural cause like other affections.*" [77]

The word "epilepsy" is derived from the Greek "*epilepsia*", which means, "to take hold of" or "to seize". The first documented incidence of epilepsy was more

than 3,000 years ago in ancient Babylonia, where the condition was referred to as "*miqtu*". In both Ancient Greece and Babylonia, people saw epilepsy as a supernatural, but perhaps holy phenomenon [68].

Epilepsy has been viewed differently by various cultures. During World War II, the Nazi Eugenics Laws mandated that persons with epilepsy must be sterilized [59]. Epileptics were highly discriminated against and treated very differently than the average person. In fact, this type of discrimination still exists in today's society [5, 28].

In modern-day Africa, people with epilepsy are finding it difficult to cope with the social consequences of the disorder as well as the exclusion from education and employment. This is further compounded by the fact that a large proportion of people there with epilepsy are not taking Anti-Epileptic Drugs (AEDs). The lack of health services makes it very difficult for people with epilepsy to be a functional part of society [63].

Discrimination is equally present in western culture as well. Up until 1980, various states in America forbade epileptics to marry [5]. The Americans With Disabilities Act of 1990 made it illegal to discriminate against people with epilepsy in the workplace [28]. Those whose seizures can be effectively controlled are not considered disabled under the act. The ability to predict and control seizure activity would dramatically improve the lifestyle of people with epilepsy, hence the motivation of this research.

1.2 Epilepsy Today

Epilepsy is classified as the second most serious neurological condition known to man, after stroke. It affects nearly 50 million people around the world, which is

approximately 1% of the world's population. As many as one out of ten people will have a seizure sometime during their lives, but the majority will not have epilepsy as the underlying cause of the convulsions lie outside the brain. Approximately one in 50 people will have some form of epilepsy at some point in their life. About 75 people are diagnosed with epilepsy every day. Only about 3-5% of people with epilepsy will be affected by flashing lights (photosensitive epilepsy). It's a common misconception that all epileptics are affected by quick visual stimuli; this is actually quite uncommon. Only about two out of 10,000 people in the general population have this condition [65].

At the present time, it is estimated that from 1.5 to 2 million people in the United States alone have an active form of epilepsy [36]. Today, there are many people living with epilepsy whose lives have been greatly improved by modern science. In fact, many people who have had epilepsy have also been able to have successful careers or become very famous. Alfred Nobel [23], founder of the Nobel Prize and the inventor of dynamite, had epilepsy. Other famous/well-known people with epilepsy include Fyodor Dostoevsky [24], Neil Young [7], George Gerswhin [71], Philip K. Dick [69], Napoleon I [22], Muhammad [15], and Joan of Arc [23]. These people, and many others, give the word 'disability' a different meaning.

1.3   Epileptic Seizure Prediction

One of the most debilitating aspects of epilepsy is the lack of warning before a seizure occurs. As a result, common day to day tasks such as driving or using a knife become much more hazardous. In some patients, seizures occur hundreds of times a day, but they can also be as infrequent as once every few years. These limitations can have a debilitating affect on quality of life, and hinder basic everyday

activities.

The ability to predict seizures could lead the way to novel diagnostic and therapeutic methods for the treatment of patients with epilepsy [37]. Real-time prediction of epileptic state transitions and early onset prediction provide time to administer preventive interventions possibly terminating the seizure before it happens. Also, a lead-time could give epileptic patients enough time to remove themselves from harms way. It would give them the ability to do things that people without epilepsy take for granted.

1.4   Goal: Epileptic Seizure Control

Once a robust seizure prediction algorithm is in place, a device that could somehow control the seizure would be possible. An implanted system that activates a mini-AED [65] delivery system to deliver the medications directly into the epileptogenic focus would be feasible. A so-called "brain pacemaker" which appropriately stimulates the vagus nerve [61] at the predicted onset time could also reduce the frequency and ferocity of epileptic seizures. However, if the true prediction rate is too high and the false prediction rate is too low, these medications and therapeutic treatments would be administered too often. This would result in a whole slew of clinical side-effects, most of which would be neuropsychological [65]. A careful balance of all the influential parameters is required for the optimal automated seizure prediction algorithm which would put us on the path to the ultimate goal: Seizure Control [56].

1.5 Organization

In *Chapter 1*, a brief introduction to the problem is given, and the goals of this thesis are provided.

Chapter 2 provides some background information on the epileptic condition as well as some information on Electroencephalogram (EEG) signals in relation to epilepsy. An overview of the time series analysis techniques is given. Also, a literature review of the current progress in the field of seizure prediction is presented.

Chapter 3 outlines the proposed system for seizure state prediction. Also presented here are some new methods and algorithms that were created during the course of this research to contribute to the field of seizure prediction.

Chapter 4 discusses the results of the seizure prediction algorithms that were proposed in Chapter 3. The optimized feature set is presented for different types of EEG signals, and the strength of the dynamic real-time classification system is discussed.

Chapter 5 summarizes the conclusions of this research and provides a discussion about the research methods. Possible ways to improve the results and algorithms are given as possible future work.

Chapter 2

Background

This chapter gives an overview of epilepsy and provides its relation to EEG signals. The different states of a seizure are provided along with a discussion about the brain waves during each period. In the following section, all of the numerical methods that were implemented in the system are provided. Lastly, a literature review of the current state of epileptic seizure prediction is provided.

2.1   The Neurological Condition: Epilepsy

Epilepsy is a serious neurological disorder characterized by recurrent unprovoked seizures due to abnormal or excessive neuronal activity in the brain. This condition is characterized by chronic abnormal bursts of electrical spikes in the brain. The region of seizure generating tissue, known as the *epileptogenic focus*, can be the result of structural abnormalities in the brain which may or may not be genetic. In cases where the cause is known, the epileptic condition is referred to as *symptomatic epilepsy*. In cases where there is no identifiable cause, but by deduction a genetic basis is presumed, the condition is referred to as *idiopathic epilepsy*. Cases that don't fit into either of these two categories are considered to be *cryptogenic epilepsy*

[14].

Epileptic seizures are characterized by uncontrollable movements such as shaking of the arms or legs, known as convulsions. Some may lose consciousness, which may consist of a complete collapse or the patient simply gazing into space. Fainting spells with incontinence, followed by excessive fatigue, is common in more serious types of seizures. Distorted perceptions, odd sounds, and sudden feelings of fear for no apparent reason are characteristic of the "aura", which is felt just prior to the seizure. These "auras" can happen anywhere from a day before the seizure to just a few seconds prior. A select few can predict that they will have a seizure by themselves just by understanding the "aura" [55].

Seizures are usually treated with medications known as Anti-Epileptic Drugs (AEDs). AEDs attempt to stop the occurrence of seizures, but do not serve as a cure for epilepsy. With the correct AEDs, approximately 70% of people with epilepsy could have their seizures controlled or stopped [4]. For people who cannot control their seizures with AEDs, epileptic surgery is quite common. This type of surgery involves the resection of sections of the brain that comprise the epileptogenic focus. It has been shown to greatly reduce the severity of epileptic seizures in patients, or in some cases, stop them completely [55]. Vagus Nerve Stimulation (VNS) is a treatment of epilepsy which attenuates seizure frequency, severity, and duration by chronic intermittent stimulation of the vagus nerve. It is intended for use as an adjunctive treatment with AED medications. The patient is implanted with a VNS therapy system which directly stimulates the vagus nerve at predetermined time intervals [51].

2.2 Epilepsy & EEG signals

An electroencephalogram (EEG) is a recording of the electrical activity in the brain at different frequencies. It was first developed in 1924 by Hans Berger, a german psychiatrist, who revealed the practical and diagnostic use of this test [21]. Special sensors are placed strategically around the head that are connected to a machine which records the electrical impulses, either on screen or on paper. Trained neurologists are able to look at the different frequencies in the EEG and recognize patterns in it which provide information about the epileptic condition [14].

2.2.1 Brain Waves

Raw EEG signals are usually described in terms of the four basic brain waves: Alpha [7.5-13] Hz, Beta [13-30] Hz, Delta [0-3.5] Hz, and Theta [3.5-7.5] Hz. These bands represent the most prominent activity in the brain [58].

Alpha Waves

Alpha waves are comprised of brain signals of the frequency range [7.5-13] Hz. Healthy alpha waves promote mental resourcefulness, aid in the ability to mentally coordinate, and enhance the overall sense of relaxation and fatigue. With healthy alpha waves, one can move quickly and efficiently to accomplish tasks at hand. Alpha waves appear to bridge the conscious and subconscious mind and are quite prominent in normal relaxed adults. It is present throughout most of an individual's life, but especially beyond the 13th year when it dominates the resting tracing (normal EEG tracings) [58].

The most important of the recorded waves in a normal adult EEG are the

occipital alpha waves, which are best obtained from the back of the head when the subject is resting quietly with the eyes closed, but not asleep. Interference in this band can be caused by opening the eyes or excitement.

These waves are blocked by both excitement or by opening the eyes.

Beta Waves

Beta waves are considered "fast" brain activity (above 13Hz). These waves are seen on both sides of the brain in a symmetrical distribution and is most evident in the frontal region of the brain. This band may be absent or reduced in areas with cortical damage. It is generally the normal rhythm in those who are alert, anxious, or who have their eyes open [58].

Delta Waves

Delta waves are comprised of the lowest frequencies in the brain, specifically from [0-3.5] Hz. They usually occur in the deep sleep state and in some abnormal processes in the brain. These waves are the dominant rhythm in infants up to one year of age and are present in stages III and IV of sleep. The delta band tends to be the highest in amplitude and also the slowest of the waves. Increasing delta waves mean a decrease in our awareness of the physical world, which is a characteristic property of seizure activity. Our unconscious mind is also represented through delta waves. Peak performers decrease their Delta waves when high focus and peak performance are required [58].

Theta Waves

Theta activity is classified as rhythmic, slow waves from the frequency range of [3.5-7.5] Hz. It has connection with creativity, intuition, learning, daydreaming,

fantasizing, and is a repository for memories, emotions, and sensations. The presence of these waves reflect the state between wakefulness and sleep. Theta waves are abnormal in awake adults, but are perfectly normal in children up to 13 years old and during sleep. Theta waves are also observed during anxiety, pain, behavioral activation, and behavioral inhibition. The appearance of excess Theta waves are an indicator of abnormalities in the brain [58].

2.2.2   Epileptic States

The different stages of an epileptic seizure are referred to as *ictal* states. These states represent the different stages of an epileptic seizure in its most general sense [52].

1) Interictal State

The *interictal* state refers to the normal resting state containing no seizure activity, but the EEG is still characterized by the epileptic condition (irregular neuronal activity). Given the possibility of seizures, the chronic interictal period is interesting because of the presence of natural homeostatic mechanisms that prevent seizure generation. It is unsure which factors or mechanisms try to maintain homeostasis in the brain, and it is also unsure if these mechanisms differ for various types of seizures and epileptic syndromes. This period comprises more than 99% of patients' lives. In this way, the interictal period can be used by neurologists to diagnose an epileptic condition. The EEG tracings would normally exhibit small spikes and other abnormalities known by neurologists as subclinical seizures. These are not real seizures, but rather little hints from the brain that something is abnormal [52].

2) Preictal State [Seizure Onset]

The *preictal* state refers to a period of time occurring before a seizure, but does not refer to the normal state of the brain. This state defines that a seizure is going to occur within a certain period of time [52]. The presence of a preictal period is still being debated by several researchers. Lehnertz and Litt state that in some conditions, the transition can take a considerable amount of time, opening the potential for the application of electrophysiological techniques to predict seizure onset anywhere from minutes to hours before occurrence [39].

Onset of a clinical seizure is defined as the time at which the transition between the interictal and preictal state occurs. It is characterized by a sudden change in the frequency characteristics of the EEG. The alpha band has a tendency to decrease in frequency and increase in amplitude. The transition from the preictal to the ictal state, for a focal epileptic seizure, consists of a gradual change from chaotic to ordered waveforms. In each type of epilepsy, the transitional period may differ, and the transition can have significantly different characteristics [52].

3) Ictal State to Seizure Termination

The period of time during which the seizure is in its activation period is referred to as the *ictal* state. It is characterized by an EEG tracing that exhibits significantly higher amplitudes and frequencies. There is an immediate alteration in synchronization and rhythmicity that takes place across widespread areas of the cerebral cortex [57]. The patterns that are normally seen throughout the resting tracing suddenly become extremely erratic and unpredictable. Loss of consciousness and involuntary muscle twitching during this period is very common. Other symptoms such as incontinence are also common during this state. The patient typically has no control over their body at this point and convulsions tend to be prominent [52].

4) Postictal State

The end of an epileptic seizure represents a transition from the ictal state back to an individual's normal, or interictal state. This is referred to as the *postictal* state and signifies the recovery period of the brain. Focal or generalized neurological deficit, ranging from postictal depression to aphasia or paralysis is prevalent during this state. This period is associated with a difficulty in thinking clearly and a variety of other cognitive defects. The postictal state could last from seconds to hours depending on the severity of the seizure and the efficacy of the AEDs. Disturbances or aftershocks are seen in the EEG, which may just be the presence of natural mechanisms acting to terminate the seizure and restore homeostasis [52].

Often postictal deficits are a consequence of the natural mechanisms that act to terminate a seizure suggesting that interventions designed to exploit these same homeostatic events could exacerbate postictal dysfunction [52].

Attention and concentration is generally very difficult during this period. Poor short term memory and decreased verbal and interactive skills are noticeable. Postictal migraine headaches are very common due to the pressure resulting from cerebral edema. At this point, patients are unaware that they have had a seizure, but these symptoms are evidence enough for an experienced epileptic [52].

2.3   Time Series Analysis

A time series is an ordered sequence of values of a variable at equally spaced time intervals. Time series analysis is comprised of methods that attempt to identify the nature of a phenomenon represented by a sequence of observations. This requires that patterns of observed time series data are properly identified and mathematically described.

2.3.1 Sampling

Sampling is the operation of transforming a time signal from the continuous to the discrete time domain. Let $x_a(t)$ be a continuous-time signal that is sampled uniformly at $t = nT$, generating the sequence $x[n]$ where

$$x[n] = x_a(nT), \quad n = 0, 1, 2, 3, \ldots \tag{2.1}$$

with T being the *sampling period*. The reciprocal of T is referred to as the sampling frequency f_s. In order to reconstruct the original signal completely and accurately, the *Nyquist-Shannon Sampling theorem* must be adhered to. This theorem states that a band-limited signal can be reconstructed perfectly from the discrete sampled signal given that the sampling rate is more than twice the frequency. Therefore, the frequency that is half of the sampling rate f_s, also known as the *folding frequency* or the *Nyquist frequency*, is the highest frequency that can be represented in a sampled signal without being affected by aliasing or under-sampling. In equation (2.2), also known as the *Nyquist Condition*, the *Nyquist frequency* is denoted as Ω_m.

$$\Omega_T \geq 2\Omega_m, \quad where \quad \Omega_T = 2\pi/T \tag{2.2}$$

"*Aliasing*" is the phenomenon in which high-frequency components of a time function can translate into low frequencies if the sampling rate is too low. Thus, any component in $x(t)$ higher than the folding frequency is aliased (or folded) into the frequencies that are below the folding frequency [16]. Figure 2.1 illustrates the effects of aliasing.

As seen in Figure 2.1, the reconstructed signal contains frequencies that are lower than those in the original signal due to aliasing. The sampled signal does not

Effects of Aliasing
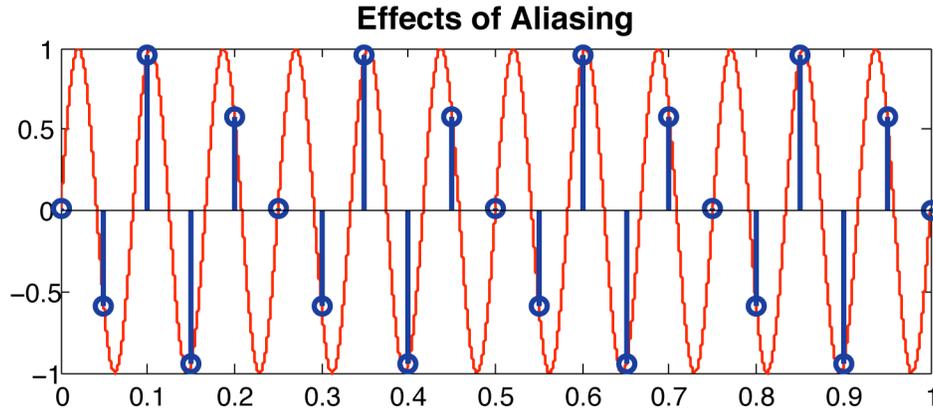
Figure 2.1: Effects of aliasing on a discretely sampled signal.

satisfy the *Nyquist Condition*.

Brain waves are an example of a continuous time signal which, when sampled with an EEG machine, becomes a discrete time series in the form of equation (2.1). In general, when any signal is discretely sampled, 'information' is lost from the resulting time series. The frequency of the sampled signal must be high enough to preserve the frequency information from within the signal. The problem with EEG signals is that it is difficult to tell what the appropriate sampling frequency is because of its unpredictable nature. According to Castellaro, most of the prominent frequencies in EEG signals lie between 0 and 46Hz [6]. This means that the lowest feasible sampling rate would be $f_s = 92$Hz. To satisfy the *Nyquist condition*, all EEG signals used in testing this prediction system were collected at a minimum of 250Hz.

2.3.2   Windowing

A window function is a function that defines a signal within a defined interval and is zero everywhere outside. It is multiplied by the original signal to give a clearer perspective of the signal in the frequency domain. The most basic window function

is the *Rectangular Window*, also known as the *Dirichlet Window*, stated as:

$$w_R[n] = \begin{cases} 1, & -M \geq n \geq M, \\ 0, & \text{otherwise.} \end{cases} \tag{2.3}$$

EEG signals cannot be classified at every discrete point collected in an EEG signal. There is just not enough information at each time point. The system must be given a window of the signal with enough information to make a sensible judgement about the state of the signal at each moment in time. The length of the window must also be at least the Nyquist rate or else it will not be able to recognize the relevant frequencies that are present. But it is also unclear how much of the signal is necessary to give the system enough information to properly identify classes within the signal.

EEG signals tend to be arbitrary by nature, and with some epileptic states, the frequency component of the signal can vary with time depending on the severity of the condition. During the ictal period, the frequency components of an EEG signal become extremely erratic and unpredictable. Due to this irregularity, a rectangular window is more appropriate for frequency analysis of these signals.

2.3.3 Frequency Domain Techniques

2.3.3.1 Discrete Fourier Transform

The Discrete Fourier Transform (DFT) is the link between the time domain representation $f[nT]$ and the frequency representation $F[k\omega_0]$ of a signal that is discrete and periodic in both the time variable nT and the frequency variable $k\omega_0$ [33]. The Discrete Fourier Transform is defined as:

$$X[k] = \sum_{n=1}^{N} x[n] \omega_N^{(n-1)(k-1)} \tag{2.4a}$$

$$\omega_N = e^{j(-2\pi)/N} \tag{2.4b}$$

For each value of k, it looks like the computation of $X[k]$ should require N multiplications, but since $X[k]$ will be a complex number, the computation of an N-point DFT will actually require N^2 multiplications.

The Inverse Discrete Fourier Transform (IDFT) will recover the discrete-time sequence as defined in the following equation.

$$x[n] = \frac{1}{N} \sum_{k=1}^{N} X[k] \omega_N^{-(n-1)(k-1)} \tag{2.5}$$

The Fast Fourier Transform (FFT) is a set of algorithms that permit the fast and accurate computation of the Discrete Fourier Transform. The most common FFT algorithm is the Cooley-Tukey Algorithm [9]. This algorithm is popular because of its ability to rapidly compute the DFT. The discovery of this algorithm contributed substantially to the expansion of DFT applications and also to the development of real-time digital signal processing techniques.

2.3.3.2   Power Spectra Analysis

The Power Spectra of a signal shows the frequencies that contain the signal's power and is represented by a distribution of power values as a function of frequency. The equation for the power spectra of a given signal is defined as:

$$P(f) = \lim_{T \to \infty} \frac{1}{T} \left\{ E[|F_T(f)|^2] \right\} \tag{2.6}$$

where $F_T(f)$ is the Fourier transform of a signal, $E[\cdot]$ denotes the expectation or average, f is the frequency in hertz, and T is the period. The FFT is also helpful in reducing the number of computations needed to compute the power spectra of a given signal [75].

2.3.3.3 Mel Frequency Cepstral Coefficients

The Mel-Frequency Cepstrum Coefficients (MFCC) are very effective in applications such as automatic speech recognition and in the modeling of pitch and frequency content of audio signals [76]. These coefficients $c[k]$ are calculated from the FFT power coefficients and are based on the spectral information of a given signal. These coefficients are unique in the sense that they are modeled to capture only the most important parts of the signal to model the spectral energy distribution. In Mel-Frequency Cepstrum Analysis, the frequency bands are based on the *mel-scale*, which attempts to model the response of the human auditory system. The cepstral coefficients are computed in the equation given by

$$c[k] = IDFT\{log|DFT\{x[n]\}|\} \tag{2.7}$$

The non-linear mel-scale is defined such that a tone of 1000 Hz equals 1000 mels. A tone which is aurally perceived twice as high will have a Mel value that is twice as high. The mel-scale is approximately linear below 1000 Hz and logarithmic otherwise. In this way, the filter bank can be constructed so that the first 13 filters are linearly spaced with a frequency interval of 133.33 Hz between the center frequency of each filter bank [40]. The subsequent 27 filters are spaced logarithmically by frequency with a factor of 1.0711703.

The algorithm for calculating the Mel-Frequency Cepstral Coefficients is given as follows:

1. Decompose the signal into windows using the Hamming window.

2. Compute the DFT of the resulting signal.

3. Find the magnitude of the resulting DFT.

4. Restructure the FFT output into the MFCC filter bank outputs, *i.e.* smooth the spectra and scale it with respect to the mel-scale.

5. Apply a logarithm with base 10 to the outputs.

6. Compute the Discrete Cosine Transform (DCT), de-correlating the Mel-spectral vectors and reducing dimensionality.

2.3.3.4   The Discrete Wavelet Transform

The Discrete Wavelet Transform (DWT) of a signal $x[n]$ is calculated by passing the signal through a series of filters. The signal is simultaneously decomposed using a low pass and a high pass filter. The outputs of the high pass filter give the detail coefficients, and the outputs of the low pass filter give the approximation coefficients. Since, for every succeeding level, half the frequencies are removed from the sample, half the samples can be discarded according to the *Shannon-Nyquist Theorem*. In this sense, the filter outputs are all down-sampled by 2. These filters are known together as the *Quadrature Mirror Filter (QMF)*. The process of generating the set of subsequences $u_r[n]$ for $0 \leq r \leq L$ is known as the Discrete Wavelet Transform, given as

$$u_r[n] = \begin{cases} \displaystyle\sum_{m=-\infty}^{\infty} h_{r,1}[2^{r+1}n - m]x[m], & 0 \leq r \leq L - 1 \\ \displaystyle\sum_{m=-\infty}^{\infty} h_{r,0}[2^r n - m]x[m], & r = L \end{cases} \tag{2.8}$$

where r is the current level, L is the number of levels, and $h_{r,0}$ and $h_{r,1}$ refer to the low and high-pass filters respectively. More precisely, the wavelet at level r is related to the impulse response $h_{r,1}$ of its corresponding high-pass filter $H_{r,1}$, for the rth bandpass output. The impulse response $h_{1,1}[n]$ of its corresponding high-pass filter $H_{1,1}[n]$, is referred to as the *mother wavelet*. Lastly, the impulse response $h_{r,0}[n]$ for its corresponding low-pass filter $H_{r,0}[n]$, is known as the *scaling function* at level r.

2.3.3.5   The Hilbert Transform

The Hilbert Transform is useful in calculating instantaneous attributes of a time series relating to the amplitude and frequency. The instantaneous amplitude is the amplitude of the complex Hilbert Transform, and the instantaneous frequency is the derivative of the instantaneous phase angle. A function and its Hilbert transform both have the same energy and therefore the energy can be used to measure the accuracy of the approximated Hilbert transform. The Hilbert Transform is denoted by the equation

$$\widehat{x}(t) = \frac{1}{\pi}P\int_{-\infty}^{\infty}\frac{x(\tau)}{t-\tau}d\tau \tag{2.9}$$

where P is denotes the Cauchy principal value [8]. The Hilbert transform may be thought of as the convolution of the original signal $f(x)$ with the impulse response $h(t) = \frac{1}{\pi t}$, as such

$$\widehat{x}(t) = H\{x\}(t) = (h * x)(t) = \frac{1}{\pi}\int_{-\infty}^{\infty}\frac{x(\tau)}{t-\tau}d\tau \tag{2.10}$$

This transform is very common in signal processing because of its ability to give instantaneous frequency information from its relation to the phase angle.

2.3.4 Signal Modeling Techniques

2.3.4.1 Linear Predictive Coding Taps

Linear Predictive Coding Taps (LPC) is a popular tool used frequently in audio and speech analysis because of its ability to encode decent quality speech at a low bit rate [42]. It has the unique capability of being able to represent the spectral envelope of a signal in a compressed form. An efficient algorithm known as the Levinson-Durbin algorithm is used to estimate the linear prediction coefficients from a given waveform. In linear predictive coding, a signal is modeled by the following equation:

$$\tilde{x}(n) = a_1 x(n-1) + a_2 x(n-2) + \ldots + a_M x(n-M) = \sum_{i=1}^{M} a_i x(n-i) \qquad (2.11)$$

where M samples of the signal are used to predict the present sample of the signal while minimizing the Mean Squared Error (MSE) over the entire signal. $\tilde{x}(n)$ is the prediction of $x(n)$, $x(n-1)$ is the i-th step of the previous sample, and $\{a_i\}$ are the linear predictive coefficients. The coefficients $\{a_i\}$ are determined by using an adaptive Finite Impulse Response (FIR) filter along with the Least Means Squares (LMS) algorithm. The error between the actual and predicted sample can be expressed as

$$\varepsilon(n) = x(n) - \tilde{x}(n) = x(n) - \sum_{i=1}^{M} a_i x(n-i) \qquad (2.12)$$

2.3.4.2 Auto-Regressive Modeling

An Auto-Regressive (AR) model is quite simple and useful for describing the stochastic behavior of a given time series. Efficient algorithms for parameter estima-

tion make AR modeling very attractive in real-time signal processing applications. The AR model is represented by the equation

$$H(z) = \frac{\sqrt{e}}{A(z)} = \frac{\sqrt{e}}{1 + a_2 z^{-1} + \ldots + a_{p+1} z^{-p}} \qquad (2.13)$$

where the model order is represented by p, and the vector a contains the normalized estimate of the AR system parameters, $A(z)$, in descending powers of z.

2.3.4.3 Iterative Enveloping & Derivatives

Qualitatively, the envelope of a signal $s[n]$ is the boundary within which the signal is contained, when viewed in the time domain. Envelope detectors take an input signal and produce an output signal that is essentially an "envelope" of the original signal. Envelope detection is most commonly used in circuit design as signal rectifiers to convert an AC signal into a DC signal. This method of enveloping finds all the peaks and valleys of a signal by finding when the first derivative of $s[n]$ changes signs. The time point n that precedes the change of signs is assigned to be a peak or valley. If the derivative is 0 at n, n is assigned to be a peak or valley. A peak is defined when the first derivative goes from positive to negative, and a valley is defined when the first derivative goes from negative to positive. The derivative of a discrete signal is approximated by

$$\frac{ds[n]}{dn} = \frac{s_y[n] - s_y[n-1]}{s_x[n] - s_x[n-1]} \qquad (2.14)$$

where $s_y[n]$ is the y-component and $s_x[n]$ is the x-component of signal $s[n]$.

The peaks and valleys of the signal are stored in vectors which perfectly envelope the signal $y[n]$. Enveloping can be done recursively on the resulting envelope to reveal layered frequencies in a signal. Using more than one recursion can provide

a better envelope depending on the spectral density of the original signal. When this computation is done on a windowed signal, it can reveal information about the specific frequency components within in the window.

The second and third derivatives are computed by recursively applying equation (2.14) to the original signal $y[n]$ for the order of derivatives required.

2.3.5 Chaos & Coherence Techniques

2.3.5.1 Phase Shift by Cross-Correlation

The phase shift of a signal is just a measure of horizontal shifting. If the phase shift is positive, the horizontal shift is to the right. If the phase shift is negative, the horizontal shift is to the left. The normal form of a continuous sinusoidal signal is given in the form

$$y = A\sin(B(x - C)) + D \tag{2.15}$$

The phase shift of the signal is determined by finding the change that is made to the value of x, which is the value C.

In the case of discrete signals, there are different methods of calculating the essential phase difference. The method used here involves estimating the cross-correlation sequence of two independent signals. The true cross-correlation sequence is a statistical measure defined as

$$R_{xy}(m) = E[x_{n+m}y_n^\star] = E[x_n y_{n-m}^\star] \tag{2.16}$$

where x_n and y_n are two independent signals, and $E[\cdot]$ is the expected value operator. A common estimate based on N samples of $x[n]$ and $y[n]$ is the deterministic cross-correlation sequence for discrete signals shown in the following set of

equations.

$$\hat{R}_{xy}[m] = \begin{cases} \dfrac{1}{N} \displaystyle\sum_{n=0}^{N-m-1} x[n+m]y^{\star}[n] & 0 \leq m \leq N-1 \\[3em] \dfrac{1}{N} \displaystyle\sum_{n=0}^{N+m-1} x[n]y^{\star}[n-m] & -(N-1) \leq m \leq -1 \end{cases} \qquad (2.17)$$

The expression $\hat{R}_{xy}$ essentially reports the lag between the two signals $x[n]$ and $y[n]$, which, since the discrete time signals in question are all in the time domain, give an estimate of pure time delay, or phase difference, between the two signals [13].

2.3.5.2 Entropy Estimation

In information theory, the entropy, H, of a discrete random variable X is a measure of the amount of uncertainty associated with the value of X. The theory of entropy was developed by Claude Shannon in 1948 to address the classical engineering problem of the transmission of information over a noisy channel [62]. Shannon's *source coding theorem* establishes that, on average, the number of bits needed to represent the result of an uncertain event is given by its *entropy*. The shannon entropy quantifies, in the sense of an expected value, the average information contained in a message, usually in units such as bits.

The information entropy of a discrete random variable X, which in the case of EEG signals is a time series, is defined by

$$H(X) = E[I(X)] \qquad (2.18)$$

$E[\cdot]$ is the expected value operator, and $I(X)$ is the information content of X.

$I(X)$ itself is a random variable due to association. If p denotes the probability mass function of X, then the information entropy can be written as

$$H(X) = -K \sum_{i=1}^{n} p(x_i) \log_b p(x_i) \tag{2.19}$$

where x_i refers to the indices of the time series, b is the base of the logarithm used, and n is the length of the time series. The K variable merely accounts for the choice of a unit of measure.

2.3.5.3   Magnitude Squared Coherence

The Magnitude Squared Coherence estimate is a function of frequency with values between 0 and 1 that indicate how well x corresponds to y at each frequency. The coherence is a function of the power spectral density of x and y (P_{xx} and P_{yy}) and the cross power spectral density (P_{xy}) of x and y [31]. The coherence function is denoted as

$$C_{xy}(f) = \frac{|P_{xy}(f)|^2}{P_{xx}(f)P_{yy}(f)} \tag{2.20}$$

where x and y must be of the same length. The power spectra are calculated using Welch's averaged periodogram method [74], which involves sectioning the time series, taking the modified periodogram of these series, and then averaging these modified periodograms.

This estimate is only useful for datasets with more than one channel as the coherence of any signal with respect to itself if always 1. In the case of data sets containing 6 and 32 channels, the magnitude squared coherence was calculated for each channel against every other channel independently.

2.3.5.4   Lyapunov Exponents

In mathematics, Chaos theory is used to describe the behavior of certain dynamical systems, specifically, systems whose states change over time that may exhibit dynamics that are highly sensitive to initial conditions [11, 67]. The Lyapunov exponents of a dynamical system characterizes the rate of separation of infinitesimally close trajectories denoted as Z. In a discrete time signal, a trajectory is an ordered set of points in a signal. Quantitatively, two trajectories in a signal with initial separation δZ_0 diverge as

$$|\delta Z(t)| \approx e^{\lambda t}|\delta Z_0| \tag{2.21}$$

where λ is the Lyapunov exponent.

The rate of separation can be different for different orientations of the initial separation vector thereby creating a whole spectrum of Lyapunov exponents $\{\lambda_1, \lambda_2, \ldots, \lambda_n\}$. It is common to just refer to the Maximal Lyapunov Exponent (MLE) because it determines the predictability of a dynamical system. A positive MLE is an indicator that the system is chaotic. The maximal Lyapunov exponent is defined as

$$\lambda = \lim_{t \to \infty} \frac{1}{t} \ln \frac{|\delta Z(t)|}{|\delta Z_0|} \tag{2.22}$$

2.4   Classification Techniques

Classifiers are used to map feature space X to a discrete set of labels C. Classifiers are further subdivided into supervised and unsupervised classifiers, the former having a training set consisting of feature vectors and their outputs, the latter simply

having a set of feature vectors where the outputs are determined by the algorithm.

2.4.1 Principal Component Analysis

The central idea of Principal Component Analysis (PCA) is to reduce the dimensionality of a dataset consisting of a large number of interrelated variables while retaining as much variation as possible. This is achieved by transforming to a new set of variables, the principal components, which are uncorrelated, and which are ordered so that the first few retain most of the variation present in all of the original variables [29, 64].

Let $\{x_1, x_2, x_3, \ldots, x_M\}$ be a set of M column vectors (dimensions) of N trials. The mean of each dimension is represented as

$$\bar{x} = \frac{1}{N} \sum_{i=1}^{M} x_i \qquad (2.23)$$

The first step is to subtract off the mean which is done as follows

$$\Phi_i = x_i - \bar{x} \qquad (2.24)$$

Next, form the matrix $A = [\Phi_1 \Phi_2 \ldots \Phi_M]$ and then compute the covariance matrix to characterize the scatter of the data

$$C = \frac{1}{M} \sum_{n=1}^{M} \Phi_n \Phi_n^T = AA^T \qquad (2.25)$$

Once this is done, the eigenvalues and eigenvectors must be calculated from the covariance matrix as follows

$$(C - \lambda I)u = 0 \qquad (2.26)$$

which results in the eigenvalues of $C : \lambda_1 > \lambda_2 > \ldots > \lambda_N$ and the eigenvectors of $C : u_1, u_2, \ldots, u_N$.

Since C is symmetric, $\{u_1, u_2, \ldots, u_N\}$ form a basis, (i.e., any vector x, or actually $(x - \bar{x})$, can be written as a linear combination of the eigenvectors):

$$x - \bar{x} = b_1 u_1 + b_2 u_2 + \ldots + b_N u_N = \sum_{i=1}^{N} b_i u_i \tag{2.27}$$

The last step is to reduce the dimensionality, which is done by keeping only the terms that correspond to the K largest eigenvalues (K being the desired dimensionality), as such

$$\hat{x} - x = \sum_{i=1}^{K} b_i u_i \quad where \quad K << N \tag{2.28}$$

2.4.2 Linear Discriminant Analysis

Linear Discriminant Analysis (LDA), originally developed by R.A. Fisher in 1936, is a classical method of classification. It is a statistical technique to classify samples into mutually exclusive and exhaustive classes based on a set of measurable features [47]. LDA transforms the data into a new feature space where the classes can be separated more easily. It then maximizes the between-class to within-class ratio to form distinctions between the different classes.

Let $\vec{x}_i$ be a set of N column vectors containing i samples of dimension D. The mean of the dataset is

$$\vec{\mu}_x = \frac{1}{N} \sum_{i=1}^{N} \vec{x}_i \tag{2.29}$$

where each of the samples belongs to one of V classes $\{C_1, C_2, C_3, \ldots, C_V\}$. The mean of class V containing N_V members is denoted as $\vec{\mu}_{xv}$. The between-class

scatter matrix S_B is defined as

$$S_B = \sum_{v=1}^{V} N_v (\vec{\mu}_{xv} - \vec{\mu}_x)(\vec{\mu}_{xv} - \vec{\mu}_x)^T \tag{2.30}$$

The within-class scatter matrix S_W is defined as

$$S_W = \sum_{v=1}^{V} \sum_{\vec{x}_i \in C_v} (\vec{x}_i - \vec{\mu}_{xv})(\vec{x}_i - \vec{\mu}_{xv})^T \tag{2.31}$$

The transformation matrix that repositions the data to be most separable is the matrix W that maximizes

$$\frac{\det(W^T S_B W)}{\det(W^T S_W W)} \tag{2.32}$$

In this sense, W is the set of generalized eigenvectors of S_B and S_W where $W = [\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_D]$. This gives a projection space of dimension D. The goal is to maximize the inter-class to intra-class ratio as in the equation

$$W = \max[S_w^{-1} S_b] \tag{2.33}$$

The transformation function found by maximizing the ratio W rotates the axes of the original data such that the classes are now projected onto a new basis where the difference between the classes are maximized. When testing data is given to the algorithm, the test vectors are transformed and the Euclidean distance (shown in equation (2.34)) of the test vectors from the means of the class are calculated. A test vector is a member of a class when that vector has the smallest euclidean distance to the class mean.

2.4.3 K-Nearest Neighbor

The K-Nearest Neighbor (KNN) algorithm is based on finding the number of K-nearest neighboring samples around an unknown sample. The class of the sample is determined by the class that has the highest percentage of neighbors.

In the training phase, the training data, which is represented as a set of features is stored in a multidimensional feature space. This multidimensional space is divided into regions by location and the classes given to each sample. Let $\vec{x}_i$ be a set of N column vectors containing i samples of dimension d where each of the samples belong to one of V classes $\{C_1, C_2, C_3, \ldots, C_V\}$. During the testing phase, a test vector is given and distances from the new sample $\vec{y}_i$ to all the samples in the training set $\vec{x}_i$ are calculated. Most commonly, the Euclidean distance is used, though other types of distance measures may be used as well. The Euclidean distance between two points for a d dimensional space is given by

$$D(\vec{x}_i, \vec{y}_i) = \sqrt{\sum_{i=1}^{d} (\vec{x}_i - \vec{y}_i)^2} \qquad (2.34)$$

Based on the calculated distances, the number of k nearest samples to the new sample is selected. The class of the new sample is predicted based on finding the class with the maximum number of K-nearest samples to the set of testing data.

2.4.4 Support Vector Machines

Support Vector Machines (SVMs), which are a relatively new method for learning separating functions in pattern recognition tasks, originated from the Statistical Learning Theory (SLT) developed by Vapnik and Chervonenkis [72]. Unlike other classification techniques which attempt to minimize the error of classification, SVMs incorporate structured risk minimization which minimizes the upper bound on the

generalized error [54]. The idea is to determine from an infinite number of planes dividing two classes, the one which will have the smallest generalization error. This separating hyper-plane created in an n-dimensional space maximizes the margin between the two classes, thereby maximizing the classification potential [70].

SVMs are based on the concept of decision hyper-planes that define decision boundaries. A decision hyper-plane is one that separates a set of objects having different class memberships. In cases when the given classes cannot be separated in the original input space, the SVM nonlinearly transforms the original input space into a higher dimensional feature space. This transformation can be achieved by using various nonlinear mappings: polynomials, sigmoidal (as in multilayer perceptrons), spline functions, etc. After this nonlinear transformation, finding the linear optimal separating hyperplane is trivial. The optimization problem to solve will be of the same kind as the calculation of a separating hyper-plane in the original input-space for linearly separable classes [32].

2.5 Prediction of Ictal States using EEG analysis

2.5.1 The Preictal State

The shift from the interictal condition (relatively asymptomatic), to the ictal condition (seizure state, loss of consciousness) is considered a *state transition*. The question is: Are these changes during the transition gradual or abrupt? A gradual transition would provide the possibility of predicting the ictal state prior to its occurrence. An abrupt change, however, would not provide enough mathematically deducible data.

Evidence of a gradual transition has been reported as early as the 1970s, where Viglione and Walsh [73] reported changes in EEG characteristics using linear meth-

ods several minutes prior to seizure onset. Seigel et al. [66] were able to find a difference in the distribution of spikes during the interictal period just minutes before onset. Junling and Dazong [30] proposed a linear seizure predictor which tracks the time evolution of the slow wave energy larger than some preset threshold in scalp EEGs. They were also able to detect a pre-seizure phase for several minutes prior to the ictal state. Niederhoefer and Tetzlaff [49] used prediction error profiles to forecast seizures. They were able to show distinct changes minutes before seizure onset. Using nonlinear techniques, Andrzejak et al. [1] were able to show that EEG analysis provided strong evidence that it is indeed possible to detect the presence of a pre-seizure state, but the state could occur from hours to just minutes before the ictal state. Litt et al. [38], who also used EEG analysis for seizure onset prediction, states that there is considerable evidence that seizures develop over up to hours prior to the occurrence of the seizure. Niederhofer et al. [48] studied a generalized EEG-signal analysis based on spatiotemporal linear prediction methods and found a distinct increase in the values of the predictor coefficients just prior to ictal onset. The presence of the preictal phase exhibits the ability to forecast and gives hope to research on seizure prediction, but there is a lack of consensus to this question.

Maiwald et al. [41] proposed that both seizure detection and prediction methods should be evaluated with respect to the true positive rate, TPR_{max}, and the false prediction rate, FPR_{max}. The TPR_{max} refers to the number of true predictions in a time interval, while FPR_{max} refers to the number of false predictions in a time interval. Increasing the TPR_{max} to 100% would also dramatically increase FPR_{max} and result in a large number of false alarms during the interictal period.

2.5.2 Methods & Techniques

This section of the literature review discusses the different mathematical and signal processing techniques that have been used by several researchers in the field of seizure prediction. The techniques used are comprised of both linear and non-linear methods.

Wavelets are a popular tool to use in EEG signal analysis because of their ability to provide both time and spectral information. Aziz et al. [3] built a Very-Large-Scale Integration (VLSI) seizure prediction microsystem that was tested on the hippocampus of a mouse which used a Wavelet Artificial Neural Network (WANN) prediction algorithm. This real-time spectral analysis device was able to predict induced seizures, but the false-prediction rate was higher than desired. Schuyler et al. [61] also developed a seizure prediction algorithm based on wavelet decomposition of the EEG signal and found that the signal's high frequency components contain information relevant for seizure prediction.

Chaos theory and coherence measures are widely used tools, especially for use in seizure prediction algorithms. McSharry et al. [43] discuss the use of variance and correlation dimensions on EEG recordings in seizure prediction algorithms. They show that there is evidence of a state that exists before the ictal state. However, Lai et al. and others [26, 35, 53] state that Lyapunov exponents computed from a time series were shown to be ineffective due to statistical fluctuations and noise. As such, its ineffectiveness would be especially exaggerated in a high-dimensional noisy system such as the brain. Golovko et al. [17] used an Independent Component Analysis (ICA) approach for signal extraction, calculated the Lyapunov exponents, and then used an Artificial Neural Network (ANN) for seizure forecasting. ICA is an computational technique for separating a multivariate signal into additive subcomponents by maximizing the statistical independence of the estimated com-

ponents.

Entropy is also useful in measuring the degree of order in a time series. Molteni [45] states that a change in the entropic content within the EEG may mirror a real modification in cortical functional organization. Mishra et al. [44] studied measures of entropy within the preictal period of seizures that were induced in rats and found that the time series became more ordered (measured with Shannon Entropy) during the ictal period.

Several people have provided evidence that there is a difference in the synchronization between different parts of the brain during the ictal and interictal periods. Hively and Protopopescu [21] use Phase-Space Dissimilarity Measures (PSDM) to extract the correlation between channels connected to different sections of the brain and found that there were differences between the interictal and ictal states. Gupta et al. [19] use Phase Locking Values (PLV), which is an index based on the average variation of the phase synchronization, to assess the amount of synchronization of a windowed time signal. They were able to show a difference in phase synchronization during periods of ictal activity. Krug et al. [34] used a Cellular Nonlinear Networks (CNN)-based approach to measure the strength of generalized synchronization in EEG recordings from epilepsy patients. They were able to show the usefulness of CNN-based synchronization approximation for the development of miniaturized seizure prediction devices. Nagaraj et al. [46] developed a device that measures synchronization using cross-correlation between different recording sites in the brain. They were also able to show a change in synchronization before the ictal period.

Other numerical methods that have been used are filter coefficients and methods of Auto-Regression. Haas et al. [20] developed an algorithm for real-time detection of epileptic seizures based on FIR filter coefficients and LPC that achieved zero false

negatives and zero false positives, however, the average seizure onset prediction time was 15.5 seconds. Isaksson et al. [27] showed the usefulness of auto-regressive models for spectral parameter analysis of the EEG.

2.5.3 Feature Space Optimization

An important aspect of any seizure prediction algorithm is the feature set. However, this is an aspect that seems to have been overlooked by most researchers. A comparison of the strength of each of the numerical methods would be useful in finding the best methods of seizure prediction. D'Alessandro et al. [10] developed an algorithm that uses a genetic algorithm to select the optimal feature set. They concluded that a patient specific feature set was required for seizure prediction.

2.5.4 State-of-the-Art

The following state-of-the-art seizure prediction algorithms have shown the most potential in this field up until now. Gopalsami et al. [18] proposed an implantable Surface Acoustic Wave (SAW) microsensor for early detection of the seizure state based on the local temperature changes in the brain's epileptogenic zones. Iasemidis et al. [25] developed and patented an automated seizure prediction system based on the convergence and divergence of short-term Lyapunov exponents. The system was able to predict seizures 82% of the time with a false prediction rate of 0.16/hour. Seizure onset was predicted, on average, 71.7 minutes before ictal onset, however, the system produced an average of one false warning every 6.25-8.33 hours. Osorio et al. [50] developed a real-time automated seizure detection system which could predict seizure onset with a mean of 15.5 seconds. This system uses several digital signal processing techniques used in time-frequency localization and image processing.

As seen from the literature review above, many different theories have been applied in an effort to detect seizures before they occur. Aspects of chaos theory and synchronization of EEG channels were very popular methods, and many used approaches that involved different types of frequency analysis, wavelets being most common. Different feature optimization techniques were explored to find out what mathematical techniques were useful in the determination of the preictal state. All of these investigators, however, focused their research on one or two major mathematical techniques. Perhaps a combination of techniques would improve the performance of these algorithms.

2.6 Summary

In this chapter, background information about epilepsy, EEGs, and brain waves are given. Several techniques to decompose and model a time series signal are discussed and four classification techniques are provided to be used in the seizure prediction algorithm discussed in this thesis. Lastly, a literature review is given discussing the determination of the preictal state, feature space optimization, and common techniques used for predicting seizure onset.

Chapter 3

Methods

This chapter provides a description of the automatic seizure prediction system proposed in this thesis. First, a discussion about the characteristics of the data is provided followed by the methods that were developed to solve this problem. A description is given of the system that was implemented and tested with the data that was provided. A closed-loop prediction algorithm is presented that makes decisions about the current epileptic state based on past predictions, and a feature optimization algorithm is presented to determine the feature sets that most improve the accuracy of predicting the different epileptic states.

3.1 EEG Data

In total, 24 EEG datasets containing both ictal and interictal data were provided for analysis. These 24 sets can be further subdivided into 6 channel, and 32 channel sets. The scheme of the locations of surface electrodes used is based on the standard international 10-20 system [2].

Four EEG datasets were provided by the Krembil Neuroscience Center in Toronto, Canada. Each of these sets are approximately 1 hour long and contain 32 chan-

nels of EEG data in the standard international format. The first dataset provided was an example of one severe occurrence of a seizure (possibly a tonic-clonic [65]), and the second dataset was an example of a complex partial seizure [65] in one hemisphere of the brain followed by a generalized seizure [65] several minutes later. Both of these data sets were sampled at 500Hz. The third and fourth data sets contained several minutes of interictal EEG data as the "baseline", and were both followed by episodes of ictal activity. These two data sets were sampled at 250Hz.

Patient	Sex	Age	Foci	Type	Cases
1	F	15	Frontal	SP,CP	4
2	M	38	Temporal	SP,CP,GTC	3
3	M	14	Frontal	SP,CP	5
4	F	26	Temporal	SP,CP,GTC	5
5	F	16	Frontal	SP,CP,GTC	5
6	F	31	Temp./Occ.	CP,GTC	3
7	F	42	Temporal	SP,CP,GTC	3
8	F	32	Frontal	SP,CP	2
9	M	44	Temp./Occ.	CP,GTC	5
10	M	47	Temporal	SP,CP,GTC	5
11	F	10	Parietal	SP,CP,GTC	5
12	F	42	Temporal	SP,CP,GTC	4
14	F	41	Front./Temp.	CP,GTC	2
15	M	31	Temporal	SP,CP,GTC	4
16	F	50	Temporal	SP,CP,GTC	5
17	M	28	Temporal	SP,CP,GTC	5
18	F	25	Frontal	SP,CP	5
19	F	28	Frontal	SP,CP,GTC	4
20	M	33	Temp./Par.	SP,CP,GTC	5
21	M	13	Temporal	SP,CP	5

Table 3.1: Epileptic Patient Statistics: Freiburg Data

Seizure Types: SP, Simple Partial; CP, Complex Partial; GTC, Generalized Tonic-Clonic seizure origin.

The Freiburg Seizure Prediction Project in Freiburg, Germany provided invasive EEG recordings of 21 patients suffering from intractable focal epilepsy. The data was recorded at the Epilepsy Center of the University Hospital of Freiburg,

Germany. In eleven patients, the epileptic focus was located in neocortical brain structures, in eight patients, the epileptic focus was in the hippocampus, and in the last two in both. The EEG data was originally recorded with 128 channels using a 256 Hz sampling rate, and with a 16-bit A/D converter. For each of the patients, 2 sets of data were separated and defined as 'ictal' and 'interictal'. The 'ictal' data sets contained at least 50 minutes of preictal data prior to the seizure as well as the seizure itself, and the 'interictal' sets contained approximately 24 hours of continuous interictal EEG recordings. Due to technical reasons, the data was separated into blocks with a temporary pause of 1-3 seconds between each. For these data sets, six contacts of all implanted grid, strip, and depth electrodes were selected by visual inspection of the raw data by a certified epileptologist. Three of the contacts were chosen from the epileptogenic focus, and the remaining three were selected as not involved or involved latest during the seizure. Patient 13 was dropped from testing due to insufficient data prior to the ictal state.

3.2   System Description

In this thesis, a system is proposed that can use multiple channel real-time EEG signals to predict the onset of an epileptic seizure. The system is given a select number of EEG channels as input and reports back the corresponding epileptic seizure state at every second. In this system, the classifications are done as a simulation of real-time dynamic predictions and is dependent upon past predictions that were made. As a result, the sensitivity must be controlled such that seizures aren't predicted more often than they actually occur. This closed-loop system controls sensitivity by several methods which are explained in this chapter. The system can be functionally reduced to the block diagram seen in Figure 3.1.
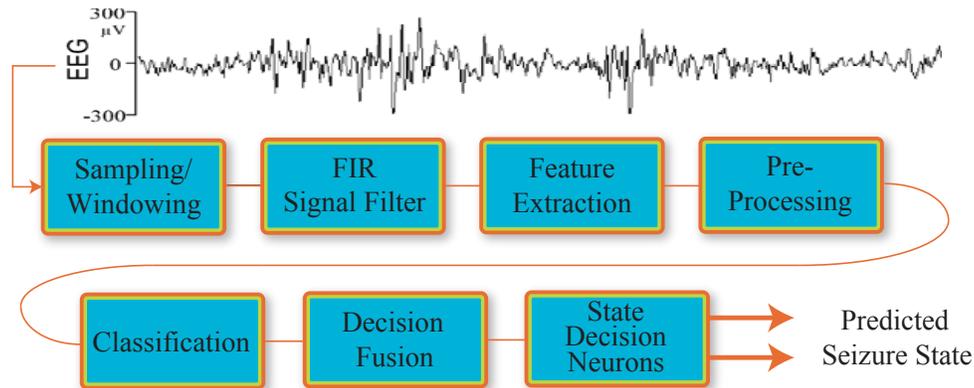
Figure 3.1: Block Diagram of the EEG Classification System.

The overall system is given a multi-channel EEG stream in segments of 3 seconds every second, and a set of features are extracted at each time point and denoted as a sample. These samples are taken every second such that the subsequent window taken overlaps. As a result, the samples collected show a more gradual transition from one epileptic seizure state to the next. A rectangular window is applied to each 3-second segment such that there is minimal distortion in the frequency response (some distortion will be present due to Gibb's Phenomenon). An FIR signal filter is applied to decompose the incoming EEG stream into its respective brain waves. Features are extracted from the incoming data streams starting from the beginning to the end of the EEG such that it simulates a real-time scenario. If a sample is extracted that contains mathematical anomalies resulting in values of NaN, the sample is simply discarded and skipped over. The training data used is 80% of a new random permutation of the entire training set for every classification performed, and the testing data is the sample that was extracted from the current window. Once the classifiers have each made a prediction, a decision fusion algorithm (discussed in Section 3.3.3) uses a set of rules to come up with an initial prediction. This prediction is given to the state decision neurons, discussed in Section 3.3.2.2, which

use a closed-loop algorithm to determine if a state change is necessary. The states defined in the system correspond directly to the epileptic states described in Section 2.2.2, and is listed in Table 3.2.

State #	Epileptic State
2	Interictal
3	Preictal
4	Ictal
1	Postictal

Table 3.2: Defined Epileptic States in transient EEG signals

Since the postictal and interictal states have signal characteristics that are similar (both represent non-ictal states), it was necessary to place the states next to each other (*i.e.* States 1 and 2). This way, if State 2 is misclassified as State 1, or vice versa, then the average of several classifications will also be in the range of States 1 and 2. If these states were defined as States 1 and 4, the average of several classifications would result in increased misclassifications of these states as States 2 or 3, which is incorrect.

3.2.1 Feature Extraction

One of the most important steps in creating a signal classification system is generating mathematical representations and reductions of the input data which allow the input signal to be properly differentiated into its respective classes. These mathematical representations of the signal are, in a sense, a mapping of a multidimensional space (the input signal) into a space of fewer dimensions. This dimensional reduction is known as "feature extraction". Ultimately, the extracted feature set should preserve only the most important information from the original signal. The mathematical transforms that were selected for this system are listed in Table 3.3.

Set	Mathematical Transform	Feature Number	# of Features
1	Linear Predictive Coding Taps	1-4	[4]
2	Fast Fourier Transform Statistics	5-10	[6]
3	Fast Fourier Transform Subbands	17-75	[59]
4	Power Spectra Analysis	11-16	[6]
5	log(FFT) Analysis	76-81	[6]
6	Mel Frequency Cepstral Coefficients	82-85	[4]
7	Wavelet Decomposition	86-89	[4]
8	Enveloping & Peak Analysis	90-99	[10]
9	1st, 2nd, & 3rd Derivatives	100-102	[3]
10	Phase Shift Correlation	103-108	[6]
11	Shannon/Log Entropy	109-110	[2]
12	Coherence Estimate	111-128	[18]
13	Hilbert Transform Statistics	129-138	[10]
14	Auto-Regressive Parameters	139-145	[7]
15	Lyapunov Exponents	146-151	[6]
Total	-	-	151 Features

Table 3.3: Mathematical transforms available to the system to analyze incoming EEG streams.

3.3 Experimental Methods

The methods described in this section were developed to study and enhance the classification of EEG signals for the problem of seizure onset prediction. First, a feature set optimization algorithm is presented which is used to do a feature set study to reveal the mathematical transforms that are most useful in predicting the preictal state. After this, a set of algorithms are given that became the framework of the seizure onset prediction system described in this thesis.

3.3.1 Stepwise Feature Optimization

In order to find the features with the most potential, an algorithm was implemented to approximate individual feature strength with respect to every other feature. The strength of a feature was determined by the accuracy with which the preictal state was classified as an average of several classifications. Similar to Cross-Validation by

Elimination (explained in Section 3.3.2.1), this algorithm repartitions the feature set, performs a set of classifications, finds the best feature sets to drop, and then adjusts the feature space to only contain features that improve the accuracy.

Let $\{f_1, f_2, \ldots, f_K\}$ be the feature sets where K denotes the set number located in the first column of Table 3.3.

Algorithm 1 Stepwise Feature Optimization

1: Evaluate the accuracy of the classification using all K feature sets.

2: Dropping one feature set at a time, repartition the feature space into K, K-1 feature subsets and save the accuracy of each subset at position K in vector φ along with the resulting accuracy.

3: Denote the index of φ with the maximum accuracy as ρ, and drop all the features listed in φ from ρ to K from the final feature space.

The resulting feature set φ has an accuracy similar to the accuracy found at position ρ in φ. Undertraining and overtraining must still be taken into consideration since it can have an effect on the accuracy of a prediction.

3.3.2 Post-Classification Validation

The two methods in this section were developed to complement the classification algorithms and enhance their classification potential for noisy dynamical systems that change state over time. The first method, which is called *Cross-Validation by Elimination*, is used to classify samples by testing the amount of correlation (determined by the accuracy of classifications) each sample has to every state and then remove classes that are least correlated to improve classification accuracy. The algorithm isolates each of the classes, compares the prediction results, and then makes a final decision based on a function of the independent predictions. This method puts testing samples that were weakly classified into classes that make

more sense based on a comparison of the prediction accuracy of each class.

The second method, *State Decision Neurons*, is used to automatically make decisions about when to transition to the next defined state. This algorithm, when used in conjunction with a set of classifiers, enables the system to make decisions based on previous predictions, a closed-loop system if you will. When there are three or more states to distinguish between in a noisy system, state decision neurons are useful in determining the appropriate moments to transition to another state.

3.3.2.1 Cross-Validation by Elimination

Cross-Validation by Elimination (CVE) is an algorithm used to correct misclassifications by repartitioning the training set in order to remove correlation between $(N > 2)$ different classes. Since it requires an additional $(N! - 2)$ classifications, which can be very computationally expensive depending on the classification method, it is best used as a method of verification when more basic classifiers' predictions begin to deviate from the current state. The decision to execute this algorithm is made through the a decision fusion algorithm discussed in Section 3.3.3. When a classification is performed using a particular algorithm and the resulting classification is different than the current state of the signal, you've either reached an exceptionally noisy part of the signal or the classifier is predicting a state transition. This method provides a better estimate of the prediction by removing classes (states) that are potentially weaker one by one, and then determining the class with the highest classification accuracy. This algorithm is only useful if there are three or more classes to differentiate between.

Let $\{C_1, C_2, \ldots, C_k\}$ be the set of classes in the training data; for this example we will set k = 4.

Algorithm 2 Cross-Validation by Elimination

1: Repartition the class memberships in the training data such that the first class, or class C_1, is defined as 1 and the rest of the classes, $\{C_2, C_3, C_4\}$, are all redefined to 2.

2: Execute the desired classification algorithm with the resulting repartitioned training data and save the classification accuracy in vector R, whose indices correspond to the class defined as 1.

3: Repeat steps 1-2 for each of the other classes until vector R is populated.

4: Drop the class corresponding to the lowest accuracy in vector R.

5: Repeat steps 1-4 until there are two classes remaining in vector R.

6: Perform one last classification to determine the performance of each of the two remaining classes.

If there is only one entry in the testing vector given to the classifier, then the vector R will only contain 0 (0%) or 1 (100%) (not classified, or classified). One can also implement a function to order the dropping of classes (in step 4) based on the current state of the signal rather than merely dropping the class with the lowest accuracy. For example, if the current state is 2, and States 3 and 4 were classified at 0%, and State 4 was more improbable than State 3, a function could be implemented in place of step 4 to choose to drop State 4 rather than State 3.

3.3.2.2 State Decision Neurons

State Decision Neurons (SDN) are useful in classification problems involving discrete time signals that change between states over time. It essentially represents each state with a population of neurons, but only the collective 'active' population can make the ultimate decision as to if a state transition should occur. Each population of neurons represents a unique state of the signal that has defined weights to pass activity to the appropriate population of neurons (a state transition). The amount of neurons in each population is a parameter that needs to be optimized

since it controls the number of past predictions to consider. Figure 3.2 shows an example of an implementation with 32 neurons (in each population) with four unique states to transition between. The numbers beside each of the arrows show the number of neurons in the population that have to predict the state that the arrow is pointing to in order to make the associated state transition.
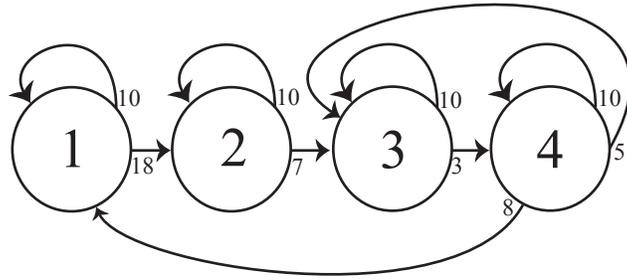
Figure 3.2: Example of State Decision Neurons.

As predictions are made about the state of the signal (based on the methods described in Section 3.3.3), they are immediately passed to the active state neuron. As predictions are made, the neurons fill up and the oldest predictions are dropped to make room for new ones. Based on the predictions stored by the neurons of the active population, it can make the decision to pass activity to another population of neurons when it defines the signal as another state for a defined proportion of the neurons (recent predictions). These percentages are what make up the weight matrix W_{ij}. The weight matrix is an NxN matrix where N is the number of states possible.

$$W_{ij} = \begin{bmatrix} w_{11} & w_{12} & w_{13} & w_{14} \\ w_{21} & w_{22} & w_{23} & w_{24} \\ w_{31} & w_{32} & w_{33} & w_{34} \\ w_{41} & w_{42} & w_{43} & w_{44} \end{bmatrix} \qquad (3.1)$$

The weight matrix stores the weights required to change from state i to state j. For example, the weight to move from State 1 to State 2 is located at index w_{12}. The weights represent a threshold for a transition from state i to state j to occur. In this way, a weight that is 0 defines that it is impossible to move from state i to state j. A weight that is approximately 1 defines that state neuron i is very sensitive to a state transition going to state j. For example, if w_{12} is .2, state neuron 1 will require that at least 80% of its neurons are storing predictions of 'State 2' before actually making the transition to State 2. The weight matrix for the state decision neurons shown in Figure 3.2 is given by

$$
W_{ij} = \begin{bmatrix} 0.6875 & 0.4375 & .68 & 0 \\ 0 & 0.6875 & 0.7813 & 0 \\ 0 & 0 & 0.6875 & 0.9063 \\ .75 & 0 & 0.8438 & 0.6875 \end{bmatrix}
\tag{3.2}
$$

where the non-zero values in the matrix represent all the possible state transitions.

One might notice that the w_{ij} values above the diagonal represent forward state transitions, while values below the diagonal represent backward state transitions. In this example, setting w_{41} to a non-zero value allows for the system to reset back to State 1 and continue looking for state transitions. Also, any matrix values adjacent to any side of the diagonal transition states in a sequential manner.

Generating the Weight Matrix

The weight matrix can be calculated manually or automatically. Automatically calculating the weights requires a two-step optimization problem defined by equations (3.3) and (3.4).

$$b = \max y_i[n] \tag{3.3}$$

$$w_{ij} = \frac{([\min y_j[n] > b] + \epsilon)}{c} \tag{3.4}$$

where $y_i[n]$ and $y_j[n]$ are defined as $y_s[n]$ (known as state curves) for state s, c is the number of neurons, and n represents the time points of the EEG signal where samples have been extracted. Equation 3.5 is used to determine the state curves $y_s[n]$. The function $\delta_s[n]$ is 1 when state s is selected at time point n, and 0 elsewhere. ϵ is a user defined threshold to adjust the sensitivity of any state transition. It is useful to set parameter ϵ to some value in order to control the true and false prediction rates.

$$y_s[n] = \begin{cases} \sum_{i=1}^{n} \delta_s[n] & 0 \le n \le c \\ \sum_{i=n-c}^{n} \delta_s[n] & n > c \end{cases} \tag{3.5}$$

As an example of the application of this theory, Figure 3.3 shows a set of possible state decision neuron tuning curves. These curves are useful in manually or automatically determining the weights required for the system. Once these weights have been determined, if the input system (in this case, EEG signals) continues to behave in a similar manner, the state decision neurons will be able to make the state transitions at more appropriate times.

As seen in Figure 3.3, State Neuron 1 is the active population up until t_2 where the threshold for a transition to State 2 was passed (weight w_{12}), whereby state neuron 1 passed activity to state neuron 2. Equations (3.3) and (3.4) attempt to find a weight that is higher than the maximum number of neurons predicting State

1 before the transition to State 2 for the duration of State $1([0, t_2])$. The example shown is idealistic, as most of the time there are other states which classify as interference, but depending on the strength of the classification methods, the correct state should have the maximal classification, and the next state should be selected sporadically at times close to the state transition in anticipation of the next state change.
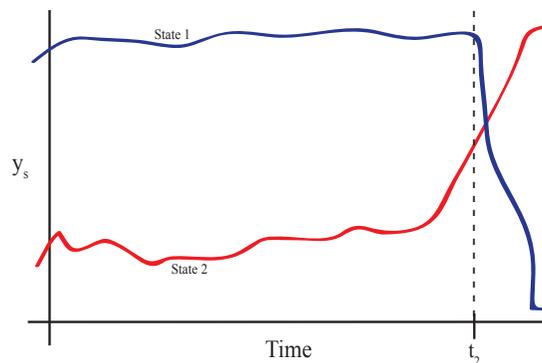
Figure 3.3: $y_s[n]$ curves for automatic determination of weight matrix.

3.3.3 Fusion of Classification Methods

The epileptic seizure onset prediction system proposed in this thesis uses a series of different classifiers at every time point to predict the state of the epileptic condition. The use of multiple classifiers requires a method to combine the predictions in order to make a final classification. Each sample taken is given to four classifiers: LDA, KNN, SVM, and CVE. The method used here involves an initial voting system followed by a decision tree to come up with a prediction of the current state.

Since there are four classifiers, a majority vote may not require all of the classifiers for a decision to be made. For the system proposed in this thesis, the classification module begins by first using the LDA and KNN algorithms, since they are
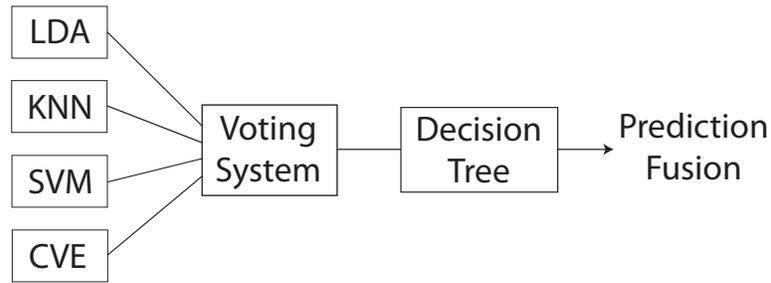
Figure 3.4: Decision Fusion Block Diagram.

the least computationally expensive. If LDA and KNN both present predictions that are different from the current state or the next most probable state, then CVE is used to give an additional prediction in hopes of obtaining a majority vote. If there is still no majority, SVM is used to determine if the current state or the state that has the majority vote (from the other classifiers used) is most probable.

At this point, a voting matrix tabulates the votes of each of the classifiers used for each of the unique states and an initial prediction is made based on the state with the majority vote. A decision tree is then used to alter the initial prediction such that it represents the most probable state based on both the current state and the initial prediction. For this seizure prediction emulator, it was necessary to define that the progression of the states must happen in a sequential and increasing manner. Given this condition, the only other requirements are to define a way to reset the system if: 1) the seizure doesn't occur or 2) the seizure occurs and the system transitions into the postictal state. Given that states can only transition in a sequentially increasing manner, if State 4 has two or more votes while the signal is in State 2, the initial prediction passed on from this step is State 3. This alters state neuron 2 so that a transition to State 3 is more probable, making a transition to State 4 possible sooner. In this way, a set of rules maintained that the progression of state transitions occurred in a sequentially increasing manner.

3.4   Experiments

A set of experiments were performed to determine characteristics about the response of ictal EEG data to several mathematical techniques. Finally, an emulator of a system that could take in a multi-channel EEG stream and return seizure status indicators was created. The first experiment was to create and test a framework that could be used to transition between a number of states based on the predictions of multiple classification algorithms. This initial testing was done using the EEG data provided by the Krembil Neuroscience Center in Toronto, Canada. All subsequent experiments use the data provided by the Frieburg Seizure Prediction Project in Germany. This was done because it was easier to assess the accuracy of the system using several cases of seizure from the same patient.

The second experiment focused on determining a feasible duration of the preictal state that best expose EEG characteristics that lead up to the ictal state. The third experiment was to find an optimal Seizure Prediction Horizon (SPH) for each of the patients given. Subsequent to an alarm, a time interval is necessary during which no seizure is supposed to start, to be regarded as a correct prediction [60]. In this experiment, training and testing data are unrelated and from different seizure cases, but all recorded from the same patient. Both the second and third experiments also find the optimal feature sets that maximize the accuracy of the prediction. These experiments show the relative strength of each of the classification algorithms and also approximate the behavior of the state transitions.

3.4.1   Framework for EEG Signal Classification

In this experiment, a framework for predicting seizure onset was created and tested. As a test of the methods presented, a first trial was done using the tonic-clonic

seizure case given by the Krembil Neuroscience Center in Toronto. Tonic-Clonic seizure activity is generally easy to see visually in an EEG as the amplitudes of the signal start to peak and a different set of frequencies are prevalent. This served as an easy indicator of the workings of the seizure prediction system.

Since there is no gold standard number of unique states to distinguish between in ictal pattern recognition, an initial guess of 7 states were defined. This guess was made by visually determining sections of EEG signal that were most unique. To keep any confounding variables constant, the training and testing data used were taken from the same seizure case where 80% was used as training, and 20% was used as testing (random new permutations for every classification). The states that were initially defined are shown in Table 3.4. The distinction made between the Seizure Onset Period and the Preictal Period was done to determine if there was a more gradual transition between the preictal and ictal seizure states.

State #	Epileptic State
1	Normal, Calm EEG
2	Seizure Onset Period
3	Preictal
4	Ictal
5	Ictal (full seizure state)
6	Recovery
7	Aftershocks

Table 3.4: Initially defined epileptic states.

The results of this experiment were able to show the efficacy of the state decision neurons for making state transitions and the decision fusion which was used to improve the classification. Also, a module was created to segment the multichannel EEG signal, apply a window function, and pass it on to the system at the appropriate time intervals. This made it possible to emulate a more realistic scenario.

3.4.2   Detection of the Preictal State

The preictal state is a phenomenon that is impossible to see by basic means. This experiment tries to determine a feasible duration of the preictal state for which there is sufficient difference between this state and the interictal state. As a sub-study, the stepwise feature optimization algorithm is used in this experiment to determine the feature sets that result in the highest accuracy predicting the preictal state. This experiment was designed with the following question in mind: Where, within the 100 seconds prior to seizure onset, is the seizure mathematically visible? This time frame was given to localize the preictal state to a region of time that was not unreasonably short or long, but just enough for seizure intervention methods to be successfully executed.

For this and all subsequent experiments, the data provided by the Freiburg Seizure Prediction Project in Germany is used for testing. In this experiment, the training and testing data were partitioned from the same dataset; 80% was used as training data and 20% was used as testing data. The states were redefined as seen in Table 3.2 for this, and all subsequent experiments. The epileptic state definitions for all the EEG streams (specifically, States 1, 2, and 4) were defined by the respective researchers who provided the data for this research. State 3, or the preictal state, was initially defined to be one second before State 4. In this experiment, the duration of the preictal state was iteratively increased to 100 seconds, and a series of classifications were done at each step for each brain wave. At each time step, the feature set was optimized using the Stepwise Feature Optimization algorithm discussed in Section 3.3.1, and the best features for each iteration were saved. This gave an estimate as to how long a feasible preictal state (one that could be predicted) would be for each of the patients, and for each type of epilepsy.

3.4.3 Determination of the Seizure Prediction Horizon

The Seizure Prediction Horizon is an important parameter since it estimates the amount of time left before a transition to the ictal state. It essentially allows the patient to know how much time is left before seizure onset so that the appropriate measures can be taken. This experiment attempts to find the seizure prediction horizon for each patient based solely on the independent classifiers using the decision tree discussed in Section 3.3.3.

The definition of the states over the EEG signals for the data used as the training set were given by the researchers who gathered the data in Freiburg, Germany. Similar to the previous experiment, the preictal state was initially defined to be one second long and was iteratively increased to 100 seconds for each brain wave. A classification was performed at each step, and the results were tabulated to determine the optimal seizure prediction horizon. This experiment was much more difficult since the training set was independent of the testing set. No data about the testing set used was given to the classifiers.

Using the results of this experiment, new state definitions (classes) were defined for each of the seizure cases, and an estimation of the weight matrices for the state decision neurons were calculated. As mentioned before, the weight of a state neuron is a defined threshold transitioning from state i to state j. The weight of w_{23} was estimated as the accuracy of State 3 with the preictal state defined at the optimal seizure prediction horizon for each case, which is determined in this experiment. Also, as before, the Stepwise Feature Optimization algorithm was performed at each step and the best features at each iteration were saved to determine the feature sets that helped isolate interictal and preictal EEG data.

3.4.4   Emulator: Dynamic Classification Over Time

Using the knowledge gained from the previous experiments, a balance of all the algorithms, features, and other parameters were combined in order to come up with a prediction model capable of detecting the preictal state. In its fundamental form, this last experiment is an emulation of what would happen in a real situation and designed with a realistic point of view to make sure that the feasibility of constructing such a device with this algorithm would be preserved. In addition, the seizure prediction horizon was empirically defined as 65 seconds to safely provide enough preictal data for all of the seizure cases. This seizure prediction horizon was dynamically defined as each of the training sets were constructed for each case that was tested. For each testing case, the training data consisted of all other seizure cases from the same patient.

This emulator can be functionally reduced to the modules shown in Figure 3.1. An incoming signal is simulated by a module that supplies the emulator with the most recent 3-seconds of multi-channel EEG data every second. This segment is then sampled at 250 Hz and a rectangular window function is applied. An FIR filter is applied to the incoming EEG stream to decompose the incoming signals into their respective brain waves. However, due to time constraints, only the original signals (unfiltered) are tested with the system. All of the feature sets described in Table 3.3 are extracted from the signals as the testing sample. Since there are feature sets that compare each of the channels to each other, the emulator is supplied data segments from all of the channels at once for each 3-second segment. Following this, a PCA reduction of a new permutation of the training data is done where the training data used is a combination of all the seizure cases from the same patient except for the case being used as testing. PCA is also applied to the testing sample using the same parameters so that the testing data is properly scaled with the training data.

The LDA and KNN classifiers are then executed with the respective training and testing sets described, and their predictions are given to the decision fusion module. The decision fusion module then decides if the CVE or SVM classifiers should be used in order to obtain a majority vote on the epileptic state (see Figure 3.4). These predictions are tabulated and an initial prediction is given to the decision tree. The decision tree makes corrections to the initial prediction based on a set of heuristic rules that make sure the states transition in sequential increasing order. This fusion of the prediction methods is then given to the state decision neurons which use the past 32 predictions (for this system, the neuron populations each contain 32 neurons) to determine the final prediction of the current epileptic state and make decisions about when to make state transitions based on the defined weight matrix. Every testing case assumed that each of the other cases in the same patient had already occurred, and the average of the optimal weights for each of those cases were used for the testing case. This emulator was given 140 seconds of data prior to ictus up until 21 seconds after the ictal state and was tested to determine the accuracy with which each of the state transitions were made.

3.5   Resources

Since the use of many linear and non-linear techniques are applied in the prediction system, the use of a computing cluster was necessary for this research. All of the major data processing and experiments were carried out on the SHARCNET computing cluster at the University of Waterloo. This allowed for a very high number of concurrent experiments on many different data sets. With the use of this system, the computation time reduced from approximately 4 to 5 months of CPU time down to several days with the use of hundreds of concurrent processors

at any time.

All of the computational code for the seizure prediction system was written in MATLAB such that it could be executed on a Mac, PC, or any of the most popular Linux distributions. The coordination of files and job submissions for each of the experiments were done by a series of Bash shell scripts. Some scripts were also written to watch for failing jobs and resubmit them to the queue. Some code was also written in C++ for the initial processing of the signal data into smaller segments to make it easier for the computer to handle.

3.6 Summary

In this chapter, the methods and experiments used in this research are discussed. A description of the system developed is given and its core prediction algorithms are presented. All of the experiments are described and the resources that were needed are explained. The results of these methods and experiments are presented in the next chapter.

Chapter 4

Results

In this chapter, the results of the seizure onset detection system are presented as a progression of the research and experiments described in Chapter 3. Some initial testing was done using the data provided by the Krembil Neuroscience Center with which a framework was developed for epileptic seizure prediction. In this initial testing, seven classes that characterized the progression of seizure states were defined (as discussed in Section 3.4.1) for the system. From the initial testing of the framework developed, it was concluded that there were only four states with mathematically separable characteristics that were useful for seizure onset prediction. Following this, each of the major classification algorithms were analyzed for their ability to predict the preictal and ictal states using the data provided by the Freiburg Institute. The decision fusion method is added to show the improvement of the classification accuracy. The results of the experiments are then presented to find the duration of the seizure prediction horizon that resulted in the maximum accuracy in predicting the preictal state. These experiments also provided a study of the feature space to determine the mathematical techniques that were able to separate the four epileptic states defined.

Finally, the results from the seizure onset prediction emulator are presented. The response of the techniques that were outlined in Chapter 3 are shown. All of the seizure cases in the Freiburg datasets were tested using the system presented in this thesis, and the prediction accuracy for each case was tabulated.

4.1 Framework for EEG Signal Classification

This section presents the results of the preliminary testing done on the framework used to classify EEG signals. As described in Section 3.4.1, these initial tests were done using a tonic-clonic seizure case provided by the Krembil Neuroscience Center in Toronto, Canada. To keep any confounding variables constant, the training and testing data used were from the same case where 80% was used as training, and 20% was used as testing. The states defined for this testing are shown in Table 3.4.

As discussed in Chapter 3, there are four major classification algorithms throughout the system to determine the current state. Each classifier has its own strengths and weaknesses in terms of the accuracy of predicting each of the states that were defined. Initial analysis of each of the classifiers revealed the states that were more easily differentiable and also gave information about the response of each of the classifiers for every state. The results of the SVM classifier for the major transitions are shown in the following figures. The accuracy is shown for a reduction of the entire feature space to both two (2D) and five (5D) vectors using PCA.

Figure 4.1 shows that both 2D and 5D hyperplanes exist that are capable of isolating samples into their corresponding classes (states). There was also significant improvement in prediction accuracy as the feature space was increased from two (2D) PCA vectors to five (5D).

Within the framework of the seizure prediction system, the LDA and KNN
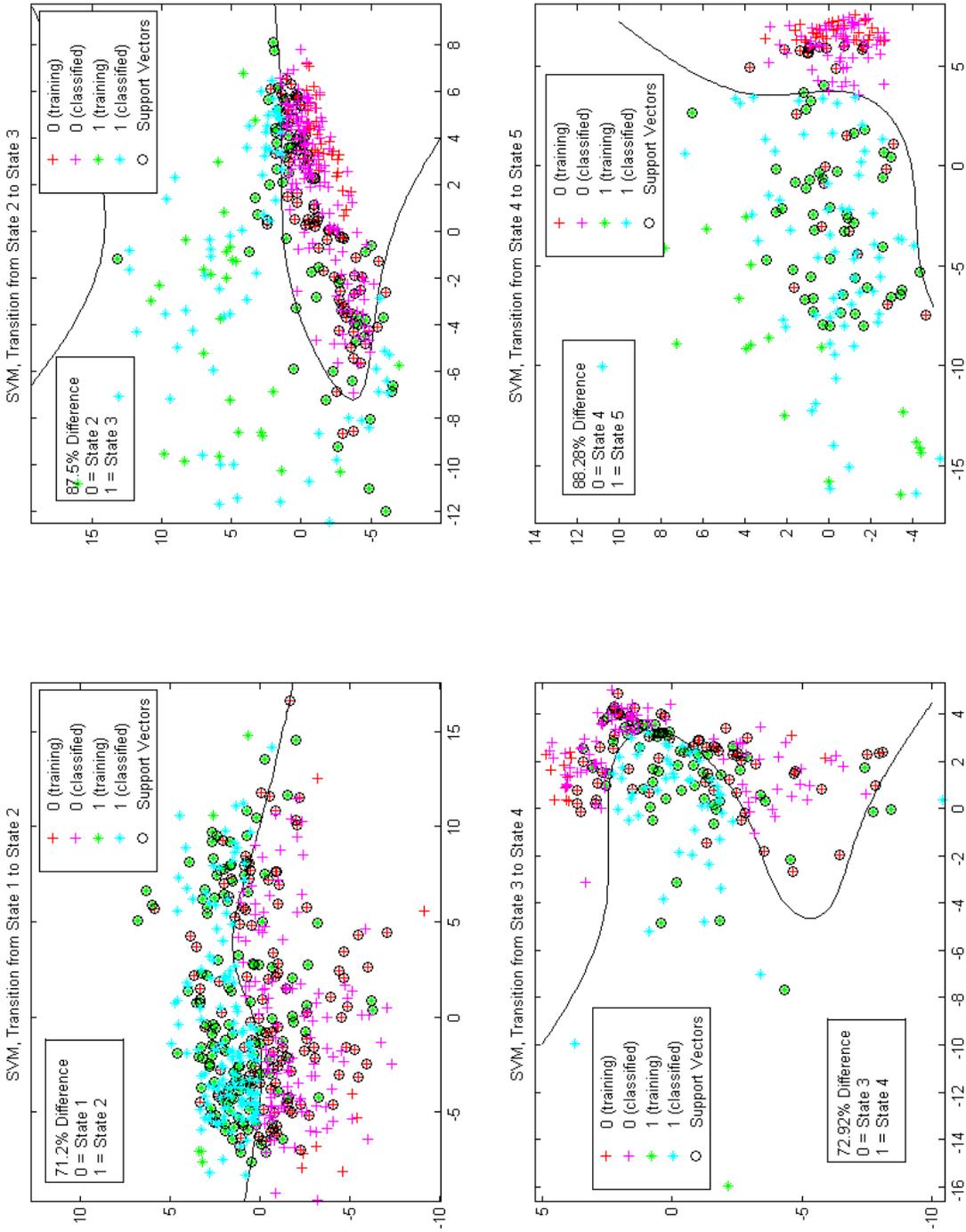
Figure 4.1: SVM State Transition Prediction Accuracy
State 1-2: 2D-71.2%, 5D-78.13%; State 2-3: 2D-87.5%, 5D-88.45%
State 3-4: 2D-72.92%, 5D-75.61%; State 4-5: 2D-88.28%, 5D-94.53%

classification algorithms are the primary prediction methods with which subsequent methods are dependent (see Section 3.3.3). It is important to note the strengths and weaknesses of these algorithms for each of the state transitions in order to establish the rules for the decision fusion module of the system. In the field of artificial intelligence, a confusion matrix is a visualization tool used to determine how a classifier responds to a set of classes. The results of these classifiers are given in the form of a confusion matrix where the rows denote the true states, and the columns identify the predicted states.

State	1	2	3	4	5	6	7
1	**89.579**	8.992	0.209	0	0.019	0.533	.667
2	8.696	**82.609**	0	0	0	0	8.69565
3	3.333	50	**40**	3.33	3.33	0	0
4	0	4.317	7.194	**88.489**	0	0	0
5	0	0.962	0	0	**54.808**	43.269	0.962
6	0	0	0	0	7.547	**75.472**	16.981
7	0	2.736	0	0	0	5.471	**91.793**

Table 4.1: Confusion Matrix for LDA classification algorithm

State	1	2	3	4	5	6	7
1	**93.398**	3.107	1.748	0	0.388	0.583	.777
2	62.963	**11.111**	3.703	0	0	0	22.222
3	35.714	3.571	**35.714**	21.429	0	0	3.571
4	0.714	0	0	**97.857**	1.429	0	0
5	5.263	1.051	0	1.053	**65.263**	14.7368	12.6316
6	12	2	0	0	14	**38**	34
7	5.743	1.014	1.689	1.014	2.702	2.027	**85.811**

Table 4.2: Confusion Matrix for KNN classification algorithm

The confusion matrices shown in Tables 4.1 and 4.2 show that the LDA classification algorithm was able to predict State 2 with 82.6% accuracy, while the KNN algorithm was only able to predict it with 11.1% accuracy. It is obvious that the KNN classification algorithm cannot be trusted to detect State 2, however, it is more reliable than LDA in predicting State 4 (ictal state). In this way, a set of rules were developed which make up the decision tree portion of the the decision

fusion module which would re-validate predictions that were unreliable (using SVM and CVE algorithms) and come up with a more suitable classification. These results also show that the LDA and KNN classification algorithms have significantly different responses in distinguishing the defined epileptic states.

This initial testing provided an initial understanding of the closed-loop response of the SDN framework. With 32 neurons in each population, the past 32 predictions made by the decision fusion module are considered in the decision to make a state transition. The SDN tuning curves, which are shown in Figure 4.2, are used to calculate the weight matrix for each patient.
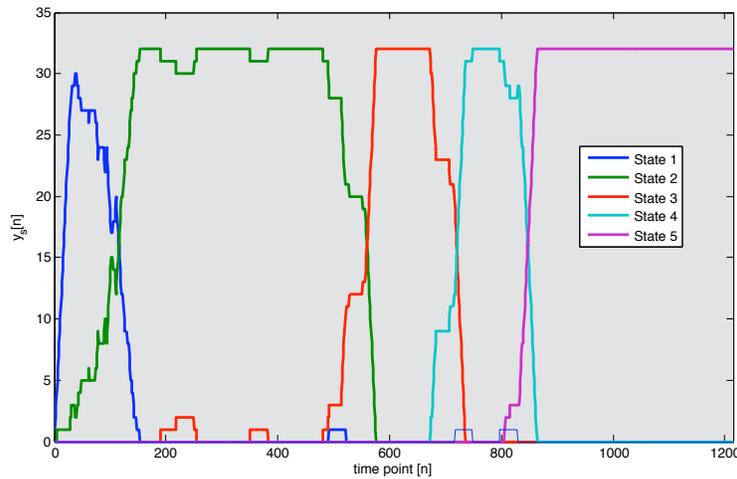
Figure 4.2: Krembil Data, SDN Tuning Curves.

Figure 4.2 shows that when past predictions are taken into consideration, each of the states have a distinct transition point that is discernible by the SDN framework. It is noticeable that the only transition that is gradual is from State 1 to State 2, illustrating that the transition between the non-seizure state and seizure onset has no definite boundary like the other states. As a result, the point on the gradient that defines the transition between the interictal and seizure onset states should be carefully chosen such that there is enough difference in the signal to not be confused

with the interictal state.

Figure 4.3 shows the progression of state transitions made by the system for the seizure case described for this experiment. The predicted state transitions happen at the times marked near each state transition, and the actual transitions occur according to the color coded regions of the EEG.
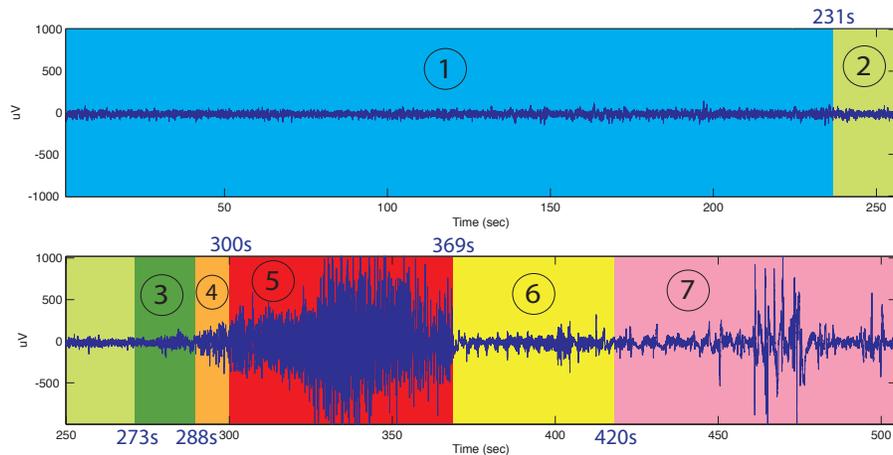
Figure 4.3: Initial Overall Results.

The simulation predicted seizure onset (State 2) at 231 seconds, which is approximately 57 seconds prior to ictus. This initial framework was defined to recognize 7 states including post-ictal aftershocks, however, less severe epileptic seizures are not usually followed by after-shocks. Due to this irregularity, it was determined that the definition of aftershocks as a state would cause more confusion to the classification algorithms and result in predictions that are less reliable. This initial testing on the framework of the system guided the intricate changes that were needed to build a system that could predict the onset of seizure attacks.

4.2 Independent Classifiers

In this section, the results of the KNN classification algorithm are provided to understand the system response to the preictal and ictal states. Each of the patients in the Freiburg dataset were tested to see the prediction accuracy for each of the brain waves: alpha, beta, delta, and theta (which were described in Section 2.2.1). The states have now been redefined to those listed in Table 3.2. For each patient, a PCA reduction of the entire feature space down to 19 vectors was used as the training set, which consisted of all the seizure cases from the same patient apart from the testing case. This test was a simple static classification of the entire testing data (no time dimension) to determine a generalized prediction accuracy.

Figure 4.5 shows the statistics of the prediction accuracy for each of the different brain waves for both State 3 and State 4. Table A.1 shows the prediction accuracy for each of the patients using the original signal and the four major brain waves described earlier, and Table A.2 shows the prediction accuracy for each of the different types of epilepsy in the same fashion, both of which are in Appendix A.
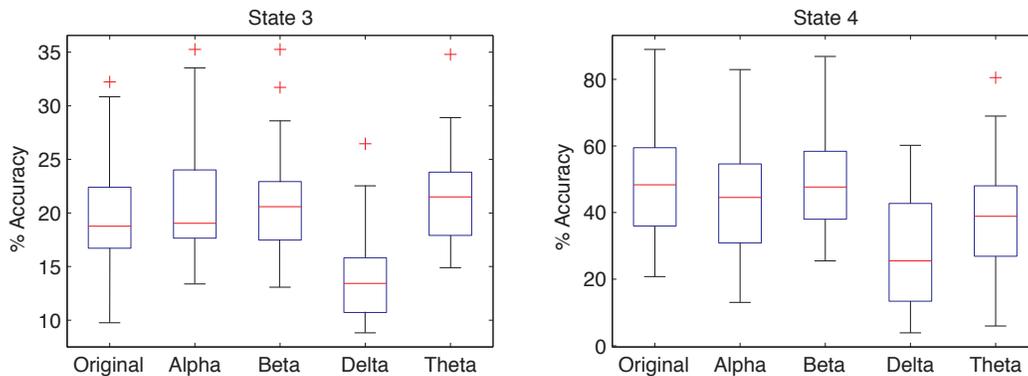
Figure 4.4: Classification accuracy without decision tree using all brain waves.

Tables A.1 and A.2 show that the prediction accuracy for these states while just using the classifier with static data are not high enough to be considered reliable

for either state. Figure 4.5 shows the statistics of the prediction accuracy for each of the different brain waves after using the decision tree. Tables A.3 and A.4 show the same results with the addition of the decision tree, both of which are located in Appendix A.
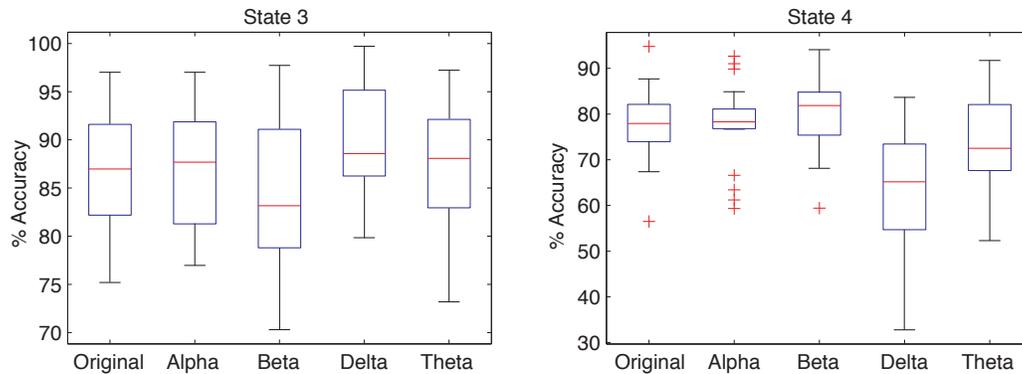
Figure 4.5: Classification accuracy using the decision tree for all brain waves.

The figure in this section show that the decision tree improves classification accuracy dramatically improving the overall reliability of the seizure prediction system to reasonable levels. Figure 4.5 shows that delta waves had the strongest response for all the patients during State 3, and Beta waves had the strongest response for all the patients during State 4. This suggests that these two brain waves may contain more information in their respective frequency ranges that are needed for epileptic seizure prediction.

4.3 Detection of the Preictal State

As described in Section 3.4.2, this experiment attempts to determine a feasible duration of the preictal state for which there is sufficient difference between this and the interictal state. As stated earlier, the training and testing data are taken from the same seizure case, but divided such that 80% was used as training and

20% was used as testing. The duration of the preictal state is iteratively increased from 1 to 100 seconds, and a series of classifications are performed at every point using each type of brain wave. At each time step, the feature set was optimized using the Stepwise Feature Optimization algorithm discussed in Section 3.3.1, and the best feature sets for each iteration were saved to determine the features that resulted in the highest classification accuracy.

4.3.1 Feature Selection Study

It is useful to know how each feature set influences the overall accuracy of the prediction system in order to optimize the prediction algorithm for the system. The prediction accuracy of each of the feature sets were tabulated for each of the seizure cases of each patient at the duration of the preictal state that yielded the highest accuracy (average times listed in Table B.1). The feature set numbers refer to the numbers listed in the first column of Table 3.3 and will be used throughout this chapter. Tables 4.3 and 4.4 show the results for both patient and epileptic type using each of the classifiers independently.

Tables 4.3 and 4.4 both emphasize the strengths and weaknesses of certain feature sets. Feature sets 14, 3, and 12 yielded strong influence in prediction accuracy while feature sets 1, 2, 6, and 10 showed the weakest response.

The prediction accuracies of the different brain waves were tabulated for each of the feature sets as an average of the results from all of the classification algorithms that were implemented and are shown in Tables 4.5 and 4.6.

Since a reduction of the alpha band denotes a loss of consciousness, an initial hypothesis was that the same would follow coming into a seizure and continue into the ictal state. From Table 4.6, it can be seen that the band with the highest classification accuracy is the alpha band, which suggests that the alpha band has

Patient	KNN		LDA		SVM	
	Max	Min	Max	Min	Max	Min
1	3(81%)	1(65.8%)	12(86.8%)	4(74%)	12(79.3%)	1(70.8%)
2	7(79%)	1(64.3%)	14(88.3%)	5(76.7%)	14(79.3%)	1(68%)
3	14(69.8%)	10(54.2%)	9(82.8%)	1(75.4%)	3(79.8%)	1(66.4%)
4	14(74.2%)	1(61%)	9(82.6%)	2(76%)	15(81%)	4(68.4%)
5	3(79.6%)	10(66%)	12(84.4%)	2(76.2%)	14(78.4%)	6(70.2%)
6	14(80.3%)	10(68%)	12(81.7%)	10(73.6%)	14(80.3%)	1(67.3%)
7	9(80.5%)	1(67.5%)	11(86.5%)	2(79%)	9(81.5%)	2(73%)
8	3(79.5%)	2(65%)	3(86.5%)	1(74.5%)	14(89.5%)	2(76.5%)
9	3(79%)	6(64.8%)	12(81.8%)	15(71.3%)	15(82.8%)	1(71%)
10	14(73.3%)	1(60%)	12(87.5%)	8(80.3%)	14(83%)	1(72.5%)
11	15(80.8%)	2(70.5%)	12(82%)	1(71.3%)	14(80.5%)	1(70.8%)
12	12(82%)	6(67%)	12(88.3%)	2(78.7%)	13(88.7%)	1(71.3%)
14	3(80.3%)	10(70.3%)	13(84.8%)	1(74%)	15(79.8%)	6(68.8%)
15	3(81%)	2(69.3%)	12(85%)	2(74.7%)	14(86%)	2(76%)
16	14(83%)	1(69.6%)	3(81%)	1(71%)	14(82.2%)	1(71%)
17	15(83.8%)	1(71%)	14(85.2%)	2(78%)	14(89.8%)	1(74.6%)
18	14(81.3%)	2(71.3%)	12(87.3%)	1(79.5%)	12(81.25%)	1(68%)
19	14(81.3%)	4(70%)	12(82.3%)	8(70.3%)	14(88.3%)	1(73.3%)
20	3(81.8%)	1(69.8%)	12(84.4%)	6(76%)	14(83.8%)	1(71%)
21	14(75.8%)	6(64%)	12(86.3%)	8(76.5%)	12(80.3%)	6(70.8%)

Table 4.3: Feature sets that resulted in the highest prediction accuracy for each patient.

Type	KNN		LDA		SVM	
	Max	Min	Max	Min	Max	Min
Frontal	14(78.9%)	1(69.1%)	12(84.7%)	1(77.1%)	14(82.1%)	1(72.7%)
Temporal	14(78.4%)	1(66.5%)	12(84.2%)	2(78.4%)	14(72.9%)	1(72.4%)
Parietal	15(80.8%)	2(70.5%)	12(82%)	1(71.3%)	14(80.5%)	1(70.8%)
FT	3(80.3%)	10(70.3%)	13(84.8%)	1(74%)	15(79.8%)	6(68.8%)
TO	14(79%)	10(66.4%)	12(81.7%)	2(72.9%)	14(80%)	1(69.2%)
TP	3(81.8%)	1(69.8%)	12(84.4%)	6(76%)	14(83.8%)	1(71%)
All	14(79.24%)	1(68.6%)	12(80.4%)	1(74.4%)	14(80.3%)	1(70.27%)

Table 4.4: Feature sets that resulted in the highest prediction accuracy for each type of epilepsy.

Patient	Original		Alpha		Beta		Delta		Theta	
	Feat.	%	Feat.	%	Feat.	%	Feat.	%	Feat.	%
1	12	86.75	12	81.25	14	80.33	14	76.8125	14	76.75
2	14	88.33	14	86.167	14	85.889	14	84.4167	14	83
3	9	82.4	14	81.3	14	81.8	14	79	14	79.16
4	9	82.6	14	79.4	14	80.53	14	78.7	14	78.6
5	12	84.4	12	83	12	82.667	12	79.3	12	79
6	12	81.667	14	80.833	14	80.22	14	80.167	14	80.53
7	11	86.5	11	85	11	83.667	12	80.5	12	79.9
8	14	89.5	14	88.25	14	84.667	12	83.625	14	83.4
9	15	82.75	12	82.125	12	82	14	79.313	14	79.4
10	12	87.5	12	86.125	12	85.083	12	82.625	12	82.45
11	12	82	3	81.25	3	81.583	12	79.313	12	79.6
12	13	88.667	14	86	12	85.889	12	83.667	12	83.267
14	13	84.75	14	83.25	14	83.167	14	81.813	14	81.75
15	14	86	14	82.33	14	84.111	14	81.083	12	81.067
16	14	83	14	83.1	14	81.2	14	80.75	12	80.92
17	14	89.8	14	84.8	14	84.533	14	83.25	14	83.52
18	12	87.25	3	85.125	12	86	12	81.375	12	81.95
19	14	88.333	14	88.167	14	86.111	14	84.167	14	84.4
20	12	84.4	12	83.6	12	83	12	82.7	12	82.48
21	12	86.25	12	83.25	12	83.417	12	78.938	12	78.7

Table 4.5: Feature sets resulting in the highest prediction accuracy in each brain wave for every patient.

Type	Original		Alpha		Beta		Delta		Theta	
	Feat.	%	Feat.	%	Feat.	%	Feat.	%	Feat.	%
Frontal	12	84.689	12	83.477	12	83.019	12	81.59	14	80.925
Temporal	12	84.187	14	83.194	14	83.148	14	82.291	14	81.74
Parietal	12	82	3	80.875	3	81.111	14	79.38	14	79.024
Temp./Occ.	12	81.708	12	80.344	12	80.174	14	79.374	14	79.493
Temp./Par.	12	84.4	12	83	12	82.733	12	82.725	12	82.676
Frontal/Temp.	13	84.75	14	84	14	83.722	14	82.62	14	81.886
All	12	83.757	12	82.52	12	82.28	12	81.154	12	80.415

Table 4.6: Feature sets resulting in the highest prediction accuracy in each brain wave for each type of epilepsy.

significant EEG alterations going into a seizure.

Table 4.7 gives a different permutation of the results to give a comparison of all the feature sets and their relative strengths and weaknesses. The results in bold show the maximum of their respective columns. As seen from this table, the same features that were deemed strong or weak before are shown similarly from this point of view.

Feature	KNN			LDA			SVM		
	Min(%)	Max(%)	StdDev	Min(%)	Max(%)	StdDev	Min(%)	Max(%)	StdDev
1	60	73	3.6211	71	83	3.67	67.33	84	3.848
2	64.25	75	3.409	71.5	86	3.592	70.67	79.2	2.414
3	63.5	82	4.613	74.34	86.5	2.802	73	84.67	2.979
4	63	77.25	3.632	74	84.5	3.154	68	83	4.186
5	61.5	78	3.692	74.25	85.25	3.008	70	82.34	3.55
6	61	79.2	4.698	73	84.34	3.251	68.75	82.6	3.995
7	64.75	79	3.661	75.5	86	2.75	69.75	86	4.427
8	63	78	3.99	70.34	84.6	3.28	69.67	86.67	4.729
9	66.5	82.2	4.674	74.34	85.5	3.156	72.5	83.5	3.34
10	61	82.8	5.12	73	86	3.158	70	87	4.696
11	65	80.4	4.26	75.8	86.5	3.142	74.4	87.34	4.444
12	63.25	82	5.049	77.8	**88.34**	3.145	74	88.5	3.83
13	66.75	79.5	4.17	74	86	3.137	73	88.67	4.263
14	73.25	83.4	2.641	76.67	**88.34**	2.98	76.25	**89.8**	4.254
15	67.25	**83.8**	4.253	71.25	87	3.618	74.34	88.2	4.264

Table 4.7: Prediction accuracy statistics of each of the feature sets for each classification algorithm.

Table 4.7 is useful in ranking the feature sets in terms of their relative strengths and weaknesses. It shows that feature sets 1 and 8 prove to be the weakest pair, while feature sets 12, 14, and 15 have the most potential in the prediction of the preictal state. Also, from the relative maximums listed for each of the classification algorithms, it is apparent that the LDA and SVM classifiers are more accurate in the recognition of the preictal state. This was also seen earlier in Section 4.1 in the results of the independent classifiers.

4.3.2 Optimal Duration of the Preictal State

Understanding how the preictal state behaves is necessary in order to train the system to recognize the transition between the interictal and preictal states (transition from State 2 to State 3). The optimal duration of the preictal state should be long enough that it has its own signal characteristics that are closer in relation to the ictal state, but are not enough to actually be classified as the ictal state.

Table B.1 in Appendix A shows the average duration of the preictal state for each of the patients which offer the maximum accuracy in the recognition of the preictal state. The overall statistics for the optimal duration of the preictal state are shown in Figure 4.6.
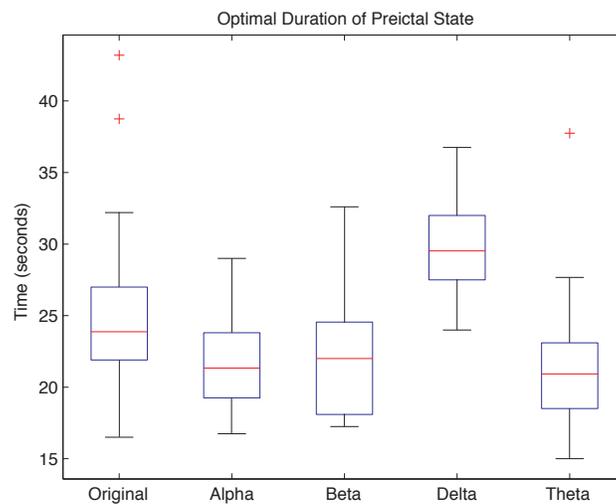
Figure 4.6: Optimal Duration of the preictal state using each brain wave.

Figure 4.6 shows that the duration of the preictal state that offers the maximum classification accuracy lies between 20 and 30 seconds before ictus. It is also clear that the delta band allows for the longest duration of the preictal state. The preictal response from a different point of view is presented in Figure 4.7, where the results of this experiment for one seizure case (Patient 19, case 69) are shown.
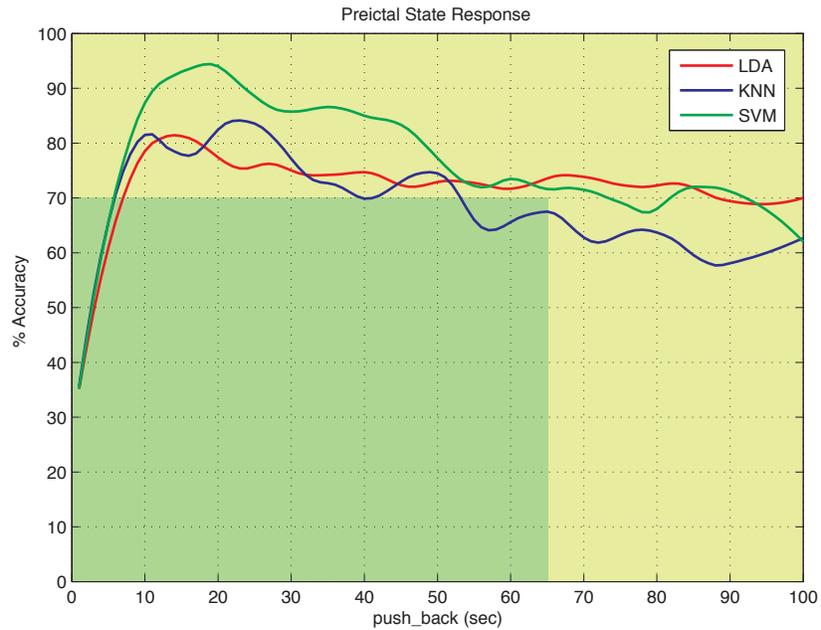
Figure 4.7: Determination of the Preictal State.

It can be seen in Figure 4.7 that the maximum classification accuracy occurs between 15 and 25 seconds for all three classification algorithms, and that the overall accuracy tapers off gradually after this point. All the seizure cases from the Freiburg dataset followed trends similar to this, and there was a noticeable tendancy for the mean accuracy of all the classification methods to drop below 70% after the preictal period reaches a duration of 65 seconds. As a result, this point served as an empirical guideline to maximize the accuracy of the seizure classification algorithm for all possible cases.

4.4 Determination of the Seizure Prediction Horizon

This experiment was used to determine the capability of predicting the preictal state. The prediction capability can be measured when the testing case used is independent of the seizure cases used as the training set. In order to reduce the number of confounding variables, the training set was limited to data collected from the same patient as the testing case. EEG waves are riddled with attributes that are unique to an individual, and using data from other people is unnecessary if significantly better data is available directly from the source, the patient.

4.4.1 Feature Selection Study

Similar to the previous experiment, the duration of the preictal state was iteratively increased from one to 100 seconds and a series of classifications were performed at each point for all the brain waves. Also, as in the previous experiment, the Stepwise Feature Optimization algorithm was executed at every time point and the best feature sets were saved at each iteration. Tables 4.8 and 4.9 show the feature sets that resulted in the highest classification accuracy for each patient and for each type of epilepsy that was processed.

Tables 4.8 and 4.9 both emphasize the strengths and weaknesses of certain feature sets. These results differ slightly from the results seen in the previous experiment due to the fact that the training data is completely independent of the testing data. Feature set 13, or the Hilbert transform, proved to be the most useful feature amongst all the patients processed with feature sets 10 and 12 just behind in terms of strength. Feature sets 1, 3, and 4 performed the worst relative to all of the other feature sets used.

Patient	KNN		LDA		SVM	
	Max	Min	Max	Min	Max	Min
1	10(59.5%)	3(49%)	12(57.75%)	1(47.75%)	11(43.75%)	3(27.25%)
2	7(61.67%)	1(51%)	15(67.33%)	1(51%)	9(70.33%)	1(41.67%)
3	14(63.4%)	15(57.6%)	11(47.2%)	12(39.4%)	14(52%)	3(43.2%)
4	15(67.6%)	2(57.4%)	15(56.2%)	8(50%)	10(63.2%)	14(58.8%)
5	15(55.4%)	8(48.4%)	12(51%)	1(43%)	9(49.2%)	6(40.8%)
6	13(63.67%)	6(51.33%)	6(36.67%)	1(28.67%)	10(67%)	2(53%)
7	9(59%)	4(40%)	12(52.5%)	6(40%)	5(53%)	2(40%)
8	9(58%)	3(35%)	6(31%)	3(15%)	13(85%)	3(70%)
9	12(71.75%)	2(60.75%)	7(24.25%)	1(17.25%)	6(69%)	4(56.5%)
10	10(65%)	6(54.75%)	15(35.5%)	12(28.25%)	11(50.75%)	12(44.5%)
11	3(73.75%)	12(60%)	5(63.25%)	13(50.5%)	14(46%)	3(33%)
12	13(60.67%)	12(51.67%)	10(42%)	3(31.67%)	8(57.67%)	12(43.67%)
14	12(69.75%)	14(55.75%)	9(42.75%)	4(31.75%)	15(39.75%)	3(30.25%)
15	12(75.33%)	4(62.33%)	14(54.33%)	2(40.33%)	5(59.67%)	3(47%)
16	15(58.2%)	12(42.8%)	7(32.2%)	3(22.6%)	8(57.8%)	12(48.8%)
17	10(75.4%)	11(69%)	8(57%)	1(31%)	11(60.8%)	3(51.6%)
18	13(51.25%)	2(36.5%)	7(60.5%)	3(50%)	7(48.75%)	12(39%)
19	6(46.67%)	3(31%)	13(37%)	3(23%)	14(52.67%)	3(41%)
20	14(70.8%)	2(59.4%)	9(38.8%)	12(27.2%)	14(60%)	12(42.4%)
21	11(52.75%)	12(40%)	10(39.75%)	14(28.75%)	3(58%)	14(46.25%)

Table 4.8: Feature sets that resulted in the highest prediction accuracy for each patient.

Type	KNN		LDA		SVM	
	Max	Min	Max	Min	Max	Min
Frontal	13(53.29)	3(44.53)	10(44.32)	3(37.73)	13(52.23)	3(44.98)
Temporal	13(61.11)	4(55.2)	10(45.73)	1(39.21)	13(56.79)	3(51.97)
Parietal	3(73.75)	12(60)	5(63.25)	13(50.5)	14(46)	3(33)
TO	12(66.54)	6(57.29)	6(29.83)	1(22.96)	10(67.75)	2(57.63)
TP	14(70.8)	2(59.4)	9(38.8)	12(27.2)	14(60)	12(42.4)
FT	12(69.75)	14(55.75)	9(42.75)	4(31.75)	15(39.75)	3(30.25)
All	13(60.33)	4(55.7)	10(43.1)	1(38.03)	13(54.34)	3(48.32)

Table 4.9: Feature sets that resulted in the highest prediction accuracy for each type of epilepsy.

Similar to the previous experiment, the prediction accuracies of the different brain waves were tabulated for each of the feature sets as an average of the results from all of the classification algorithms and are shown in Tables 4.10 and 4.11.

Patient	Original		Alpha		Beta		Delta		Theta	
	Feat.	%	Feat.	%	Feat.	%	Feat.	%	Feat.	%
1	10	59.5	13	58.875	13	58.333	13	60.375	14	61.9
2	9	70.333	13	70.667	9	68.444	15	68.417	13	67.4
3	14	63.4	14	70	14	68	14	70.05	14	70.04
4	15	67.6	3	64.5	15	65.8	13	66.9	15	66.24
5	15	55.4	15	62.6	15	60.73	15	58.35	15	59.44
6	10	67	15	67.33	14	67.78	13	65.33	13	65.93
7	9	59	9	57.75	11	62.167	11	59.75	11	55
8	13	85	13	78.25	13	61.5	13	62.38	13	65.5
9	12	71.75	12	74.5	12	74.83	15	69.875	15	69.75
10	10	65	10	67.125	10	61.917	10	61.563	12	63.8
11	3	73.75	10	66.25	13	68.33	13	67.5	15	66.65
12	13	60.67	10	61.167	15	59.22	15	58.167	15	61.8
14	12	69.75	3	72.875	12	70.25	12	70.313	12	71.8
15	12	75.33	14	73.33	12	69.11	12	72	14	73.13
16	15	58.2	8	61.9	13	61.533	13	63.9	8	64.2
17	10	75.4	10	72.2	14	70.33	14	71.2	14	71.2
18	7	60.5	15	53.75	13	54.417	10	54.75	15	52.15
19	14	52.67	14	54.5	14	56.11	14	57.167	14	58.4
20	14	70.8	14	69.1	14	68.53	14	68.4	13	69.48
21	3	58	3	61	13	63.167	10	65.813	13	64.3

Table 4.10: Feature sets resulting in the highest prediction accuracy in each brain wave for every patient.

Type	Original		Alpha		Beta		Delta		Theta	
	Feat.	%	Feat.	%	Feat.	%	Feat.	%	Feat.	%
Frontal	13	53.289	13	54.535	13	55.03	13	55.5	13	56.005
Temporal	13	61.11	13	62.23	13	62.35	13	62.46	13	62.636
Parietal	3	73.75	13	69.25	13	68.94	13	68.58	13	68.17
TO	10	67.75	15	67.58	15	67.875	15	67.787	15	67.738
TP	14	70.8	14	69.95	14	69.478	14	69.208	14	69.119
FT	12	69.75	3	70.938	12	70.58	12	70.516	12	70.773
All	13	60.33	13	61.302	13	61.508	13	61.69	13	61.92

Table 4.11: Feature sets resulting in the highest prediction accuracy in each brain wave for each type of epilepsy.

The overall results shown in Table 4.11 do not follow with the initial hypothesis about a reduction in alpha band patterns. Rather, it shows a higher prevalence in the theta band. Since the theta band is not supposed to be seen in normal adult

EEG tracings, the fact that it performed the highest hints that there is change in theta band activity during the ictal state. These results also show that feature set 13, or the Hilbert Transform, was the strongest feature set using independent training data.

Table 4.12 gives a different permutation of the results to give a comparison of all the feature sets and their relative strengths and weaknesses. The results in bold show the maximum of their respective columns.

Feature	KNN			LDA			SVM		
	Min(%)	Max(%)	StdDev	Min(%)	Max(%)	StdDev	Min(%)	Max(%)	StdDev
1	38.33	70.4	9.433	17.25	55.5	11.55	36	70.5	8.387
2	36.5	73.8	9.426	19.25	63	11.98	35.75	73.5	9.63
3	31	74.2	11.98	15	54.33	12.447	27.25	70	10.587
4	40	70.4	8.819	19.5	58.75	12.136	38.5	72	9.311
5	38	71	8.548	19.75	63.25	12.127	37	76	10.08
6	45.75	72.4	7.676	23	58.5	10.88	36.5	73.5	10.45
7	39.33	69.6	7.63	24.25	60.5	10.87	37.75	74	9.488
8	42	73.2	8.679	19.5	58.25	12.026	35.5	75.5	10.17
9	43.25	73.4	7.94	22	58	11.84	33	75	10.716
10	42.5	**75.4**	9.19	18.5	59.25	11.56	34.75	73	10.347
11	38.67	71.5	8.238	22.5	57.33	11.04	38	75.5	9.283
12	40	75.33	10.619	19	57.75	12.606	35.5	70	10.06
13	42.33	72.25	7.698	20.75	58.25	10.961	35.5	**85**	11.16
14	42.67	71.67	8.696	16.5	55.75	13.016	36	71.5	9.405
15	41.33	72.33	9.049	20	**67.33**	12.647	39.75	76	9.63

Table 4.12: Classifier results for all features.

As seen in Table 4.12, Feature set 3, or the Frequency Subbands, performed the worst, while feature sets 10, 13, and 15 performed the best. The results are similar to those of the previous experiment, but since the difficulty of the problem to solve in this experiment is significantly higher, the feature sets that stand out here are more distinguishable. It is also noticeable that the strength of the KNN classifier is higher than that of the LDA classifier in the classification accuracy of the preictal state. These results are different from the findings from Section 4.1 and should be held with more weight as it reflects a possible real-world application environment.

4.4.2 Determination of the Seizure Prediction Horizon

The seizure prediction horizon is basically an estimation of the duration of the pre-ictal state for which it can be recognized by the seizure prediction system. Knowing the trends of this variable is important so that when seizure onset is actually detected by the system, a forecast can be made as to the amount of time left before ictus. The decision tree is added in this experiment to improve the accuracy of the system similar to the way it is done in the emulator. Figure 4.8 shows the results of this experiment for one seizure case (patient 19, case 69).
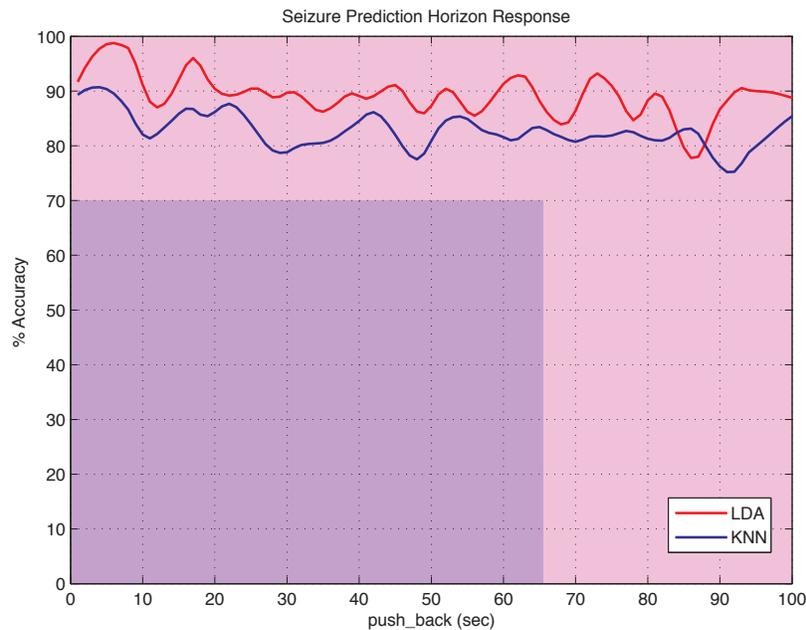
Figure 4.8: Finding the Seizure Prediction Horizon.

Figure 4.8 shows that the decision tree really improves the classification accuracy for the preictal state for this particular case. In fact, with the addition of the decision tree, the average combined prediction accuracy of all the classification algorithms doesn't drop below 70% for the duration of the experiment (100 seconds).

4.5 Emulator: Dynamic Classification Over Time

As a final test of the seizure prediction system, a program was written which could emulate the input of real-time EEG data and output the predicted seizure state flags as they occur. All of the parameters and knowledge that were gathered in the previous experiments are put to use in this final test and optimized to better suit a real-time scenario. For this test, the training data is a collection of all the seizure cases from one patient except for the one used as testing data (the input EEG stream to the system). The duration of the preictal state for all the cases that comprise the training data are all set to 65 seconds. In this way, the baseline for seizure onset prediction was set to 65 seconds prior to the ictal state. This variable was kept constant so that it was possible to measure the relative strength of the prediction system with several different seizure cases.

The SDN tuning curves (described in Section 3.3.2.2) shown in Figure 4.9 show the overall response of the emulation using data from patient 19, case 69 (Freiburg data). It shows that the state transitions are clear and maintains a relative strength of each state to a finite upper bound (the total number of neurons).

Figure 4.9 shows that State 4 begins to show signs approximately 70 seconds prior to the ictal state. Since there is no prediction of State 4 prior to this point, it shows that there are hints in the EEG of an upcoming ictal state which can be considered a preictal state. It also shows that the transition between the interictal and preictal states is gradual, not immediate, like the other state transitions. Additionally, the system was able to predict the end of the seizure approximately 10 seconds before the end of the ictal state. This is an assertion of the brain's natural homeostatic mechanisms that draw the brain waves back to the normal interictal state.
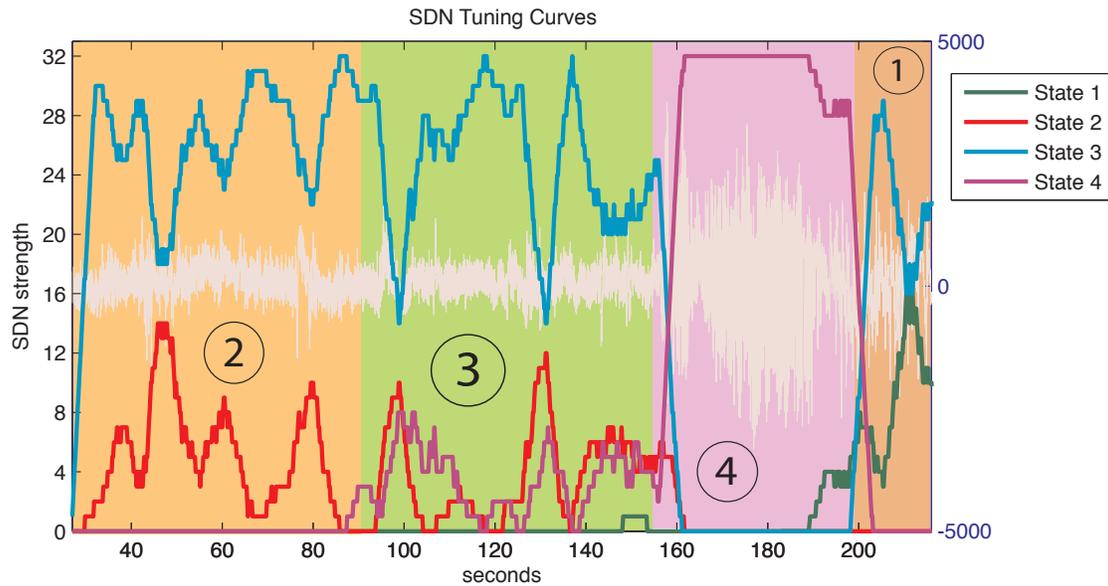
Figure 4.9: Patient 19, Case 69: SDN Tuning Curves.

All the patients in the Freiburg data set were processed in which the splice containing the ictal activity also contained continuous data for at least 140 seconds prior to the ictal state. Table 4.13 shows the final results of the emulator for each of the 71 cases that were processed. The columns 'S4 Error' and 'S1 Error' show the error, in seconds, for the prediction of State 4 and State 1, respectively. The accuracy is defined by the mean of the prediction rates for each of the four states.

Patient	Case	Onset	S4 Error	S1 Error	% Accuracy
1	7	60	4	2	90.3
1	8	72	8	6	82.45
1	13	38	1	0	79.19
1	15	79	1	0	91.41
2	15	58	-2	-10	90.8
2	18	91	12	0	85.34
2	21	63	9	4	94.75
3	105	102	5	-1	84.38
3	125	69	15	-68	81.29
3	127	48	2	1	93.15
3	174	77	5	3	92.46
3	176	26	5	5	76.61
4	204	82	-1	3	88.26
4	304	35	3	8	79.3
4	317	36	8	-1	81.59
4	323	75	2	0	95.64
4	328	85	19	6	81.53
5	13	56	1	11	80.78
5	24	27	0	2	84.35
5	26	86	2	0	86.8
5	33	29	1	3	83.42
6	11	80	5	0	93.05
6	23	35	2	1	84.41
6	30	35	2	1	87.47
7	36	59	30	21	69.2
7	372	44	-6	4	84.25
8	46	57	-5	21	69.68
8	69	58	-2	9	83.23
9	26	71	1	3	97.62
9	36	68	3	-50	79.07
9	38	62	3	-6	95.94
9	71	77	1	10	84.63
10	137	83	-18	-25	78.59
10	162	102	15	7	81.53
10	166	109	12	8	79.72
10	200	77	9	-6	88.12

Patient	Case	Onset	S4 Error	S1 Error	% Accuracy
11	24	96	14	3	91.91
11	25	65	0	3	95.8
11	31	69	1	6	90.64
12	15	89	5	-3	87.27
12	17	32	-3	1	85.32
12	19	103	7	2	74.37
14	7	40	5	21	63.35
14	15	112	4	21	56.86
14	17	70	13	9	87.44
14	19	42	13	21	64.12
15	9	47	3	21	67.7
15	41	75	9	21	69.06
15	88	93	7	21	63.77
16	41	65	5	3	96.3
16	48	68	3	7	80.94
16	83	60	6	4	94.06
17	110	20	2	5	78.54
17	115	56	3	7	96.13
17	123	99	5	2	84.52
18	11	64	4	4	92.35
18	13	78	4	1	89.55
18	14	69	5	3	79.12
18	19	44	3	-4	82.65
19	69	36	1	2	86.34
19	90	66	3	2	95.54
19	91	58	1	-3	84.52
20	13	110	2	0	82.08
20	25	87	1	0	90.99
20	55	67	2	0	96.81
20	82	48	1	1	93.27
20	94	77	3	-1	92.15
21	6	84	-2	10	81.2
21	11	66	4	-42	67.2
21	44	77	19	21	63.73
21	51	69	17	-10	90.68

Table 4.13: Accuracy statistics for each case processed through the seizure onset prediction emulator.

Table 4.13 shows that the emulator was able to predict seizure onset in the majority of the seizure cases. A select number of cases had difficulties in prediction, but this could be due to many different factors. For example, it can be calculated that Patient 19 had an average accuracy of 88.8%, an average onset prediction time of 53.33 seconds, and the average error for the State 4 and State 1 transitions are 1.667 and 0.33 seconds respectively. Patient 21, on the other hand, had an average accuracy of 75.7025%, an average onset prediction time of 74 seconds, and the average error for the State 4 and State 1 transitions are 9.5 and -5.25 seconds respectively. While the average onset prediction time was higher for Patient 21, the overall accuracy was not nearly as reliable as Patient 19. The type of epilepsy, specifically the way in which each type of epilepsy leads to the ictal state, could have significantly different tracings in the EEG leading up to the ictal state. Also, interference caused by characteristics of the damaged cortical areas could reduce accuracy for an individual patient.

4.6   Summary

Overall, the results of testing the seizure prediction algorithm show that the system has the potential to predict an oncoming seizure approximately 65 seconds prior to the ictal state. The mean accuracy of the system is 84.17%, minimum is 56.86%, maximum is 97.63%, and the standard deviation is 9.43. The average onset prediction time is 65.8551 seconds prior to the ictal state. The techniques that were most useful in the feature space were the Phase Shift Correlation features, the Hilbert Transform Statistics, and the Lyapunov Exponents.

Chapter 5

Summary & Conclusions

In this chapter, a summary of the research will be presented showing the strengths and weaknesses of the seizure onset prediction system. Following this, a possible implementation of such a system is given. Lastly, some additional areas for research in the field of seizure onset prediction are given.

5.1   Summary

This thesis presented several new methods and algorithms to enhance the prediction of ictal states. Several experiments were proposed to determine the behavior of a possible preictal state. A study of the feature sets that maximized the accuracy was completed. Finally, an emulator was developed to test the proposed algorithms in a real-time environment.

The results of the experiments performed show that there are definite warning signs in the EEG prior to an ictal state that can be predicted using appropriate mathematical methods. The feature space study showed that the chaos and coherence feature sets were the most valuable in seizure onset prediction. The results from the emulator show that the existence of a preictal state begins to show ap-

proximately 65 seconds prior to an ictal state, as seen in Figure 4.9. Overall, the prediction system was successful at predicting seizures prior to ictus. A few cases showed some anomalies or irregularities, but this shows that seizure activity within one patient may not always be the same every time depending on the type of seizure or just patient anomalies.

Since there is usually cortical damage causing the seizures, there is an indeterminate number of possible confounding variables. For this reason, the prediction system was designed to learn and predict seizures from within the same patient.

5.2   Implementation

In order to make this system accessible to epileptic patients, a small device with this algorithm could be implemented. Six probes, three on the epileptogenic focus, and three on the opposite lobe, would have to be installed on the patient which would input the signals into the processing unit, possibly by a wireless protocol such as bluetooth or WiFi. The processing unit would have to be attached to the body in a discrete manner, such as a belt or something that can be worn at all times.

At the time point of each state transition, the appropriate flags are set on the processing unit which trigger indicators. For a state transition to seizure onset, an indicator should go off and the unit should begin to vibrate. At this point, if it is possible to administer AEDs directly to the brain, it should be done either manually or automatically through an implanted drug reservoir. If the patient then transitions to an ictal state regardless of the treatment, an alarm should sound to warn others and appropriate medical attention should be called upon, automatically. The device should record the EEG data for each seizure that happens

and use it later for learning. It should also store some metadata about the seizure along with some vitals such as body temperature and heart rhythm. This type of data could help doctors better understand a patient's condition as well as the progression and development of the condition over time.

Since the system has to learn from the patient, the first few seizures will be predicted by a training set of EEG data from the same type of epilepsy. Once enough seizures have occurred for the device to properly learn the EEG characteristics of the patient, it will replace the initial data with data collected directly from the patient. Once the device has enough data, it should be able to predict seizures and also keep replacing its own data with fresh data to compensate for changes in the EEG of a patient over time.

5.3   The Next Steps

With this research, seizure prediction has been reduced to less than 10 variables. Improving the method of optimizing all the necessary variables would make the system work more efficiently and would result in less time with a specialist. Also, using ICA to remove signal artifacts (*i.e.* eye movements) from the EEG may have a positive effect on accuracy. Lastly, research must be done on the transition from the ictal state back to the interictal state to see how long after a seizure the EEG returns completely back to normal, allowing the system to learn to reset itself accurately and automatically. These additions would help reduce the number of false positives and make the implementation more robust and reliable.

APPENDICES

Appendix A

Tables: Independent Classifiers

Patient	Original		Alpha		Beta		Delta		Theta	
	State 3	State 4	State 3	State 4	State 3	State 4	State 3	State 4	State 3	State 4
1	15.915	36.803	17.361	13.002	15.972	25.505	12.974	3.907	17.172	6.034
2	20.664	59.635	17.866	59.050	22.412	58.244	10.415	38.406	16.919	52.754
3	18.586	61.018	33.535	65.828	24.394	71.538	15.950	9.552	24.647	62.406
4	20.455	72.512	22.879	78.362	31.717	79.545	14.574	60.176	18.260	68.995
5	17.813	44.650	21.061	30.699	20.96	45.312	13.651	15.310	22.475	32.573
6	16.174	40.749	25.152	43.3	23.232	46.106	13.629	12.403	22.172	35.035
7	17.102	35.156	15.707	39.829	18.99	39.774	11.01	33.612	14.899	27.564
8	20.627	44.486	15.556	32.5	20.606	41.532	9.755	20.953	18.232	30.741
9	30.838	59.318	35.253	53.195	28.586	54.883	18.129	25.065	34.798	42.772
10	17.15	28.469	19.141	23.649	13.131	27.888	15.692	21.564	22.020	26.213
11	23.899	59.6	22.273	45.766	16.919	63.613	16.309	55.377	23.687	47.004
12	21.136	51.997	30.101	48.405	15.455	55.115	8.823	47.6	25.909	46.982
14	18.259	53.830	18.498	47.728	20.96	49.662	11.302	47.194	20.96	45.809
15	18.954	54.414	18.939	48.858	20.581	49.249	11.848	47.036	18.308	46.055
16	27.903	59.094	18.889	55.977	19.495	58.586	22.538	37.984	19.343	49.054
17	23.66	38.918	21.67	33.321	22.626	36.267	14.846	36.709	23.939	33.837
18	32.235	88.976	26.111	82.918	35.253	86.899	26.461	25.887	28.889	80.512
19	9.758	30.581	13.384	15.41	18.030	32.03	9.703	7.643	17.597	5.964
20	16.312	35.113	18.889	19.311	19.95	45.29	13.227	10.64	23.395	14.205
21	15.351	20.78	17.424	31.101	13.068	28.06	8.95	14.363	15.537	22.7

Table A.1: Patient results for classification accuracy without decision tree using all brain waves.

Type	Original		Alpha		Beta		Delta		Theta	
	State 3	State 4	State 3	State 4	State 3	State 4	State 3	State 4	State 3	State 4
Frontal	19.156	51.086	21.168	40.06	22.536	50.469	14.749	13.875	21.502	36.372
Temporal	20.264	46.775	20.29	46.506	19.719	48.081	13.188	37.494	19.46	41.573
Parietal	23.899	59.6	22.273	45.766	16.919	63.613	16.309	55.377	23.687	47.004
TO	23.506	50.033	30.202	48.247	25.909	50.494	15.879	18.734	28.485	38.903
TP	16.312	35.113	18.889	19.311	19.95	45.29	13.227	10.64	23.395	14.205
FT	18.259	53.83	18.498	47.728	20.96	49.662	11.302	47.194	20.96	45.809
All	20.14	48.805	21.484	43.41	21.117	49.755	13.989	28.569	21.458	38.86

Table A.2: Epileptic type results for classification accuracy without decision tree using all brain waves.

Patient	Original		Alpha		Beta		Delta		Theta	
	State 3	State 4	State 3	State 4	State 3	State 4	State 3	State 4	State 3	State 4
1	95.454	72.532	96.654	66.616	96.780	70.866	97.435	53.62	97.222	62.765
2	81.5	86.742	81.124	82.656	80.24	84.175	88.551	67.5	85.164	84.135
3	82.879	78.033	89.293	89.882	90.909	89.918	88.452	59.501	84.849	86.339
4	92.727	86.104	92.071	91.043	91.768	91.82	88.574	82.139	91.259	87.527
5	96.916	87.694	97.020	78.031	96.97	84.896	99.707	72.717	96.616	65.909
6	87.042	74.842	90.657	79.035	90.707	78.365	93.154	54.959	90.808	72.478
7	81.417	73.695	76.97	76.696	77.475	77.576	89.186	50.900	79.748	69.368
8	79.952	77.084	77.475	77.691	78.939	75.268	79.843	76.269	78.636	72.547
9	86.663	77.855	86.869	78.668	78.636	82.056	94.682	49.942	90.859	73.379
10	75.199	70.753	78.232	63.430	77.424	68.132	81.679	62.664	76.313	63.582
11	84.747	81.796	91.667	77.817	70.303	82.035	87.325	83.656	88.182	83.092
12	81.323	82.421	87.424	84.878	80.455	85.003	80.159	74.166	83.131	81.104
14	91.598	79.448	87.689	78.021	90.909	81.844	85.909	68.065	87.816	78.088
15	90.526	80.433	87.689	79.589	91.288	80.63	85.901	70.55	88.384	79.477
16	84.492	76.83	77.778	78.577	77.020	75.478	92.535	62.831	73.182	71.516
17	86.898	74.216	85.808	76.936	83.182	73.293	88.577	70.956	87.98	69.478
18	97.02	94.790	96.869	92.666	97.727	94.103	96.204	58.176	94.95	91.752
19	87.494	56.499	93.03	61.229	83.131	59.412	95.617	32.796	92.973	52.299
20	91.595	67.383	87.879	59.340	82.778	81.846	97.755	54.444	95.348	59.36
21	87.871	78.394	81.439	78.62	85.038	84.753	86.577	75.565	82.754	71.594

Table A.3: Patient results for classification accuracy without decision tree using all brain waves. (after decision tree)

Type	Original		Alpha		Beta		Delta		Theta	
	State 3	State 4	State 3	State 4	State 3	State 4	State 3	State 4	State 3	State 4
Frontal	89.952	77.772	91.724	77.686	90.743	79.077	92.876	58.847	90.874	71.935
Temporal	84.661	78.843	83.171	79.159	82.654	80.096	86.86	68.586	83.102	75.309
Parietal	84.747	81.796	91.667	77.817	70.303	82.035	87.325	83.656	88.182	83.092
TO	86.853	76.349	88.763	78.852	84.672	80.211	93.918	52.451	90.833	72.928
TP	91.595	67.383	87.879	59.34	82.778	81.846	97.755	54.444	95.348	59.36
FT	91.598	79.448	87.689	78.021	90.909	81.844	85.909	68.065	87.816	78.088
All	87.166	77.877	87.182	77.571	85.084	80.074	89.891	64.071	87.309	73.79

Table A.4: Epileptic type results for classification accuracy without decision tree using all brain waves. (after decision tree)

Appendix B

Tables: Detection of the Preictal State

Patient	Original (sec)	Alpha (sec)	Beta (sec)	Delta (sec)	Theta (sec)
1	29	24.5	28	30.25	21.75
2	22	20	22	28.67	27.67
3	25	18	23.2	25.4	15.6
4	27	19	23.6	28.2	26.8
5	21	16.75	18	30.75	19.5
6	21.67	24.67	21.67	33	16.67
7	22	19.5	22	24	20.5
8	25	29	18	27	24.5
9	21	17.75	20.75	31.25	17.5
10	24.75	21	18.5	28.5	22.75
11	16.5	17.25	17.25	36.75	15
12	22.34	22.34	26	25.67	18.34
14	22.	19.5	17.5	34.25	21.5
15	27	22.34	17.67	30.67	21.34
16	21.8	22.6	18.2	26.2	22.8
17	43.2	23.6	32.6	28.8	20
18	38.75	24.5	24.25	31.75	37.75
19	23	21.67	24.34	35.34	18.67
20	32.2	24	27	28	23.4
21	27	20.5	24.75	32.25	19.25

Table B.1: Optimal duration of the preictal state for each type of brain wave.

References

[1] R. Andrzejak, J. Arnhold, T. Kreuz, and F. Mormann. Nonlinear EEG Analysis in Epilepsy: Its Possible Use for Interictal Focus Localization, Seizure Anticipation, and Prevention. *Journal of Clinical Neurophysiology*, 18(3):209–222, Jan 2001. 31

[2] R. Andrzejak, K. Lehnertz, F. Mormann, and C. Rieke. Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state. *Physical Review E*, 64(6):061907, Nov 2001. 36

[3] J. Aziz, R. Karakiewicz, R. Genov, A. Chiu, B. Bardakjian, M. Derchansky, and P. Carlen. In Vitro Epileptic Seizure Prediction Microsystem. In *IEEE International Symposium on Circuits and Systems, ISCAS 2007*, pages 3115–3118, May 2007. 32

[4] K. Ballaban-Gil, C. Callahan, C. O'Dell, and M. Pappo. Complications of the Ketogenic Diet. *Epilepsia*, 39(7):744–748, Aug 1998. 7

[5] R. Beran. Epilepsy and law. *Epilepsy and Behavior*, 12(4):644–651, May 2008. 2

[6] C. Castellaro, G. Favaro, A. Castellaro, A. Casagrande, S. Castellaro,

D. Puthenparampil, and C. Salimbeni. An artificial intelligence approach to classify and analyse EEG traces. *Neurophysiologie Clinique/Clinical Neurophysiology*, 32(3):193 – 214, 2002. 14

[7] K. Chong. *Neil Young Nation: A Quest, an Obsession, and a True Story.* Greystone Books, Jan 2005. 3

[8] V. Cizek. Discrete Hilbert Transform. *Audio and Electroacoustics*, 18(4):340–343, Dec 1970. 19

[9] J. Cooley. The Re-Discovery of the Fast Fourier Transform Algorithm. *Microchimica Acta*, 93(1-6):33–45, Jan 1987. 16

[10] M. D'Alessandro, G. Vachtsevanos, A. Hinson, R. Esteller, J. Echauz, and B. Litt. A Genetic Approach to Selecting the Optimal Feature for Epileptic Seizure Prediction. In *Proceedings of the 23rd Annual International Conference of the IEEE in Engineering in Medicine and Biology Society, EMBS '01*, volume 2, pages 1703 – 1706, Istanbul, Turkey, Sep 2001. 34

[11] A. Darbyshire. Calculating Lyapunov Exponents from a Time Series. In *IEE Colloquium on Exploiting Chaos in Signal Processing*, pages 2/1 – 2/6, London, Jun 1993. 25

[12] M. Eadie and P. Bladin. *A Disease Once Sacred: A History of the Medical Understanding of Epilepsy.* John Libbey, Jan 2001. 1

[13] M. El-Gohary, J. McNames, T. Ellis, and B. Goldstein. Time Delay and Causality in Biological Systems Using Whitened Cross-Correlation Analysis. In *28th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS '06*, pages 6169 – 6172, New York, Jan 2006. 23

[14] J. Engel, T. Pedley, J. Aicardi, M. Dichter, and S. Moshé. *Epilepsy: A Comprehensive Textbook*. Lippincott Williams & Wilkins, Jan 2007. 7, 8

[15] F. Freemo. A Differential Diagnosis of the Inspirational Spells of Muhammad the Prophet of Islam. *Epilepsia*, 17(4):423 – 427, Jan 1976. 3

[16] A. Girgis and F. Ham. A Quantitative Study of Pitfalls in the FFT. *IEEE Transactions on Aerospace and Electronic Systems*, AES-16(4):434 – 439, Jul 1980. 13

[17] V. Golovko, S. Bezobrazova, S. Bezobrazov, and U. Rubanau. Application of Neural Networks to the Electroencephalogram Analysis for Epilepsy Detection. In *International Joint Conference on Neural Networks, IJCNN 2007*, pages 2707 – 2711, Jul 2007. 32

[18] N. Gopalsami, I. Osorio, S. Kulikov, S. Buyko, A. Martynov, and A. Raptis. SAW Microsensor Brain Implant for Prediction and Monitoring of Seizures. *Sensors Journal, IEEE*, 7(7):977 – 982, Jul 2007. 34

[19] D. Gupta, C. James, and W. Gray. Phase Synchronization with ICA for Epileptic Seizure Onset Prediction in the Long Term EEG. In *4th IET International Conference on Advances in Medical, Signal and Information Processing, MEDSIP 2008*, pages 1 – 4, Jun 2008. 33

[20] S. Haas, M. Frei, and I. Osorio. Strategies for adapting automated seizure detection algorithms. *Medical Engineering and Physics*, 29(8):895–909, Oct 2007. 33

[21] L. Hively and V. Protopopescu. Channel-Consistent Forewarning of Epileptic Events from Scalp EEG. *IEEE Transactions on Biomedical Engineering*, 50(5):584 – 593, May 2003. 8, 33

[22] J. Hughes. Emperor Napoleon Bonaparte: Did he have seizures? Psychogenic or epileptic or both? *Epilepsy and Behavior*, 4(6):793–796, Dec 2003. 3

[23] J. Hughes. Did all those famous people really have epilepsy? *Epilepsy and Behavior*, 6(2):115–139, Mar 2005. 3

[24] J. Hughes. The idiosyncratic aspects of the epilepsy of Fyodor Dostoevsky. *Epilepsy and Behavior*, 7(3):531–538, Nov 2005. 3

[25] L. Iasemidis and J. Sackellares. Seizure Warning and Prediction. *US Patent 6,304,775*, Oct 2001. 34

[26] L. Iasemidis, J. Sackellares, and H. Zaveri. Phase Space Topography and the Lyapunov Exponent of Electrocorticograms in Partial Seizures. *Brain Topography*, 2(3):187–201, Mar 1990. 32

[27] A. Isaksson, A. Wennberg, and L. Zetterberg. Computer analysis of EEG signals with Parametric Models. In *Proceedings of the IEEE*, pages 451–461, Apr 1981. 34

[28] A. Jacoby, J. Gorry, and G. Baker. Employers' Attitudes to Employment of People with Epilepsy: Still the Same Old Story? *Epilepsia*, 46(12):1978–1987, Nov 2005. 2

[29] I. Jolliffe. *Principal Component Analysis*. Springer, Jan 2002. 26

[30] Z. Junling and J. Dazong. A linear epileptic seizure predictor based on slow waves of scalp EEGs. In *27th Annual International Conference of the Engineering in Medicine and Biology Society, EMBS '05*, pages 7277 – 7280, Aug 2005. 31

[31] S. Kay. *Modern Spectral Estimation: Theory and Application.* Prentice Hall, Jan 1988. 24

[32] V. Kecman. *Learning and Soft Computing: Support Vector Machines, Neural Networks, and Fuzzy Logic Models.* MIT Press, Jan 2001. 30

[33] P. Kraniauskas. A plain man's guide to the fft. *Signal Processing Magazine, IEEE*, 11(2):24 – 35, Apr 1994. 15

[34] D. Krug, C. Elger, and K. Lehnertz. A CNN-Based Synchronization Analysis for Epileptic Seizure Prediction: Inter- and Intraindividual Generalization Properties. *Cellular Neural Networks and Their Applications, 2008. CNNA 2008. 11th International Workshop on*, pages 92 – 95, Jun 2008. 33

[35] Y. Lai, M. Harrison, M. Frei, and I. Osorio. Inability of Lyapunov Exponents to Predict Epileptic Seizures. *Physical Review Letters*, 91(6):068102, Aug 2003. 32

[36] I. Leppik. *Epilepsy: A Guide to Balancing Your Life.* Demos Medical Publishing, LLC, Jan 2006. 3

[37] B. Litt. Engineering Devices to Treat Epilepsy: A Clinical Perspective. In *Proceedings of the 23rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS '01*, volume 4, pages 4124 – 4128, Sep 2001. 4

[38] B. Litt, A. D'Alessandro, R. Esteller, J. Echauz, and G. Vachtsevanos. Translating Seizure Detection, Prediction and Brain Stimulation Into Implantable Devices for Epilepsy. In *Conference Proceedings in the 1st International IEEE EMBS Conference on Neural Engineering*, pages 485 – 488, Feb 2003. 31

[39] B. Litt and K. Lehnertz. Seizure Prediction and the Preseizure Period. *Current Opinion in Neurology*, 15(2):173–177, Apr 2002. 11

[40] Q. Loh and S. Emmanuel. ELM for the Classification of Music Genres. In *9th International Conference on Control, Automation, Robotics and Vision, ICARCV '06.*, pages 1 – 6, Nov 2006. 17

[41] T. Maiwald, M. Winterhalder, R. Aschenbrenner-Scheibe, H.Voss, A. Schulze-Bonhage, and J. Timmer. Comparison of three nonlinear seizure prediction methods by means of the seizure prediction characteristic. *Physica D: Nonlinear Phenomena*, 194(3-4):357 – 368, Jan 2004. 31

[42] J. Makhoul. Linear Prediction: A Tutorial Review. In *Proceedings of the IEEE*, volume 63, pages 561 – 580, Apr 1975. 20

[43] P. McSharry, L. Smith, and L. Tarassenko. Comparison of Predictability of Epileptic Seizures by a Linear and a Nonlinear Method. *IEEE Transactions on Biomedical Engineering*, 50(5):628 – 633, May 2003. 32

[44] M. Mishra, B. Jones, J. Simonotto, M. Furman, W. Norman, Z. Liu, T. De-Marse, P. Carney, and W. Ditto. Pre-Ictal Entropy Analysis of Microwire Data from an Animal Model of Limbic Epilepsy. In *28th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS '06*, pages 1605 – 1607, Jan 2006. 33

[45] E. Molteni, P. Perego, N. Zanotta, and G. Reni. Entropy analysis on EEG signal in a case study of focal myoclonus. In *30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS '08*, pages 4724 – 4727, Jul 2008. 33

[46] S. Nagaraj, A. Shah, P. Shah, V. Szeto, and M. Bergen. Ambulatory Pre-seizure Detection Device. In *Proceedings of the IEEE 32nd Annual Northeast in Bioengineering Conference*, pages 41 – 42, Mar 2006. 33

[47] V. Nhat and S. Lee. *Computer Recognition Systems*. Springer Berlin / Heidelberg, 2005. 27

[48] C. Niederhöfer, F. Gollas, and R. Tetzlaff. Dynamics of EEG-signals in epilepsy: Spatio temporal analysis by Cellular Nonlinear Networks. In *18th European Conference on Circuit Theory and Design, ECCTD 2007*, pages 296 – 299, Jul 2007. 31

[49] C. Niederhöfer and R. Tetzlaff. Prediction Error Profiles Allowing a Seizure Forecasting in Epilepsy? In *10th International Workshop on Cellular Neural Networks and Their Applications, CNNA '06*, pages 1 – 6, Jul 2006. 31

[50] I. Osorio, M. Frei, and S. Wilkinson. Real-Time Automated Detection and Quantitative Analysis of Seizures and Short-Term Prediction of Clinical Onset. *Epilepsia*, 39(6):615 – 627, Jan 1998. 34

[51] J. Pellock, W. Dodson, B. Bourgeois, R. Sankar, and D. Nordli. *Pediatric Epilepsy: Diagnosis and Therapy*. Demos Medical Publishing, LLC, Jan 2008. 7

[52] A. Pitkänen, P. Schwartzkroin, and S. Moshé. *Models of Seizures and Epilepsy*. Academic Press, Jan 2005. 10, 11, 12

[53] A. Prasad, A. Spanias, J. Sackellares, and L. Iasemidis. Predictability of epileptic seizures: a comparative study using Lyapunov exponent and entropy based measures. *Biomedical Sciences Instrumentation*, 39:129–135, Jan 2003. 32

[54] A. Pujari. *Data Mining Techniques*. Orient Blackswan, Jan 2001. 30

[55] J. Reynolds. Epilepsy: Its Symptoms, Treatment, and Relation to Other Chronic Convulsive Diseases. *Churchill*, page 360, Jan 1861. 7

[56] M. Rise. Commercialization of Seizure Prediction Technology Promises and Pitfalls of Biosignal Analysis: Seizure Prediction and Management (A case study). In *Engineering in Medicine and Biology Society, 2001. Proceedings of the 23rd Annual International Conference of the IEEE*, volume 4, pages 4129 – 4130, Sep 2001. 4

[57] J. Sackellares. Seizure Prediction. *Epilepsy Currents / American Epilepsy Society*, 8(3):55–9, Jan 2008. 11

[58] S. Sanei and J. Chambers. *EEG Signal Processing*. Wiley-Interscience, 2007. 8, 9, 10

[59] G. Scambler. *Epilepsy*. Routledge, Jan 1989. 2

[60] B. Schelter, M. Winterhalder, T. Maiwald, A. Brandt, A. Schad, J. Timmer, and A. Schulze-Bonhage. Do False Predictions of Seizures Depend on the State of Vigilance? A Report from Two Seizure-Prediction Methods and Proposed Remedies. *Epilepsia*, 47(12):2058–2070, Nov 2006. 50

[61] R. Schuyler, A. White, K. Staley, and K. Cios. Epileptic Seizure Detection. *Engineering in Medicine and Biology Magazine, IEEE*, 26(2):74 – 81, Mar 2007. 4, 32

[62] C. Shannon. A mathematical theory of communication. *ACM SIGMOBILE Mobile Computing and Communications Review*, 5(1):3–55, Jan 2001. 23

[63] G. Sharkawy, C. Newton, and S. Hartley. Attitudes and practices of families and health care personnel toward children with epilepsy in Kilifi, Kenya. *Epilepsy & Behavior*, 8(1):201–212, 2006. 2

[64] J. Shlens. A Tutorial on Principal Component Analysis. In *Systems Neurobiology Laboratory, Salk Institute for Biological Studies*, Dec 2005. 26

[65] B. Shneker and N. Fountain. Epilepsy. *Disease-a-Month*, 49(7):426–478, Jul 2003. 3, 4, 37

[66] A. Siegel, C. Grady, and A. Mirsky. Prediction of Spike-Wave Bursts in Absence Epilepsy by EEG Power-Spectrum Signals. *Epilepsia*, 23(1):47 – 60, Nov 1982. 31

[67] J. Sprott. *Chaos and Time-Series Analysis.* Oxford University Press, 2003. 25

[68] M. Stol. *Epilepsy in Babylonia.* Brill, Jan 1993. 2

[69] L. Sutin. *Divine Invasions: A Life of Philip K. Dick.* Harmony Books, 1989. 3

[70] J. Suykens, T. Gestel, J. Brabanter, B. Moor, and J. Vandewalle. *Least Squares Support Vector Machines.* World Scientific, Jan 2002. 30

[71] H. Teive, F. Germiniani, A. Cardoso, and L. Paola. The Uncinated Crisis of George Gershwin. *Arquivos de Neuro-Psiquiatria*, 60(2B):505–508, Jan 2002. 3

[72] V. Vapnik and A. Chervonenkis. On the uniform convergence of relative frequencies of events to their probabilities. *Theory of Probability and its Applications*, 16(2):264–280, 1971. 29

[73] S. Viglione and G. Walsh. Proceedings: Epileptic Seizure Prediction. *Electroencephalography and Clinical Neurophysiology*, 39(4):435–436, Oct 1975. 30

[74] P. Welch. The Use of Fast Fourier Transform for the Estimation of Power Spectra: A Method Based on Time Averaging Over Short, Modified Periodograms. *IEEE Transactions on Audio and Electroacoustics*, 15(2):70 – 73, Jan 1967. 24

[75] R. Wyatt. A Study of Power Spectra Analysis of Normal Finger Tremors. *IEEE Transactions on Biomedical Engineering*, BME-15(1):33 – 45, Jan 1968. 17

[76] D. Yu, L. Deng, J. Droppo, J. Wu, Y. Gong, and A. Acero. A minimum-mean-square-error noise reduction algorithm on Mel-frequency cepstra for robust speech recognition. In *IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2008*, pages 4041–4044, 2008. 17

[77] G. Zanchin. Considerations on "The Sacred Disease" by Hippocrates. *Journal of the History of the Neurosciences*, 1(2):91–5, Apr 1992. 1