Technological University Dublin
ARROW@TU Dublin

Dissertations

School of Computing

2018

A Performance Comparison of Neural Network and SVM Classifiers Using EEG Spectral Features to Predict Epileptic Seizures

Ian Thomas Tennant Watson
Technological University Dublin

Follow this and additional works at: https://arrow.tudublin.ie/scschcomdis

 Part of the Computer Sciences Commons

Recommended Citation

Tennant Watson, Ian Thomas (2018). *A performance comparison of neural network and SVM classifiers using EEG spectral features to predict epileptic seizures.* Masters dissertation, DIT, 2018.

This Dissertation is brought to you for free and open access by the School of Computing at ARROW@TU Dublin. It has been accepted for inclusion in Dissertations by an authorized administrator of ARROW@TU Dublin. For more information, please contact yvonne.desmond@tudublin.ie, arrow.admin@tudublin.ie, brian.widdis@tudublin.ie.

This work is licensed under a Creative Commons Attribution-Noncommercial-Share Alike 3.0 License

A Performance Comparison of Neural Network and SVM Classifiers Using EEG Spectral Features to Predict Epileptic Seizures

Ian Thomas Tennant Watson

A dissertation submitted in partial fulfilment of the requirements of
Dublin Institute of Technology for the degree of
M.Sc. in Computing (Data Analytics)

February 2018

Declaration

I certify that this dissertation which I now submit for examination for the award of MSc in Computing (Data Analytics), is entirely my own work and has not been taken from the work of others save and to the extent that such work has been cited and acknowledged within the text of my work.

This dissertation was prepared according to the regulations for postgraduate study of the Dublin Institute of Technology and has not been submitted in whole or part for an award in any other Institute or University.

The work reported on in this dissertation conforms to the principles and requirements of the Institute's guidelines for ethics in research.

Name of Candidate: _____

Signature of Candidate: _____

Date: _____

Abstract

Epilepsy is one of the most common neurological disorders, and afflicts approximately 70 million people globally. 30-40% of patients have refractory epilepsy, where seizures cannot be controlled by anti-epileptic medication, and surgery is neither appropriate, nor available. The unpredictable nature of epileptic seizures is the primary cause of mortality among patients, and leads to significant psychosocial disability. If seizures could be predicted in advance, automatic seizure warning systems could transform the lives of millions of people.

This study presents a performance comparison of artificial neural network and support vector machine classifiers, using EEG spectral features to predict the onset of epileptic seizures. In addition, the study also examines the influence of EEG window size, feature selection, and data sampling on classification performance. A total of 216 generalised models were trained and tested on a public seizure database, which contained over 1300 hours of EEG data from 7 subjects.

The results showed that ANN outperform SVM, when using spectral features ($p = 0.035$). The beta and gamma frequency bands were shown to be the best predictors of seizure onset. No significant differences in performance were determined for the different window sizes, or for the feature selection methods. The data sampling method significantly influenced the performance ($p < 0.001$), and highlighted the importance of treating class imbalance in EEG datasets.

Keywords: Artificial Neural Network, Support Vector Machine, Supervised Machine Learning, Epilepsy, Seizure Prediction, Spectral, ANN, SVM, EEG

Acknowledgements

I would like to express my gratitude to the lecturers and my fellow students at DIT for their professionalism and support over the course of my studies. In particular, my special thanks to Brendan 'Is it Finished?' Tierney, for his pragmatic advice on getting this dissertation over the line.

None of this would have been possible without the love and support of all the women in my life. Firstly, my long-suffering wife Jane, who tolerated the deadlines, missed weekends and single-parenting without complaint. You have always believed in me, through all the years we have shared. To my three angels, Kate, Emily and Rose, you were put on this world to make me a better person. Nothing will ever make up for the lost bedtime stories and family time, but I hope my dedication will serve as an example to you later in your lives. Lastly, my sister Jenny, who showed me that it was possible to overcome anything. You may be far away, but you are always in my thoughts.

Contents

List of Figures

List of Tables

Acronyms

AED Anti Epileptic Drug. 1

ANN Artificial Neural Network. 4

CSVM Cost-Sensitive Support Vector Machine. 41

DFT Discrete Fourier Transform. 28

DWT Discrete Wavelet Transform. 28

EEG Electroencephalograph. 3

FFT Fast Fourier Transform. 28

FPR False Prediction Rate. 46

HFO High Frequency Oscillations. 9

iEEG Intracranial Electroencephalograph. 15

KNN K Nearest Neighbours. 34

LVQ Learning Vector Quantisation. 58

mRMR Minimum Redundancy Maximum Relevance. 34

PCA Principle Component Analysis. 34

RBF Radial Basis Function. 40

RFE Recursive Feature Elimination. 58

sEEG Scalp Electroencephalograph. 15

SMOTE Synthetic Minority Over-sampling Technique. 36

SVM Support Vector Machine. 4

Glossary

Artefact False signal in data. 53

Focal epilepsy Seizures localised in a specific brain region. 13

Generalised epilepsy Seizures across both lobes of the brain. 13

Ictal Physiological state indicating seizure. 10

Interictal Physiological state between seizure events. 10

Morbidity Condition of being diseased. 2

Postictal Physiological state following seizure. 10

Preictal Physiological state preceding seizure. 9

Refractory epilepsy Uncontrolled drug-resistant seizures. 1

Status epilepticus Seizure duration of greater than 5 minutes. 2

Sulcus Depression or groove in the cerebral cortex. 16

Chapter 1

Introduction

1.1 Background

Epilepsy is a neurological disorder that affects approximately 70 million people worldwide. It is estimated to affect 10 in every 1000 persons, a rate of approximately 1%. The prevalence is slightly higher in males and in early childhood and adolescence (Shakirullah, Niaz, Khan, & Nabi, 2014). The risk increases after middle age and about 3% of people are diagnosed with epilepsy by age 60 (Yu et al., 2016).

Epilepsy is characterised by recurring debilitating episodes of dysfunctional brain activity (Kharbouch, Shoeb, Guttag, & Cash, 2011) that manifest as epileptic seizures. Conventional treatments for epilepsy are based on anti-epileptic drugs (AEDs), with common side-effects including sedation, dizziness, fatigue and depression. Unfortunately the medication is ineffective for 30-35% of patients with refractory epilepsy (Carney, Myers, & Geyer, 2011) and is unable to control seizures; for these patients, brain surgery is the only feasible alternative and is used to resect the region of the brain responsible for seizure generation.

However, surgery is frequently neither possible nor desirable, and always involves high risk to the patient. In a survey of patients with refactory epilepsy, Arthurs, Zaveri, Frei, and Osorio (2010) found that the neurosurgical procedures and the concomitant surgical complications made implantable cortical devices unattractive to a majority of respondents. As a consequence, about 30% of patients have no viable treatment options and must live with the continuous threat of unpredictable seizures, which can occur at any time or place. Epileptic seizure prediction is an ongoing field of research that is primarily targeted at this group of patients, and if successful, could greatly

improve their quality of life.

Seizures generally occur without warning, like a "bolt from the blue" (Mormann, Andrzejak, Elger, & Lehnertz, 2007) and are currently clinically unpredictable. Mortality and morbidity associated with epilepsy is largely related to injuries from falls after loss of conciousness and the risk of status epilepticus. The inherent unpredictability of seizures leads to significant psychosocial disability and continuing anxiety in patients. Quality of life for patients would be significantly improved if seizures could be predicted in advance, thus alerting the patient and their caregivers.

The goal of epileptic seizure research is a reliable and accurate algorithm that will underpin a therapeutic device to provide a warning alarm or medical intervention shortly before seizure onset. If seizures can be reliably predicted, the patient can be treated with fast-acting AEDs or with other treatment options, or modify their activities to reduce the risk of injury to themselves and others.

New therapeutic solutions like closed-loop electrical stimulation and automatic anti-seizure drug administration would be feasible if systems with sufficient predictive performance can be developed (Feldwisch-Drentrup et al., 2011). On-demand medication and treatments delivered prior to seizure would also help to reduce the debilitating side effects caused by regular use of AEDs (Carney et al., 2011). The development of accurate and reliable seizure forecasting could "transform epilepsy care" (Brinkmann et al., 2016, p.2), and the lives of millions of patients. However, this goal remains a challenging proposition and the subject of ongoing research.

1.2 Research Problem

In their influential review of seizure prediction, Mormann et al. (2007) noted that earlier studies with very optimistic results could not be reproduced in follow-up research with larger sets of data. According to Andrzejak, Chicharro, Elger, and Mormann (2009), initial studies using linear and non-linear methods showed promising results and claimed to reliably predict seizures. However, these findings were never substan-

tiated or replicated in subsequent research. It is widely recognised that there is a reproducibility problem with many claims in biomedical research (Brinkmann et al., 2016), which delays the development of therapies.

Early prediction studies did not use rigorous statistical testing and validation techniques. In order to improve meta-analysis, Mormann et al. (2007) and Snyder, Echauz, Grimes, and Litt (2008) made recommendations to standardise performance evaluation and statistical validation methods. However, these guidelines are not consistently followed and a diversity of techniques and performance measures is found in the literature. This negatively impacts seizure prediction research by making it difficult to compare the performance of algorithms from different studies.

Mormann et al. (2007) pointed out the lack of publicly available EEG datasets that contain continuous long-term seizure recordings. Short-duration studies seldom provide adequate numbers of seizure events, and which are clearly separated by interictal periods (Brinkmann et al., 2016). Additionally, Carney et al. (2011) stated that there was no class 1 evidence of clinical usefulness within the published research, and that this stemmed in part from the poor availability of long-term EEG recordings, and the failure to statistically validate the results. The predictive performance of initial studies was too low for clinical applications, which require classification error rates of less than 5% (Harati et al., 2014). Methods should also deliver results in real-time, before they can be included in embedded systems (Chisci et al., 2010).

Seizure prediction is complicated most notably by variations in electroencephalograph (EEG) signals across individuals. EEG patterns associated with seizures in one patient may resemble non-seizure patterns in another patients' EEG, making a generalisable approach difficult to achieve. Kharbouch et al. (2011, p. S29) argue that the principle challenge in creating a generalised predictor is the "heterogeneity of seizure EEG patterns across different patients and even within a patient".

Indeed, there may be no single algorithm can be applied generically to all patients' data that is capable of achieving adequate performance (A. H. Shoeb, 2009). Similarly, according to Feldwisch-Drentrup et al. (2010), EEG signals are highly complex and no

single measure has been sufficiently predictive. Hence, Park, Luo, Parhi, and Netoff (2011) hypothesise that the solution to the poor performance is a patient-specific algorithm based on multiple EEG features. Despite claims to the contrary, seizure prediction with high levels of sensitivity and specificity has remained elusive (Park et al., 2011). In conclusion, researchers remain far from a reliable approach for seizure prediction that can realistically be used in a practical clinical setting (Bandarabadi, Teixeira, Rasekhi, & Dourado, 2015). Further work is required to assess the relative predictive power of features and the capabilities of the machine learning algorithms (Brinkmann et al., 2016).

1.3 Research Aim and Objectives

The support vector machine (SVM) is a popular supervised learning model, and is frequently used in seizure prediction research. Although artificial neural networks (ANNs) are used less often, they are highly suitable due to their inherent parallelism and their ability to deal with multi-dimensional, non-linear inputs. Studies using SVM and ANN have reported very good results, and the two classifiers show the most potential for developing high performance seizure prediction models.

The proposed research hypothesises that ANN are superior to SVM for predicting the onset of an epileptic seizure. The study will construct supervised machine learning models and use them to make predictions on test data from long-term EEG recordings. The performance of the classifiers is quantified and compared to determine whether the ANN delivers superior results, thus proving or disproving the hypothesis.

Primary Research Question 1:

"Does an artificial neural network classifier provide superior performance compared to a support vector machine classifier, when using EEG spectral features to predict the onset of epileptic seizures?"

From this, the null hypothesis (H_0) may be expressed as:

> "An artificial neural network does not demonstrate superior classification performance compared to a support vector machine classifier, when using EEG spectral features to predict the onset of epileptic seizures".

Conversely, the alternative hypothesis (H_A) is stated as:

> "An artificial neural network demonstrates superior classification performance compared to a support vector machine classifier, when using EEG spectral features to predict the onset of epileptic seizures".

The study has four secondary objectives, which aim to evaluate the influence of four different EEG preprocessing methods on the performance of the classifiers.

Secondary Research Question 1:

The classifiers use a set of spectral features, which are well-known for their predictive capabilities. A second set of statistics-based features, and a third set combining the spectral and statistics sets are also used for training. The performance of the classifiers will be assessed to determine the most predictive set of features.

Secondary Research Question 2:

All prediction studies segment the EEG signal prior to feature extraction, but it is unclear how EEG segment size affects overall classification performance. Four different EEG windows will be assessed for their impact on performance, to determine the optimal segment size.

Secondary Research Question 3:

Once EEG features have been extracted, the study examines the effects of two feature selection methods. These reduce the feature count to a more practical level, but can also degrade the overall classification accuracy. The impact of feature selection on classification performance is evaluated.

Secondary Research Question 4:

EEG datasets generally have relatively few seizure events, and imbalanced class dis-

tributions can severely reduce classification performance. The study investigates the performance impact of upsampling the minority class, and downsampling the majority class. Their performance is compared to that obtained from using imbalanced data.

Seizure prediction is "very difficult" (Stacey, Le Van Quyen, Mormann, & Schulze-Bonhage, 2011), with many confounding variables including patients' medication, physiological state, seizure heterogeneity and the random nature of seizures. Other complicating factors include differences in data preprocessing, feature selection, choice of classifier, statistical validation and evaluation measures. The proposed research will hopefully provide fresh evidence for the importance of ANN, and the continued efficacy of EEG spectral features. It will also evaluate the influence of several preprocessing techniques on the overall classification performance, thus adding to the limited literature on this subject.

1.4 Research Methodology

A critical review of the literature is conducted to determine the most up-to-date issues, methods, features and algorithms used in epileptic seizure prediction research. Secondary data from a public EEG database is used as a labelled dataset for the supervised machine learning algorithms. This research is an empirical study, that experimentally develops multiple classification models, and quantitatively assesses their performance against a set of test data. The quantitative results are tested for significance, and the outcome is used to confirm or reject the research hypothesis. Where appropriate, the research methodology broadly follows the well-known CRISP-DM model, which is the leading data mining process model used in industry. The analysis framework includes business understanding, data understanding, data preparation, modelling and evaluation.

1.5 Scope and Limitations

Moghim and Corne (2014) draw a distinction between analysis-oriented studies and prediction-oriented studies. The former focus on analysing EEG retrospectively to understand and quantify the statistical metrics of EEG that correlate with seizure and non-seizure states. The latter focus on predictive algorithms to develop a viable seizure prediction system, and typically draw from the analysis-oriented studies to help choose the best set of discriminative features.

Accordingly, this is a prediction-oriented study that will develop multiple supervised machine learning models, using a variety of preprocessing techniques, features and classifiers. The data is drawn from a repository containing long-term continuous intracranial EEG recordings from 5 canine and 2 human subjects. The study is restricted to a standard SVM implementation and a feedforward neural network with a single hidden layer. Classification models are not optimised individually. It is worth noting that this study does not claim that spectral and statistics features are optimal for predictive performance; the main goal of the research is to compare classification performance, and not to identify the most predictive EEG features.

1.6 Dissertation Outline

The rest of the paper is structured as follows: Chapter 2 provides a critical overview of the literature and provides necessary background information on epilepsy, seizures and EEGs. It also assesses current research on EEG preprocessing, features, classifiers and evaluation methods. Chapter 3 discusses the methodological approach, with reference to techniques from the literature. Chapter 4 includes the implementation and results, and Chapter 5 discusses and critically assesses the findings. Chapter 6 concludes the paper by summarising the main points of the study, and gives some thoughts on future research directions. The full set of results are contained in Appendix A.

Chapter 2

Literature Review and Related Work

This chapter starts by introducing some important background information on epilepsy, and includes sections on seizure definitions, terminology for describing brain states, and the role of the EEG and associated datasets in research. The history of seizure detection and prediction is discussed, and some of the current challenges are highlighted. The sections thereafter describe EEG preprocessing and EEG features, and the classifiers at the heart of this research. The chapter concludes by reviewing evaluation measures and discusses the importance of statistical testing in prediction studies.

2.1 What is Epilepsy?

2.1.1 Seizure Definitions

Epilepsy "is a state of the brain, characterised by recurrent epileptic events that occur as a result of chronic structural or functional changes in the brain" (Mormann & Elger, 2013, p.S31). It can originate from trauma, inflammation or vascular events and the most frequent clinical manifestations are seizures that are characterised by sudden changes in a patient's behaviour or perceptions, or both. Seizures can affect different parts of the brain and symptoms can be subtle and unrecognisable as epileptic events. Indeed, if seizures occur in areas responsible for memory or conciousness, patients may not even be aware of the event.

The original definition of an epileptic seizure is "that of a behavioural event perceived by a patient or noted by an observer" (Gotman, 2011, p.S2). These behavioural changes can manifest as blushing, muscular twitches or short durations of inattention, tingling or strange feelings. Seizures also often have an amnestic aspect, leading to

a disparity between self-reported events and clinically observed events (Cook et al., 2013). Early electroencephalographers used these symptoms to establish the link between behavioural seizures and EEG discharges, thus leading to the common definition of a seizure as a "temporary dysfunction of the brain consisting of an excessive neuronal discharge" (Gotman, 2011, p.S2). The author further defines an *EEG seizure*, as an event recorded on an EEG, independent of the seizure's behavioural component. An EEG seizure covers a continuum and includes large, small, minimal and no seizure at all. In this context, a large seizure lasts 1 minute, and involves all channels with large-amplitude activity; a small seizure lasts 20s, with a few channels and moderate-amplitude changes; and a minimal seizure lasts 8s and involves a pair of electrodes with minimal amplitude changes.

As Gotman (2011, p.S2) concludes, "some seizures are totally unambiguous, some are very uncertain and there is continuity between the two extremes". Jouny, Franaszczuk, and Bergey (2011) agree that the precise definition of a seizure remains difficult to unequivocally establish. They further state that this uncertainty has impaired meta-analysis of competing algorithms, since the studies do not use a standard definition of a seizure, thus making it impossible to compare performance between studies. In addition, clinical seizures are a small proportion of abnormal electrical activity in the brain (Karoly et al., 2016) and the presence of other waveforms caused by subclinical seizures, interictal spikes and high frequency oscillations (HFO) have confounded attempts to develop a precise electroencephalographical definition of a seizure.

2.1.2 Brain States

It is common practices in seizure prediction studies to define four cerebral states (Moghim & Corne, 2014), shown in Fig 2.1:

1. Preictal - The state immediately before seizure. It is believed that this state contains predictive markers of the pending seizure. The state can be minutes to hours in length.

2. Ictal - The seizure state, indicated by seizure activity in the brain. It covers the period between seizure *onset* and *offset* and can last several minutes.

3. Postictal - Brain activity following the seizure offset, or after the seizure. The transition period after a seizure can show abnormal excitement in the EEG.

4. Interictal - The neurological baseline or 'normal' state between seizures. Marks the period proceeding the postictal state and preceding the preictal state.

Bandarabadi, Rasekhi, Teixeira, Karami, and Dourado (2015) maintain that the duration of the preictal period is unclear and varies from seizure to seizure. The preictal period may vary considerably from hours to seconds before the seizure. In some cases, the epileptogenic changes develop late (close to seizure onset) and in others the changes develop earlier (tens of minutes in advance of seizure). Neurologists do not generally define a preictal or postictal period (Alexandre Teixeira et al., 2014) and instead these are framed within the context of the individual study.

Rasekhi, Mollaei, Bandarabadi, Teixeira, and Dourado (2013) examined the effects of different EEG preprocessing methods and found that a preictal period of 40 minutes resulted in the best performance. This was corroborated by Bandarabadi, Rasekhi, et al. (2015), who analysed performance using preictal periods ranging from 5 to 180 minutes and found that the optimal period varied significantly from one seizure to another, with a mean period of approximately 44.3 minutes. Alexandre Teixeira et al. (2014) tested 10, 20, 30 and 40 minute preictal periods, and an average of 30.47 minutes gave the best results.

2.1.3 Evidence for the Preictal State

Seizure prediction systems are predicated on the existence of a preictal brain state. This state is associated with an increased probability of seizure and provides a physiological signal - or *biomarker* - that distinguishes the interictal and preictal states. For many years, researchers have sought evidence for the preictal state and tried to determine whether EEG variations can be identified prior to seizure (Gadhoumi, Lina,

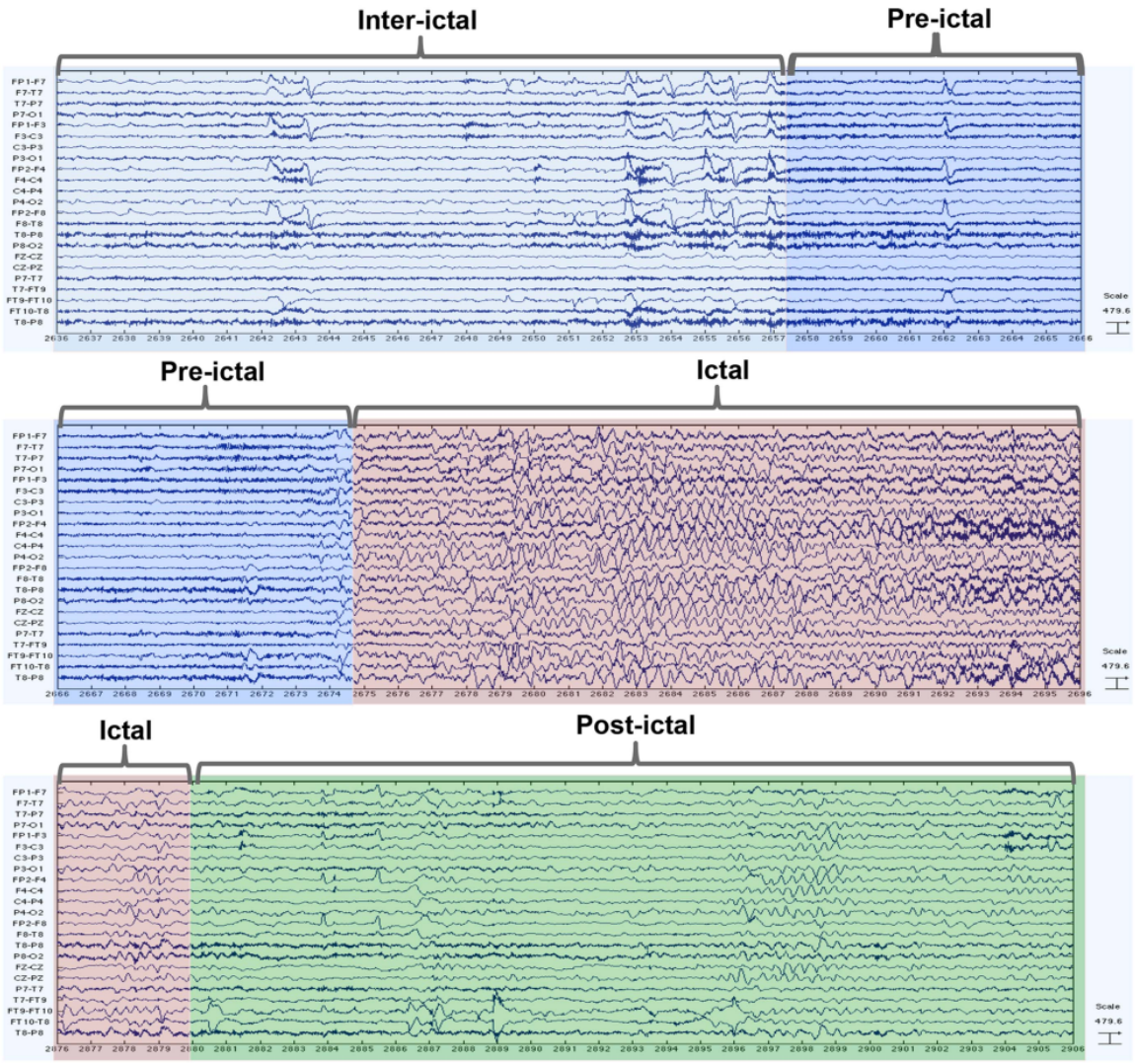

Figure 2.1. Invasive EEG with annotated brain states (Moghim & Corne, 2014)

Mormann, & Gotman, 2016).

Worrell et al. (2004) found that high-frequency gamma band oscillations (60-100 Hz) were present in the 20 minutes prior to neocortical seizures in 62% of patients in their study. The authors claimed that this offered opportunities for identifying periods with a higher probability of seizure onset. Similarly, Stacey et al. (2011, p.246) hypothesise that HFO are a "promising candidate for a seizure precursor", and may also play a role in seizure initiation. Oscillations are associated with partial seizures in patients with temporal epilepsy, with an increase in the gamma bands reported minutes prior to a

seizure. However, research into HFO has been slow, due to the inability of most EEG equipment to record at a sufficiently high sampling rate (Stacey et al., 2011).

Studies show that seizures tend to cluster in time and have underlying periodicities (Howbert et al., 2014), which suggests a fixed network responsible for seizure generation and the potential for seizure forecasting. Clinicians generally assumed that the shift from non-seizure to seizure state was an abrupt phenomenon, but Badawy, Macdonell, Jackson, and Berkovic (2009) showed that neurological effects are observed in patients up to 24 hours in advance of a seizure. Evidence has emerged that cortical hyperexcitability is a precursor to seizure onset and that seizures arise from an identifiable brain state (Cook et al., 2013). Patients also self-reported changes in conciousness, and increases in cerebral bloodflow, oxygenation and cortical excitability have been recorded leading up to a seizure (Brinkmann et al., 2016).

Evidence for the existance of a preictal state is a subject of ongoing research efforts and debate. Identification of the preictal state would also contribute to an understanding of the mechanisms that underpin individual seizures and epilepsy (Carney et al., 2011). Quantifying the preictal state in EEG signals provides a valuable tool for epileptic seizure prediction (Bandarabadi, Teixeira, et al., 2015).

2.1.4 Seizure Genesis

Epileptic seizure prediction is based on the observation of brain dynamics in EEG, which are used to classify time-series brain electrical activity into two classes - the non-seizure or interictal state and the pre-seizure or preictal state. The process of moving to a seizure state is generally not well understood (Mormann et al., 2007) and it is conceivable that different mechanisms are responsible for seizures occurring in different parts of the brain. Dynamical changes in an EEG prior to seizure and the methods for detecting them may therefore vary considerably from one patient to another. Stacey et al. (2011) concur that there is a limited understanding of the activities that generate seizures, which confounds the development of reliable seizure

prediction algorithms.

Da Silva et al. (2003) hypothesise that seizures can evolve in two different ways and use a mathematical approach to develop a non-linear model describing the brain's transition from normal to seizure state. In certain brains, the distance between 'normal steady-state' and epileptic state is very small and minor random fluctuations can trigger the transition to seizure - such seizures are not practical to predict and are normally associated with generalised epilepsy (Mormann et al., 2007). However, in other cases the distance between states is larger and critical neuronal changes occur over time, leading to a gradual transition to a seizure state. The seizures manifesting gradual transitions in state are more likely associated with focal epilepsy and are theoretically predictable if the brain dynamics can be recorded and interpreted within a window of opportunity.

Characteristics of EEGs that signal the onset of a seizure can be very different from those that signal its end (A. Shoeb, Kharbouch, Soegaard, Schachter, & Guttag, 2011). Rhythmic activity at the start of the seizure typically has a fundamental frequency within the alpha, beta, theta or delta band. Postictal EEG recordings often exhibit delta wave slowing, amplitude attenuation or a combination of these two. Prediction, detection or termination of seizures is complicated by the gradual nature with which the EEG changes at the beginning and end of seizures. It may also be masked by muscle or movement artefacts that cause unwanted signals in the EEG.

Alexandre Teixeira et al. (2014, p.326) define seizure onset as the "first electrographic sign of oncoming seizure". Prior to seizure, the brain undergoes a gradual transition from chaoticity to rythmicity (Mormann et al., 2007). When a focal seizure initiates, sychronised epileptic activity is observed in a small localised area of the cortex (Niknazar et al., 2013), and from this initial focus the seizure spreads to other brain regions. This seizure focus is the basis of resective surgery, which can cure epilepsy if the brain tissue generating seizures can be identified and removed.

2.2 The EEG

The EEG is increasingly being used in preventive diagnostic medicine. According to Harati et al. (2014), the global economic burden due to brain-related illnesses exceeded more that $2T per year in 2014. The majority of research is focused on epilepsy diagnosis and stroke but is also expanding the application of EEG to other neurological conditions such as Alzheimer's disease and sleep disorders. EEG are also commonly used to monitor blood flow during surgery, diagnose post traumatic stress disorder (PTSD) and chronic traumatic encephalopathy (CTE) in athletes (Harati et al., 2014).

The EEG reflects the electrical activity of the cerebral cortex, which is recorded using electrodes uniformly distributed on the scalp, or on the surface of the brain. The names and locations of electrodes are specified by the International 10-20 system, which uses anatomical landmarks for placement intervals. Odd numbered electrodes are placed on the left side of the head, and even numbered electrodes are placed on the right.

An EEG recording may be visualised as a chart showing electrical energy on the y-axis and time on the x-axis. An EEG channel is formed by measuring the difference in electrical potential between two electrodes. The potential is the sum of the electrical activity of millions of neurons in the cortex. Seizures are caused by abnormal or excessive synchronous neuronal activity in the brain. Since seizures are related to the electrical activity in the brain, the EEG recording is a useful biosignal for seizure detection and prediction (Niknazar et al., 2013). Following seizure onset, EEG channels exhibit rythmic activity that is composed of multiple spectral components (A. H. Shoeb & Guttag, 2010), with variation between individuals with respect to the channels involved and the nature of the rhythmic activity.

Seizure detection and prediction relies primarily on EEG analysis (Gotman, 2011). Although other behavioural aspects and physiological systems have been investigated, the most appropriate source of information for forecasting epileptic seizure is the scalp

electroencephalograph (sEEG) or intracranial electroencephalograph (iEEG) (Jouny et al., 2011). As the name implies, sEEG use electrodes placed on the scalp whereas iEEG use electrodes placed on the surface of the brain via a surgical procedure. The detection of a seizure event depends on the ability of the EEG to measure neuronal activity in the part of the brain where the epileptic discharge occurs. The discharge may be clearly visible in a scalp EEG; only visible in an invasive iEEG, or may not be visible at all if the electrodes are too far from the seizure locus (Gotman, 2011).

iEEG recordings are considered superior to scalp recordings and have higher signal-to-noise ratios, better spatial resolution and are mostly free of artefacts (Mormann et al., 2007). Similarly, Sudalaimani et al. (2015) concur that intracranial EEG data is less prone to noise compared to scalp measurements. However, there is no general agreement on which type of EEG leads to better predictive performance.

Alexandre Teixeira et al. (2014) compared results from iEEG and sEEG recordings and found no substantial differences in sensitivity. Conversely, both Rasekhi et al. (2013) and Bandarabadi, Teixeira, et al. (2015) achieved marginally better sensitivities for sEEG data, compared to iEEG. The latter study takes the view that sEEG presents a generalised spatial view of neuronal activity compared to iEEG, which provide a more restricted localised view. The authors hypothesise that the broad spatial perspective of sEEG may capture different dynamics across a wider region of the brain, leading to better predictive performance.

EEG signal sampling frequency has increased in conjunction with the development of equipment with greater computational power. These higher sampling rates may be used to characterise the higher frequency components in EEG to determine their role in seizure formation. Additionally, Alexandre Teixeira et al. (2014) found a positive correlation between higher sampling rates and predictive performance. This is due to improved signal quality and the ability to extract high-frequency components from the signal. However, sample rates above 32kHz bring their own challenges in terms of data storage and processing power (Jouny et al., 2011). As an example, 100 hours of 24-channel EEG data sampled at 32 kHz represents over 500 GB of data.

Osorio and Schachter (2011) caution that the focus on EEG data and the brain-epilepsy link has distracted researchers from the connection between brain and body. They further state that EEG and iEEG are severely limited due to the inherent cortical signal attenuation and that access to neural sources is poor - with scalp EEG, only a third of the neocortex can be surveyed, and in the case of intracranial electrodes very little activity can be recorded from the lateral and bottom walls of sulci. The authors propose extracerebral signals as an alternative to EEG, in particular cardiac and motor signals. However, the use of these signals in prediction studies is rare and the EEG recording remains the principle source of data for epilepsy research.

2.3 EEG Datasets

The poor availability of EEG data for testing is a commonly cited issue in this field of research. This section gives a brief overview of several well-known publicly available EEG seizure databases, which are frequently used in seizure prediction studies.

Seizure prediction research is limited by the lack of publicly available EEG datasets that contain continuous long-term seizure recordings (Brinkmann et al., 2016). Research based on private datasets containing data from small numbers of patients and seizures cannot be statistically validated, reproduced or compared to other studies. Most clinical records do not provide adequate numbers of seizures separated by clear, seizure-free periods for statitistical testing. Seizure events are relatively rare in most patients and months of continuous recording may be be required to characterise their individual seizure frequency. However, long-term scalp measurements are impractical due to the difficulty of maintaining leads and may be inaccurate because scalp recordings are insensitive to deep or focal seizures (Smith, 2005). Consequently, there is an urgent need for high-quality long-term continuous intracranial EEG recordings. These are a necessary prerequisite for performance analysis and will enable future seizure forecasting efforts (Schulze-Bonhage, Feldwisch-Drentrup, & Ihle, 2011).

An EEG dataset is a collection, database or repository of digital files containing anno-

tated EEG seizure data, whose labels are agreed upon by a panel of expert clinicians. Biosignals from EEG are 'highly subjective' (Asha, Sudalaimani, Devanand, Thomas, & Sudhamony, 2013) and the diversity of seizure patterns can mean that even experts sometimes cannot reach agreement on specific seizure events (Jouny et al., 2011). Under this system, seizures are identified and labelled according to the majority consensus among the reviewers.

FSPEEG & EPILEPSIAE Databases

Since its inception in 2004, the Freiburg Seizure Prediction EEG (FSPEEG) database is the most cited resource in seizure detection and prediction research, and is still used by many current publications (Moghim & Corne, 2014; Gadhoumi et al., 2016). The database was superseded by the EPILEPSIAE database in 2012, which is comprehensively reviewed by Schulze-Bonhage et al. (2011) and Klatt et al. (2012).

The EPILEPSIAE database contains continuous long-term recordings of 275 patients, and is comprised of 225 scalp EEG and 50 iEEG recordings. It claims to be the largest international epilepsy database for seizure prediction. Although funded by the European Union, the dataset is only commercially available. At least 5 days of continuous EEG is available for each patient, with a minimum of 24 channels and sample rates of up to 1024 Hz. The database is used by Feldwisch-Drentrup et al. (2011); Alexandre Teixeira et al. (2014); Moghim and Corne (2014); Bandarabadi, Teixeira, et al. (2015) and others.

TUH EEG Corpus

The Temple University Hospital (TUH) EEG Corpus is a repository of EEG recordings freely available to the research community and is described in depth by Harati et al. (2014). It consists of more than 25,000 EEG recordings from 14,000 patients and is collected from the hospital's Department of Neurology. The EEG records have between 20 and 128 channels and are sampled at 250 Hz minimum. If each channel is considered

independently, there is over 1 B seconds of time-series data. Each record contains metadata with the patient's anonymised information and is accompanied by a board-certified neurologist's report that includes medical history and medications.

CHB-MIT Scalp Database

The CHB-MIT Scalp EEG database contains records of 198 seizures recorded from 24 paediatric patients and is described by A. H. Shoeb (2009). The database is also commonly referred to as the Physio Net epilepsy database and is freely available to researchers. The EEG recordings were sampled at 256 Hz with 16-bit resolution and with 23 electrodes in most cases. The start and end of each seizure was annotated by a trained electroencephalographer, thus making the dataset suitable for supervised learning techniques. The CHB-MIT dataset is used by A. H. Shoeb and Guttag (2010), A. Shoeb et al. (2011), Sharma (2015) and Fergus et al. (2016), among others.

MSEL Epilepsy Database

The Mayo Clinic and the Mayo Systems Electrophysiology Lab (MSEL) provide a public database of iEEG recordings, comprising data taken from seven subjects. The data consists of ambulatory iEEG recordings taken from five canine subjects and presurgical iEEG recordings from two human patients. All data clips were reviewed and labelled by a board certified expert.

The suitability of using canine epilepsy as a disease model has been investigated for two decades. Berendt and Gram (1999) applied definitions of human seizures to sixty-three naturally epileptic dogs and reported that the prevalence, age of initial onset and presentation of epilepsy in humans and dogs is similar. In their research on an implantable device tested in dogs, Davis et al. (2011, p.121) noted that the test animals' "electrographic presentation appeared indistinguishable from human focal onset neocortical seizures". They further conclude that their algorithm developed for human application had comparable results on canine data, thus "suggesting quantitative sim-

ilarity between canine and human ictal patterns". Similar studies by Howbert et al. (2014) and Brinkmann et al. (2015) have reported good performance and have further validated the use of canine epileptic EEG data in research.

The MSEL data was originally used in research by Brinkmann et al. (2015) and was subsequently made available for a crowdsourcing seizure prediction competition in 2014, sponsored by the Mayo Clinic and hosted by Kaggle.com. The prediction competition is described by Brinkmann et al. (2016) and the dataset is used in research by Temko, Sarkar, and Lightbody (2015), Karumuri, Vlachos, Liu, Adkinson, and Iasemidis (2016) and Sudalaimani et al. (2017).

2.4 Origins of Epileptic Seizure Prediction

2.4.1 Seizure Detection

The feasibility of epileptic seizure prediction has been explored for over 35 years. Many of the initial studies focused on seizure *detection* rather than *prediction*, and examined which measures could be used to detect the onset of a seizure. Whereas seizure detection aims to identify seizures shortly before or after onset, seizure prediction attempts to recognise the seizure minutes in advance.

Seizure detection systems are an attractive alternative for seizure classification and epilepsy diagnosis. Conventionally, an expert must manually analyse the entire length of an EEG to determine epileptic events; it can take as long as several weeks for a review of an EEG by a certified neurologist (Harati et al., 2014), depending on resources. Automated detection algorithms reduce the requirement for highly trained clinicians and can speed up diagnosis and treatment. Kharbouch et al. (2011) developed a real-time patient-specific seizure onset detection algorithm, using temporal and spectral features. Similarly, Fergus et al. (2016) report on a generalised seizure detection system using spectral and statistical features and finally, Kumar, Sriraam, and Benakop (2008) present a seizure detection system using wavelet entropy with a

neural network classifier.

Related to detection, A. Shoeb et al. (2011) claim to have developed the first machine-learning algorithm for determining seizure *termination* from electrographic activity. In conjunction with a seizure-onset detection system, it may be used to determine seizure duration to aid in diagnosing status epilepticus, or to deliver pharmaceutical therapies to offset post-seizure symptoms.

Compared to seizure prediction systems, seizure detection systems are relatively advanced and closer to prospective clinical applications. The performance of seizure detection algorithms is typically higher, since detecting an imminent seizure is easier than predicting one minutes ahead of onset (Moghim & Corne, 2014). Eight years ago, Chisci et al. (2010) declared that the seizure-detection problem is 'practically solved'. In a recent generalised study, Fergus et al. (2016) report sensitivity and specificity of 88%, using a KNN classifier. However, an effective and reliable solution for seizure forecasting requires further research efforts.

2.4.2 Seizure Prediction

There is considerable overlap in the research methodologies between seizure detection and seizure prediction. As research shifted to seizure *prediction*, many of the detection measures were used to try and produce evidence of a preictal state, an endeavour which has 'remained elusive' (Jouny et al., 2011). The goal of epileptic seizure prediction can be simply stated as the *detection of the preictal state* (Rasekhi et al., 2013).

In the 1980's, the development of theories of non-linear systems led to new techniques that promised better results than the linear methods used to-date. Initial optimistic results using combinations of linear and non-linear methods were regarded as a proof-of-concept of the preictal state (Gadhoumi et al., 2016). However, none of the results of these early prediction studies could be replicated and claims about the superiority of non-linear measures were questioned. Initial successes were generally hard to replicate due to overtraining, and no measure was able to perform with high levels of specificity

and sensitivity (Carney et al., 2011).

Mormann et al. (2007) presented a seminal chronological review of seizure prediction since the 1970s. Their work highlighted shortfalls in methodology and poor statistical significance in many papers to-date. In particular, they pointed to inadequacies in sampling the preictal period. The study suggests guidelines for dealing with these important issues, and proposes that performance should be baselined against a random predictor. They maintain that the fundamental question is whether characteristic features extracted from an EEG are predictive of an impending seizure. They provide a useful summary of studies, their prediction measures and reported results, and describe commonly used linear and non-linear measures.

Similarly, Gadhoumi et al. (2016) have published an updated review of research from 2004 to 2014 that includes studies based on their methodological validity and statistical significance. Their paper is restricted to research using intracranial EEG recordings, which they argue are more appropriate for therapeutic devices. The review also includes a table of the studies and the features, prediction methods, EEG databases and validation methods that each uses. Ramgopal et al. (2014) conducted a review of therapeutic devices, but the focus is predominantly on seizure detection and not prediction.

Recent studies that focus on building predictive models with optimal predictive performance have been obtaining very good results. In two separate patient-specific studies, both of which used the FSPEEG dataset, Moghim and Corne (2014) and Park et al. (2011) report an average sensitivity of 91.14% and 97.5% respectively.

2.5 EEG Preprocessing

This section covers the various techniques used for preparing the EEG data. The preprocessing of an EEG recording in preparation for modelling is a non-trivial matter. Firstly, the time-series data must be extracted from the individual EEG channels. If appropriate, data from specific electrodes are selected and the data is segmented into

windows for individual analysis. Each window is then processed to extract features that characterise the underlying signal. The features are bound into a feature matrix, and the values are generally normalised. Feature selection is then used to determine the most predictive variables, and the feature matrix is reduced accordingly. Finally, the data is partitioned into testing and training sets, and data sampling methods are applied to the data to balance the class distributions in the training data.

2.5.1 Electrode Selection

Brinkmann et al. (2015) maintain that no data exists to address the selection and placement of electrodes for seizure prediction. However, a limited number of studies have investigated the impact of preselecting the electrodes, and generally select the electrodes closest to the seizure focus.

The research by Kharbouch et al. (2011) uses the full set of electrodes in the data. The authors state that it is unclear what the effect of using a subset of electrodes may have on classification performance. On the one hand, the use of all electrodes may include information from other brain regions not involved in seizure onset that is not readily apparent to a clinician. The potential downside is that the inclusion of all channels may incorporate features with little discriminative power and thus reduce accuracy. Chisci et al. (2010), Rasekhi et al. (2013) and Bandarabadi, Teixeira, et al. (2015) all used six electrodes in total - three placed close to the focal area and three far from the focal area. The latter argue that this is an adequate trade-off between the information required and the need to reduce processing load on therapeutic devices.

The research by Direito (2011) examined the impact of feature selection methods and used three different electrode configurations: randomly selected; spatially distributed; and based upon seizure localisation. They found that the optimal set of electrodes is highly patient specific. The same three electrode configurations are used by Alexandre Teixeira et al. (2014) for their patient-specific classifiers. They maintain that using fewer electrodes models the conditions for a real-world clinical application,

since few patients are willing to wear large arrays of electrodes for sustained periods; six electrodes are a reasonable compromise between EEG information and patient comfort. However, their results showed that none of these combinations of electrodes led to improved performance, suggesting a "global preictal process" (Alexandre Teixeira et al., 2014, p.334).

2.5.2 Time Series Segmentation

EEG signals are considered to be non-stationary and therefore it is important to extract spectral features from a small time frame (A. H. Shoeb & Guttag, 2010). According to Aarabi, Fazel-Rezai, and Aghakhani (2009), linear methods require a signal to be stationary, which is achieved by analysing the data in smaller segments or *windows*. Carney et al. (2011) concur that the main assumption in linear modelling is the stationarity of the signal and consequently, EEG signals must be segmented. Most of the studies reviewed by Mormann et al. (2007) use a moving window analysis, where the features are calculated from a window of EEG data with a defined length.

The EEG signal is split into a sequence of consecutive windows, and a feature vector is extracted from each window in turn until the data is fully parsed. Windows may be contiguous or overlapping (usually by 50%). See Figure 2.2. The extracted feature vectors from each window are merged into a feature matrix - each EEG file is divided into multiple segments, and each segment is classified individually. A majority voting mechanism is used to classify each EEG file as interictal or preictal, based on the combined results for each window of data (Gadhoumi, Lina, & Gotman, 2013).

2.5.3 EEG Features

EEG features describe the brain's state and are extracted from the EEG time-series and assembled into a feature vector. The feature vector is then passed to a classification algorithm to determine if the input represents a preictal or interictal state. This section discusses the types of EEG features used in current research and the ongoing debate

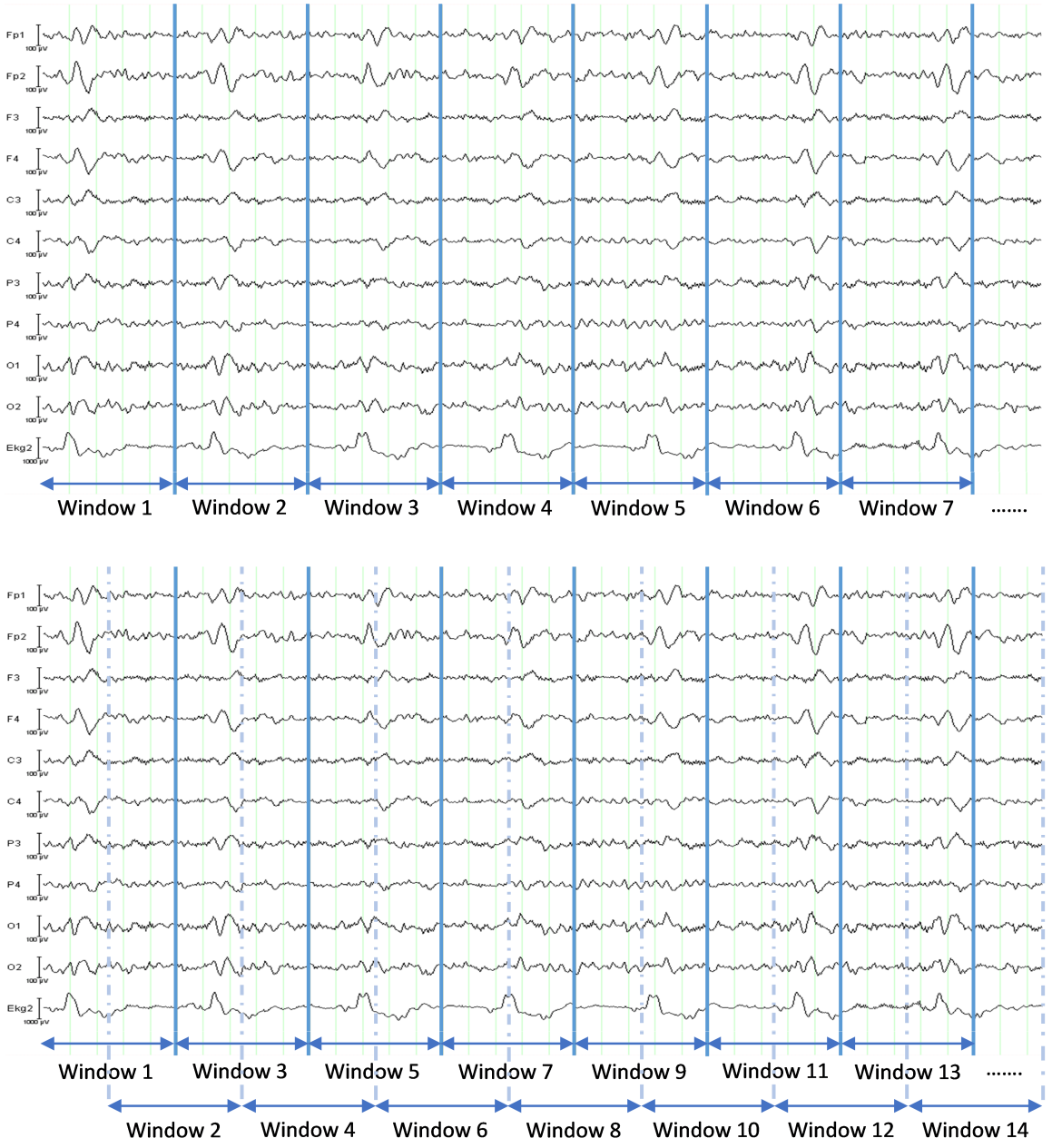

Figure 2.2. EEG signal segmentation. Time-series data is sliced into equal duration segments or windows. The top figure shows contiguous windows, and the lower figure shows overlapping windows, with 50% overlap.

concerning which features have the best predictive power. It covers spectral features in depth, and gives an overview of other commonly used features.

Types of Features

EEG features are broadly categorised as either linear or non-linear (Aarabi et al., 2009; Gadhoumi et al., 2016). Linear features are derived from the phase and amplitude of the EEG signals, whereas non-linear features are extracted using the non-linear dynamics of the signal. Non-linear theory considers the brain as a non-linear dynamical system; the system *state* describes the system at any given time, and the system *dynamics* describes the rules for how the system state evolves over time (Aarabi et al., 2009). An EEG signal can be considered as the output of a non-linear system and can consequently be characterised deterministically (Asha et al., 2013).

Features may further be considered as univariate (single channel), bivariate (dual channel) or multivariate (multiple channel), depending on the number of EEG electrodes from which they are extracted. Univariate features characterise EEG time series based on information from a single recording site. Bivariate and multivariate features are taken from two or more electrodes and represent relationships between different regions of the brain, depending on the electrode locations. Multivariate analysis considers the interactions between channels and how they correlate with each other over time.

Choosing Features

Carney et al. (2011) stated that no measure can reliably predict seizures with the required levels of specificity and sensitivity. Which feature, or combination of features, are the best at predicting seizures is a subject of continuing research (Gadhoumi et al., 2016), and the profusion of different features used in the literature is evidence of this ongoing uncertainty.

Lehnertz et al. (2001) reported that bivariate features and univariate features provide different but complementary information. Mormann et al. (2005) concur and demon-

strated that bivariate measures show synchronisation between electrodes, and that channels far from the seizure source contain important information. They conducted a broad study to compare linear and non-linear methods and concluded that a combination of univariate and bivariate measures offer the best approach for prediction studies. In a subsequent paper, Mormann et al. (2007) noted that recent studies reported that channels in remote areas carried relevant information, which supports the idea of an epileptic network across large parts of the brain, as opposed to a limited and localised epileptic focus.

Jouny et al. (2011) stated that analysis of single-channel data could not provide information on the interaction between different brain region. However, Rasekhi et al. (2013) take the opposite view and claim that studies using linear univariate measures have shown good performance, with the caveat that the results are less certain when extended to larger datasets. The researchers maintain that combining different univariate features into one feature space is a promising new research area that can be exploited by machine learning approaches.

In their review focusing on EEG features, Aarabi et al. (2009) conclude that bivariate measures are generally more effective. Bandarabadi, Teixeira, et al. (2015) lend further support to the importance of bivariate features and report that features such as phase or lag synchronisation give better functionality. The authors further speculate that univariate features are influenced the general state of the brain and thus cannot guarantee high discrimination. P. Mirowski, Madhavan, LeCun, and Kuzniecky (2009) found that a combination of linear and non-linear time and frequency-domain features are optimal. Conversely, Niknazar et al. (2013) claim their results prove that non-linear features based on dynamical system theory are superior to linear features.

The heterogeneity of seizures is due to different underlying pathologies, and EEG dynamics that signal the transition to seizure state are not necessarily consistent. A study by Alexandre Teixeira et al. (2014) found that combining features has a greater potential for correctly classifying heterogeneous seizures seen in refractory epilepsy. They reason that a feature that works well for one patient may perform poorly for

another, and a combination of features may provide a degree of robustness. This is also reported by Moghim and Corne (2014), who demonstrate that the optimal feature set varies from patient to patient.

In conclusion, the research is far from unanimous but generally supports the view that a mixed univariate and bivariate approach can provide better insights into the spatiotemporal activity responsible for seizure generation and propagation. A mixture of features provides different but complementary information, and logically a combination of both types will provide superior predictive performance. However, this comes at the cost of additional algorithmic complexity and increased processing time.

Spectral Features

Early EEG processing methods were based on simple analysis of EEG signals in the time and frequency domains (Niknazar et al., 2013). Time-domain features measure components of the EEG signal that are a function of time and are extracted using classical and advanced signal processing techniques. These features are used to characterise linear and non-linear behaviours in EEG waveforms, and describe three aspects of the signal: amplitude, variability and synchronicity (Gadhoumi et al., 2016).

The frequency-domain features are derived from a transformation of the time-domain signal and represent the spectral content of the EEG. The transformation to frequency-domain can be performed by either discrete Fourier transform (DFT) or a discrete wavelet transform (DWT) (Fergus et al., 2016). The DFT is a frequency domain representation of the input function, and is generally implemented using a fast Fourier transform (FFT) - as such, the terms DFT and FFT are often used interchangeably. On the other hand, the DWT transforms the signal in a way that captures the frequency and the temporal localisation properties, and is used to extract the signal energy of a specific frequency in a specific time interval (Moghim & Corne, 2014). Wavelet transforms are considered more suitable for functions with discontinuities and sharp peaks, and for non-periodic and non-stationary signals like EEG (Asha et al., 2013).

The power spectrum from one or more channels of an EEG is one of the more popular measures used in seizure prediction research (Carney et al., 2011). A number of studies have proved the usefulness of EEG signal spectral power in seizure detection and prediction. Calculating spectral power features is also computationally effective which makes them suitable for portable seizure warning systems (Bandarabadi, Rasekhi, et al., 2015). The central hypothesis is that an EEG signal has a distinct spectral signature that varies between the preictial and interictal states, which can be discriminated by classifiers.

Spectral power or power spectral density (PSD) is defined as the statistical estimate of the power at each frequency and is used extensively in prediction experiments (Gadhoumi et al., 2016). Spectral power shows the distribution of power of a signal within each frequency sub-band of interest. Howbert et al. (2014) and Brinkmann et al. (2015) use a *power-in-band* (PIB) feature, which is the sum of the power in each frequency band of interest. Lastly, signal energy is calculated as the mean of the signal over a given period. More formally, the signal energy is the square of the signal's amplitude divided by the period of measurement. It is used by Costa, Oliveira, Rodrigues, Leitao, and Dourado (2008) and Moghim and Corne (2014), among others.

EEG Frequency Bands

An EEG signal has multiple spectral components that combine to indicate the presence or absence of seizure activity. Studies usually consider EEG signals in terms of five well-known frequency sub-bands used by clinicians (Aarabi et al., 2009; Bandarabadi, Teixeira, et al., 2015). The sub-bands are generally denoted (in ascending frequency): Delta 'δ'; Theta 'θ' ; Alpha 'α' ; Beta 'β'; and Gamma 'γ'. However, there is considerable variation in the frequency ranges used to analyse EEG in the literature, and Brinkmann et al. (2015) note that there is no empirical or theoretical basis for using these bands for seizure predictions. See Table 2.1 for a summary of the frequency bands used in studies referenced by this paper.

The *peak frequency* is derived from the PSD, and is the frequency of the highest peak in

the distribution. This is the dominant cyclic component of the signal in the frequency domain (Fergus et al., 2016) and is a commonly used feature. However, the dominant spectral component of a signal indicating a seizure state can considerably overlap that of a signal with non-seizure activity, and key information may be contained in the other frequencies (Kharbouch et al., 2011). They further argue that the important features for characterising EEG activity are the spectral distribution, the specific channels on which they exist, and the short-term temporal evolution.

The efficacy of spectral features is widely reported. Mormann et al. (2005) claimed that higher frequencies are a useful predictor of seizure activity, and reported a decrease of delta power and a transfer of power from lower to higher frequencies prior to seizure. Bandarabadi, Rasekhi, et al. (2015) stated that spectral power features in the gamma frequency bands have demonstrated good results, and similarly Park et al. (2011) found that the gamma bands were the most discriminatory in 8 out of 13 patients. They hypothesise that this is due to an increase in iEEG spikes, and that predictive performance would improve if even higher frequencies could be assessed. Stacey et al. (2011) agree with this assessment and state that oscillations in the gamma bands (40-120 Hz) are commonly found in iEEG of partial seizures.

Sharma (2015) divided the EEG signal into the five standard sub-bands and propose spectral power in the beta band as a predictor of epileptic seizure. Furthermore, they claim that their method permits retrospective localisation of the seizure focus based on the region of the brain displaying the largest change in beta power. Bandarabadi, Rasekhi, et al. (2015) found that higher frequency features were generally more discriminative, but this varied according to the type of EEG recording. Similarly, a study by Netoff, Park, and Parhi (2009) divides spectral power into 9 frequency bands, and they argue that power in the higher frequency bands plays a key role in seizure prediction. All of the top 6 teams in the 2014 Kaggle seizure prediction competition used some form of spectral power, split across discrete frequency bands. Furthermore, five out of the six used time domain and/or frequency domain interchannel correlations, which measure sychronisation between electrodes (Brinkmann et al., 2016).

Table 2.1

Commonly used Frequency Bands in EEG Signals

Study	Feature	Frequency Band (Hz)							Other
		'δ'	'θ'	'α'	'β'	'γ'	Low 'γ'	High 'γ'	
Costa et al. (2008)	Energy								0-12.5/12.5-25/25-50/50-100
P. W. Mirowski et al. (2008)	Synchrony	<4	4-7	7-13	13-30		30-45	65-100	Beta into high and low
Netoff et al. (2009)	Power	0-4	4-8	8-12	13-30	>30			
Park et al. (2011)	Power	0.5-4	4-8	8-13	13-30	30-47	53-75	75-97/103-128	
A. Shoeb et al. (2011)	Energy								0-25 in 1 Hz bands
Rasekhi et al. (2013)	Power	<4	4-8	8-12	13-30	>30			
Howbert et al. (2014)	Power	0.1-4	4-8	8-12	13-30		30-70	70-180	
Moghim and Corne (2014)	Energy								0-12.5/12.5-25/25-50/50-100
Moghim and Corne (2014)	Power	0.5-4	4-8	8-13	13-30	30-48			
Bandarabadi, Teixeira, et al. (2015)	Power	0.5-4	4-8	8-15	15-30	>30			
Bandarabadi, Rasekhi, et al. (2015)	Power	0.5-4	4-8	8-15	15-30	30-48	52-75	75-98	Ten other bands
Brinkmann et al. (2015)	Power	0.1-4	4-8	8-12	13-30		30-70	70-180	
Sharma (2015)	Power	0.5-4	4-4.75	8-13	14-30				0.5-50/0.8-30/0.8-47
Karumuri et al. (2016)	Entropy	0.5-4	4-7	8-15	16-31	32-50			
Brinkmann et al. (2016)	Entropy	0.1-4	4-8	8-12	12-30		30-70	70-180	

Statistical Features

Statistical analysis of EEG signals is usually calculated from four well understood parameters: mean, variance, skew and kurtosis. These measures are included in studies by Fergus et al. (2016), Alexandre Teixeira et al. (2014), Moghim and Corne (2014), Rasekhi et al. (2013) and others. More formally, these are known as the first-order, second-order, third-order and fourth-order statistical moments. The mean and variance are a measure of the location and span of the signal distribution (Rasekhi et al., 2013) and skew and kurtosis are measures of the symmetry of the amplitude distribution of the signal. Skew can be positive (for left-skew), negative (for right-skew) or zero for symmetric distributions, and kurtosis is a measure of the peakedness of the amplitude distribution.

According to Mormann et al. (2005), a decrease in variance and an increase in kurtosis is associated with the preictal stage. Carney et al. (2011) maintain that statistical moments may be useful for early seizure detection of large-amplitude seizures. Whilst statistical features are included in many studies, their efficacy is generally unclear. However, these measures are easily understood and are computationally simple.

Other Features

Karoly et al. (2016) explore the relationship between interictal spikes and seizures and conclude that spike rate and spike distribution can be used to aid prediction. However, their results are patient-specific and they are unclear whether spike rate is a symptom of a pre-seizure state or a causal factor. Interictal spikes are also used in a study by Sudalaimani et al. (2015), in conjunction with electrographic seizures. Salami, Lvesque, Gotman, and Avoli (2012) examine high frequency oscillations and assert that these are better markers of seizure foci compared to interictal spikes.

The research by Aarabi and He (2012) exclusively uses non-linear features, including correlation dimension, correlation entropy, Lyapunov exponent and bivariate non-linear interdependence extracted from 10 second EEG segments. Sudalaimani et al.

(2015) use a combination of linear and non-linear features - maximum cross correlation (MCR), phase lock value (PLV) and maximum Lyapunov exponent (MLE). The Lyapunov exponent is a commonly used non-linear feature and is used to detect the transition from chaos to synchronicity.

Papers by Feldwisch-Drentrup et al. (2011), Zheng, Wang, Li, Bao, and Wang (2014) and Karumuri et al. (2016) used a mean phase coherence (MPC) feature that measures the degree of interaction between signals in pairs of EEG channels. It calculates the coherence between two signals from electrodes located at different points of the skull; high coherence indicates high connectivity between the two sites. MPC was estimated by applying a sliding window of 30 second duration and shifting in 5 second increments. Pairs of electrodes were selected based on a pre-selection method that calculated the ratio of the global variance to the local variance.

The Hjorth parameters are three time-domain measures used to quantitatively describe EEG signals. They include activity, mobility and complexity and are used in studies by Rasekhi et al. (2013) and Alexandre Teixeira et al. (2014). Increases in activity and mobility are reported in EEG during the preictal period (Mormann et al., 2005).

To conclude, there are a multitude of features used in the literature, and no clear agreement on which features or combinations of features are the most predictive. For further background reading, a good overview of EEG features can be found in papers by Aarabi et al. (2009) and Carney et al. (2011).

2.5.4 Data Normalisation

Rasekhi et al. (2013) conducted a patient-specific study to examine the effects of pre-processing and experimented with outlier treatment, smoothing and normalisation on a set of 22 linear univariate features. According to the authors, EEG features fluctuate within short periods of time and these can be affected by daily activities, masking the effects of the epileptic activity. Their study implemented feature smoothing using a 1 minute moving average and they concluded that normalisation and smoothing give

the optimal results for most patients. Smoothing reduces the impact of short-term periodic fluctuations, and should improve classification performance.

In general, an EEG's spectral power is concentrated in the lower frequency bands, which makes it difficult to compare power across the whole signal (Park et al., 2011). The solution is to normalise the power in each band with respect to the total power across all the bands. This technique is used by Bandarabadi, Rasekhi, et al. (2015) and Bandarabadi, Teixeira, et al. (2015) to normalise the spectral power feature to the range [0, 1]. Similarly, Kharbouch et al. (2011) note that time-series EEG signals have a spectral amplitude profile that is inversely proportional to the frequency. In their experiment, they remove this frequency domain trend by applying a derivative function to the time-series for each EEG channel.

According to Rasekhi et al. (2013), the normalised spectral power is a more robust measure for detecting the preictal state, and using normalised values instead of absolute values can help isolate the variations in brain signals caused by daily life. The authors compared three methods of normalisation: range normalisation; dividing by the mean of the absolute values; and using standard scores. They found that range normalisation gave better results compared to the two others.

2.5.5 Feature Selection

Feature selection is the process of determining a subset of relevant features that have the greatest predictive strength. It provides an optimal characterisation of the dataset (Direito, 2011) and helps to simplify model interpretation, reduce training times and lower the risk of overfitting. Feature selection plays an important role in constructing robust machine learning models and helps to reduce computational complexity.

On the whole, the coverage of EEG feature selection in the literature is weak. Many studies uses large numbers of features in order to optimise classification performance, at the expense of producing algorithms that cannot be practically implemented. Fergus et al. (2016) compare a number of feature selection methods to rank the discriminant

capabilities of their features. Techniques include statistical significance (p and q-*values*), principle component analysis (PCA), Linear Discriminant Analysis (LDA) and Gram-Schmidt (GS) analysis. They found that a backward search LDA gave the best results, in conjunction with a k nearest neighbours (KNN) classifier.

Bandarabadi, Teixeira, et al. (2015) contrast two feature selection techniques and assess their impact on performance. The minimum redundancy maximum relevance (mRMR) method is compared to a novel technique that evaluates an amplitude distribution histogram (ADH) of the preictal and interictal classes. The idea behind an ADH is to select features that have the maximum distance (or minimal overlap) between distributions. The authors found that the ADH method gave superior results compared to mRMR. Direito (2011) also used the mRMR method and showed that while it degrades performance compared to the full set of features, it also helps to reduce computational costs.

2.5.6 Data Partitioning

Mormann et al. (2007) note that performance results should be reported for the testing data only. They found that previous studies incorrectly optimised the algorithm using the test data - by incorporating the test data into the training process, better results were achieved than should have been the case. Stacey et al. (2011) also caution against measuring performance using 'in-sample' data and stipulate that EEG data sets must be 'strictly' divided into training and testing sets.

The validation dataset is used to compare the performance of the trained models, and can be constructed in a number of ways. The simplest technique is the holdout method, which splits the training data into a training and validation set according to a simple ratio, usually 70%/30% respectively. Fergus et al. (2016) use holdout validation with an 80/20 split for training and validation. The studies by both Costa et al. (2008) and Moghim and Corne (2014) use a 70/30 split i.e. 70% of the data is used for training the classifiers and 30% of the data is used for model validation.

In k-fold cross validation, the samples are randomly divided into k equal-sized folds or partitions, and k separate experiments are performed. In the initial iteration, the first fold is used for validation, and the other k-1 folds are used for training; the second iteration uses the second fold as the validation sample and the remaining k-1 folds for training, and so on until k evaluation experiments have been concluded (Kelleher, Namee, & D'Arcy, 2015). The results of the performance measures are aggregated to provide a set of performance figures for final model selection. Yu et al. (2016) and Howbert et al. (2014) both use 10-fold cross-validation in their methodology, whereas Brinkmann et al. (2015) use 5 folds.

Park et al. (2011) use a double cross-validation method, which divides the data into training and testing sets; the training set is further divided into a learning set and validation set. The algorithm is tested on data that has not been used during training and optimisation, and thus better reflects the predictive performance in real-world conditions. A. Shoeb et al. (2011) use leave-one-out cross validation testing when evaluating performance. If N is the total number of feature vectors comprising both preictal and interictal records, then the classifier is first trained on N-1 records. Next, the classifier is tested on the withheld record and the process is repeated N times - in this manner each record is used for training N-1 times and used for testing once. This method is also used in studies by Kharbouch et al. (2011); A. H. Shoeb and Guttag (2010); Cook et al. (2013) and Karumuri et al. (2016).

2.5.7 Data Sampling

The relative scarcity of seizures in EEG data generally results in a highly imbalanced dataset, where the preictal class has relatively fewer samples compared to the interictal class. According to Qiao and Liu (2009), standard classifiers treat all classes the same, which can lead to problems with imbalanced data sets because the minority class can potentially be ignored. Models trained on data that predict only the majority class can achieve accuracy greater than 90%, which is clearly a misleading result. The class imbalance problem is discussed by Alexandre Teixeira et al. (2014).

Conventionally, class imbalance can be addressed by reducing the majority class by down-sampling, or increasing the minority class by up-sampling. Alternatively, Qiao and Liu (2009) recommend using weighted classes, where the minority class is given a greater weight relative to the majority class. This can be accomplished with a cost sensitive classifier with different costs for each class,but this approach can only compensate to a limited extent (Alexandre Teixeira et al., 2014).

With majority class down-sampling, a reduced number of records from the majority class are randomly selected, such that the resulting number of interictal records approximately matches the smaller number of preictal records. Down-sampling can reduce redundant information, over-training and computational costs of the trained model. This technique is used Bandarabadi, Teixeira, et al. (2015) and Direito (2011) to reduce the non-preictal samples in their data to balance the classes.

Minority up-sampling increases the number of records from the minority class, such that the final number of preictal records approximately equals the greater number of interictal records. The up-sampling can be achieved by taking multiple copies of the same samples, or by interpolation using synthetic minority over-sampling technique (SMOTE). This algorithm is used by Karumuri et al. (2016) to balance the interictal and preictal classes before training. However, up-sampling can substantially increase computation times and a trade-off must be struck between the required performance and the resources available for modelling.

2.6 Classification Algorithms

Predicting the onset of a seizure is based on detecting dynamic changes in EEG signals preceding the event. EEG signals may be classified as interictal (between seizures) or preictal (preceding seizure) - classifiers aim to distinguish interictal from preictal signals in a prediction window, before disabling symptoms occur. The amount of time between a correctly predicted seizure and the seizure onset represents a *window of opportunity* in which an intervention may potentially alter seizure evolution. See

Figure 2.3. The length of this window is dependent upon the ability of the prediction algorithm to detect changes in the EEG as soon as possible. It is also a function of the preictal period, which can potentially present hours or minutes in advance of seizure onset.

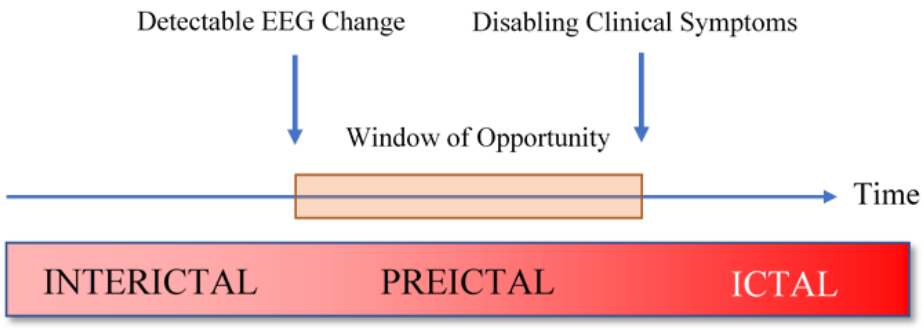

Figure 2.3. Window of opportunity for seizure prediction

Patient-Specific and Non-Specific Algorithms

Algorithms can generally be described as patient-specific or non-specific (A. Shoeb et al., 2011). Patient-specific algorithms train the prediction or detection algorithm with interictal and ictal feature vectors from an individual patient. The resultant trained classifier is then used to classify that particular patients' test data. Conversely, a non-specific algorithm is trained with data from multiple patients and is intended to derive a generalised detection or prediction algorithm. Characteristics of EEG vary considerably from one patient to another - signals associated with seizure onset in one patient may represent a normal brain state in another. However, the seizures of an individual are considerably consistent, provided that the seizure's genesis is in the same brain region. Given the heterogeneity of seizure patterns, a patient-specific algorithm will generally perform poorly if applied to other patients' data.

Costa et al. (2008) compare the results from both patient-specific and generalised classifiers. The performance of the generalised approach showed considerable degradation compared to the patient-specific approach, and the authors note that the "classifier of one patient cannot be used for another" (Costa et al., 2008, p.7). Brinkmann et al.

(2015) note the variations in preictal iEEG patterns and the subsequent need for subject specific predictive models. As a general rule, non-specific algorithms are well suited to patients with stereotypical EEG patterns and patient-specific algorithms perform best for patients with uncommon seizure patterns. A limitation of patient-specific classifiers is that they must be individually trained based on a personalised study for each patient (Alexandre Teixeira et al., 2014), and therefore could not be used 'off-the-shelf' in a seizure prediction device aimed at a wider group of patients. High-performance generalised algorithms are certainly harder to develop, but are nonetheless a more practical approach for therapeutic devices.

2.6.1 Commonly Used Classifiers

The methods for seizure prediction may be characterised as proof of principle, statistical or algorithmic (Mormann et al., 2007). The type of feature and the way it is tracked is the defining characteristic of a prediction algorithm (Gadhoumi et al., 2016). Some studies attempt to solve a ternary classification problem and aim to distinguish between interictal, preictal and ictal states. However, most studies use binary classifiers to identify the preictal and interictal classes.

There are many predictors and classification algorithms used in seizure prediction studies, and machine learning classifiers dominate. Popular machine learning algorithms include KNN, decision trees, ANN and the SVM. There are also a number of less commonly used approaches including: threshold crossing; rule-based approaches; autoregressive (AR) models; linear discriminant analysis; visibility graphs and outlier detectors. With the exception of Temko et al. (2015), ensemble models are rarely used, probably due to their high processing time and complexity. According to Alexandre Teixeira et al. (2014) machine learning classifiers produce better results than other types of predictors, particularly thresholding. They believe that their own study 'reinforces the importance of machine learning techniques' as a viable approach in epileptic seizure prediction research (Alexandre Teixeira et al., 2014, p.334).

2.6.2 Support Vector Machines

Chisci et al. (2010) list three reasons why they believe SVM are suitable for seizure forecasting. Firstly, the computational load is generally greatest during off-line model training, making SVM suitable for real-time applications. Secondly, they have good theoretical and practical capabilities and lastly, SVM can be efficiently optimised during training, allowing them to solve high-dimension problems.

By definition, the SVM is a binary classifier but it can be extended to a multi-class problem. This is achieved by reducing the multi-class problem to several two-class problems, to which the SVM classifier may be applied in turn (Alexandre Teixeira et al., 2014). During model training, the SVM transforms the instance-space and derives an optimal boundary called a *hyperplane* in the transformed space. The hyperplane is used to determine the class membership of newly observed vectors and is calculated such that it "maximises the distance... from the hyperplane to the closest transformed data instances on either side of the plane (the support vectors)" (Moghim & Corne, 2014, p9). See Figure 2.4.

The simplest SVM uses linear boundaries to separate two-class data. Classification of complex datasets with non-linear boundaries is achieved by transforming the feature space into a higher order space where linear boundaries can be used. This transformation is performed by *kernel* functions and the most commonly used is the Gaussian radial basis function (RBF) (Bandarabadi, Teixeira, et al., 2015; Moghim & Corne, 2014). The use of kernel functions means that SVM training has reasonable computational costs despite high dimensional transformations (Moghim & Corne, 2014).

Choosing the type of boundary depends on the amount of training data and the distribution of the data within the feature space (A. Shoeb et al., 2011). Non-linear decision boundaries are frequently used since linear decision boundaries may result in poor performance. Peformance of non-specific algorithms come at the cost of greater within-class diversity (due to the heterogeneity of seizure patterns) and between-class overlap (A. Shoeb et al., 2011). Consequently, SVM with linear decision boundaries

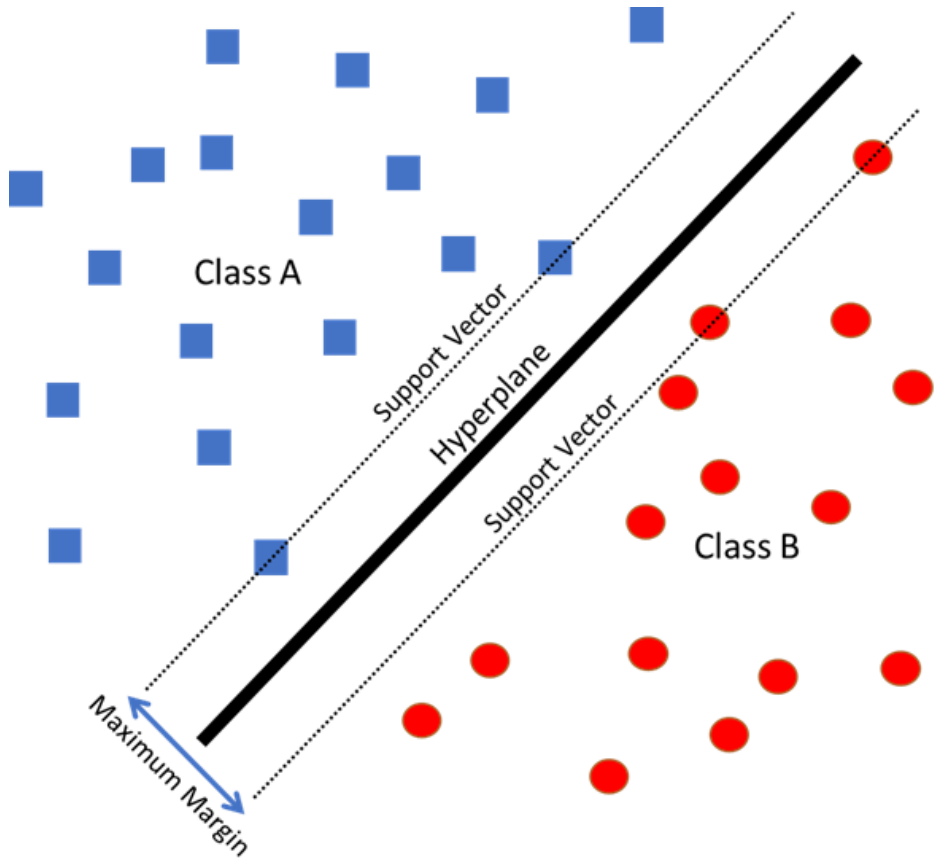

Figure 2.4. Two-class feature space with hyperplane and associated support vectors

are unable to correctly distinguish preictal from interictal patterns. For this reason, A. H. Shoeb and Guttag (2010) use a non-linear boundary with RBF kernel in their study.

EEG datasets are typically highly imbalanced, with many more interictal instances than preictal instances. Moghim and Corne (2014) used a cost-sensitive support vector machine (CSVM) which applies a weight value to each class, in order to normalise the influence of each class on the trained model. The weights are calculated using ratios of the number of interictal instances to the other class instances. The CSVM is also used by Park et al. (2011), who recommend it for handling imbalanced datasets.

Tuning a SVM with RBF kernel is performed using a parameter γ (which characterises the kernel) and a regularisation parameter C (which characterises the SVM) (Moghim & Corne, 2014). The γ parameter is used to determine the trade-off be-

tween maximising the boundary's margin width and minimising the misclassification rate (Bandarabadi, Teixeira, et al., 2015). Typically, parameter tuning is accomplished via a grid search of combinations of (C,γ) using a k-fold cross-validation run (Moghim & Corne, 2014).

2.6.3 Neural Networks

Artificial neural networks are networks of connected nodes, or *neurons* that mimic biological neural networks. Connections between neurons are analogous to *synapses* and transmit signals between neurons. Neurons receive signals, process the signals and then retransmit the transformed signals to other nodes. The neurons are typically arranged in layers and signals traverse the network from the first (input) layer to the last (output layer). The layers in between input and output can be crossed multiple times, depending on how the signals are fed forward or backward. Weights are applied to the neurons and synapses, which are used to increase or decrease the signal strength as it transits the network. A simple ANN generally includes an input layer, a hidden layer and an output layer of neurons. See Figure 2.5.

Artificial neural networks are suitable for classification problems due to their suitability for parallelism, adaptive learning and robustness. Neural networks have good potential as classifiers due to their ability to compute in real-time with high dimensional feature space (Costa et al., 2008). A prediction study by Alexandre Teixeira et al. (2014) uses two types of ANN structures - a multilayer perceptron (MLP) and a RBF network. The MLP was tested with two and three hidden layers, with up to 400 neurons in one layer; the RBF was trained with a maximum of 50 hidden neurons. The paper provides further details on initialisation and optimisation parameters for the models. Kumar et al. (2008) use a recurrent neural network for seizure detection. Since it is a feedback backpropagation network, it is able to store information from previous time intervals. The model was trained using a gradient descent backpropagation algorithm and the best results were obtained with 90 neurons, to give accuracy of 99.75%.

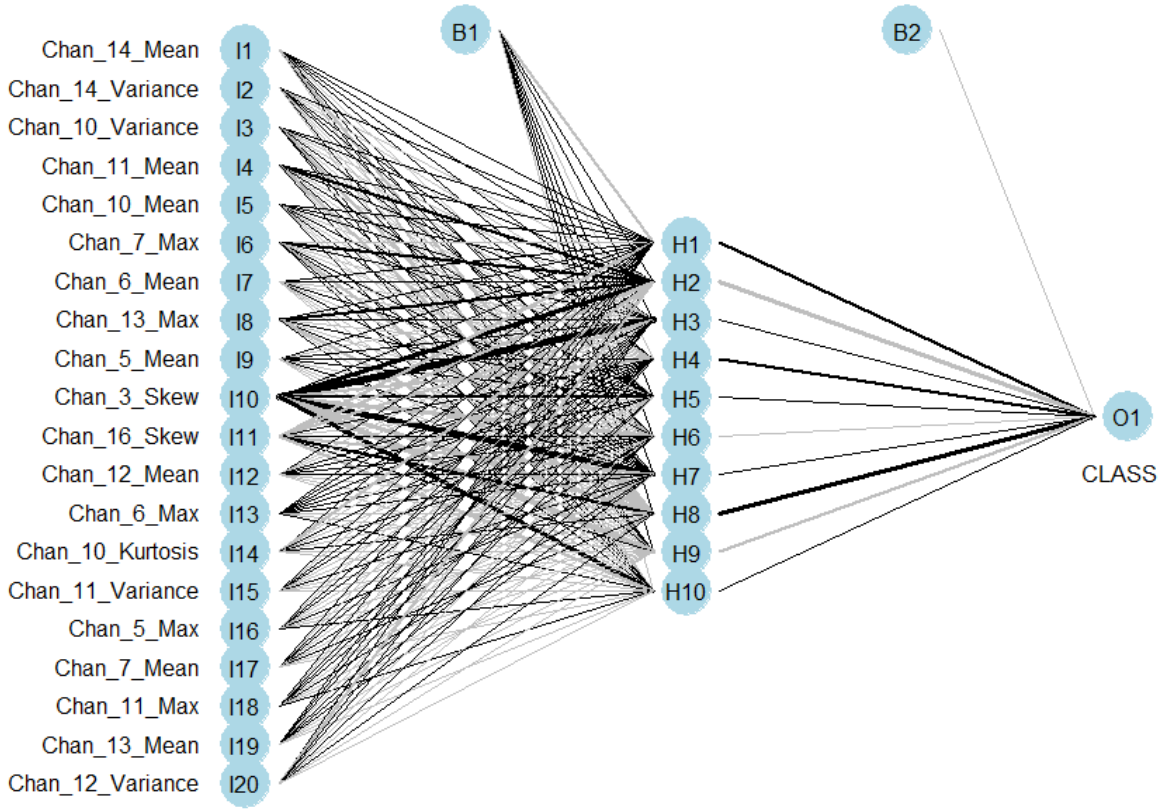

Figure 2.5. Simple artificial neural network architecture with one hidden layer. I1,..I20 represent the input layer. H1,..H10 represent neurons in the hidden layer. B1, B2 are bias neurons and O1 represents the output layer.

Costa et al. (2008) developed six EEG classifiers using six different neural network architectures. The variants include RBF; feed forward back-propagation; layer-recurrent networks; Elman networks; distributed time delay and feed-forward input time-delay back-propagation. They reported very good specificity and sensitivity results of $>$ 90%, when using patient-specific classifiers. A generalised approach showed excellent sensitivities (>97%) for the RBF network, but required more than 3000 neurons, which is computationally inefficient. The study is also quite limited, since it only uses data from two patients with a total of five seizures.

2.6.4 SVM versus ANN

The support vector machine is "one of the most powerful tools in signal processing and machine learning" (Rasekhi et al., 2013, p.13). According to Fergus et al. (2016) the support vector machine is the most commonly used classifier in seizure prediction. It is used in the following studies; Chisci et al. (2010); A. Shoeb et al. (2011); Kharbouch et al. (2011); Alexandre Teixeira et al. (2014); Sudalaimani et al. (2015) and Bandarabadi, Teixeira, et al. (2015), to name a few.

In their literature review, Gadhoumi et al. (2016) also note the popularity of SVM classifiers - principally due to their robustness - but believe their predictive success is primarily due to the correct choice of EEG features. The SVM was also used by four of the six top submissions in a seizure prediction competition hosted by Kaggle (Brinkmann et al., 2016). Scores from the competition were ranked by AUC, so sensitivity and specificity figures are unfortunately not available.

Whilst the SVM is arguably the most commonly used seizure prediction algorithm, relatively fewer studies use artificial neural network classifiers. P. W. Mirowski et al. (2008) conducted one of the first comparative studies of neural networks and SVM, and reported that the ANN achieved zero false alarms in 20 out of 21 patients, compared to 11 out of 21 for the SVM. Both SVM and ANN achieved 100% sensitivity for some patients.

In a patient-specific study, Alexandre Teixeira et al. (2014) compared two types of ANN and SVM classifiers, and found that a feedforward ANN had greater sensitivity on average compared to the SVM. However, the results of a Kruskal-Wallis test showed that the results were not significant. They also concluded that the SVM achieved lower false positive rates compared to ANN ($p < 0.01$). Conversely, Asha et al. (2013) found that SVM classifiers provide superior classification accuracy compared to ANN. However, the models were not trained on any interictal data and the training data was manually selected, so these results are difficult to assess.

On the whole, the evidence for using either SVM or ANN classifiers for seizure predic-

tion is insubstantial. Whilst SVM is more frequently implemented, some of the results are either ambiguous or statistically insignificant. Understanding which seizure classifier delivers the best results would be an important step towards the development of a viable seizure warning device.

2.7 Evaluation and Results

The definition of a seizure when assessing algorithm performance in a clinical setting is complicated. Generally, predictive performance focuses on clinical seizures (i.e. seizures with a clinical impact), and subclinical electrographic events (that have a minimal impact) are not included for consideration (Snyder et al., 2008). The number of patients in any study has an influence on the sensitivity score; studies with few patients tend to have very high sensitivity scores (Costa et al., 2008). Adding data from more patients generally decreases the sensitivity scores, due to the increased heterogeneity of the seizures. Schulze-Bonhage et al. (2011) demonstrate this dynamic by showing a negative correlation between performance and seizure numbers. This again highlights the need for public databases with long-term EEG recordings from multiple patients, and caution must be exercised when reviewing results.

2.7.1 Statistical Validity

A common issue with initial studies were doubts concerning the statistical validity of the algorithms' performance (Mormann et al., 2005). According to Andrzejak et al. (2009), many early seizure predictions models often claimed good results, but were later proven to be irreproducible or were unsubstantiated. The authors further speculate that the apparent predictive power of many algorithms is due to a failure to test them against appropriate null hypotheses.

The performance of a prediction method must be superior to a random or periodic predictor (Maiwald et al., 2004). Similarly, Snyder et al. (2008) proposed new methods for confirming significance and proposed that performance should be validated by either

Monte Carlo simulations or by comparison with a naïve (random or periodic) predictor. Andrzejak et al. (2009) examined the requirements for well-defined hypotheses for establishing predictive power, and highlighted the importance of determining the statistical significance of results.

Statistical validity can be determined by comparing any results with those obtained by a random naïve predictor that does not use any information contained in the EEG (Feldwisch-Drentrup et al., 2011). If the observed sensitivities achieved by the predictive technique exceed those of the random predictor, then the results may be considered significant. Research by Howbert et al. (2014) and Brinkmann et al. (2015) use a Poisson-process random predictor, as proposed by Snyder et al. (2008).

Moghim and Corne (2014) use several baseline and random predictors for performance comparison. The baselines are two predictors that always predict preictal, and always predict non-preictal respectively. Two random predictors are also used, which are set to be 'lucky' by setting the probability to the correct frequencies of the classes.

2.7.2 Sensitivity and Specificity

Maiwald et al. (2004) noted that half of the papers they reviewed reported sensitivity only, and did not provide any calculation of specificity. Early seizure prediction research also failed to agree on which specificity measure to use, making algorithm comparison difficult (Gadhoumi et al., 2016). As a solution, they propose that seizure prediction studies should report algorithm performance in terms of sensitivity and specificity.

Whereas sensitivity is unilaterally defined as the proportion of true predictions, there are several ways in which specificity is reported; the two most common are false prediction rate (FPR) and time under false warning. Similarly, Feldwisch-Drentrup et al. (2011) define sensitivity as the ratio of correctly predicted seizures and specificity as the rate of false predictions, whilst Andrzejak et al. (2009) define sensitivity as the number of true positives normalised by the total number of seizures.

Sensitivity is the "primary performance measure of interest..." and measures the ability of the algorithm to detect seizures (Moghim & Corne, 2014, p.12). In practical terms, a sensitivity of 90% may be interpreted as meaning that 90% of seizures that occur will be predicted, and 10% will be missed. A deficit in sensitivity score is correlated with direct danger for the patient. In contrast, poor specificity scores result in high rates of false alarms. From a patient perspective, these are arguably more benign than a failure to correctly predict a seizure event, but frequent false alarms would cause unnecessary stress to patients, and may even exacerbate their helplessness and depression (Schelter et al., 2006). Patients may also disregard future alarms, potentially placing themselves in danger (Maiwald et al., 2004).

Sensitivity and specificity are the most commonly used measures for evaluating algorithm prediction performance and are used by many studies including Kharbouch et al. (2011); Aarabi and He (2012); Feldwisch-Drentrup et al. (2011); Alexandre Teixeira et al. (2014); Moghim and Corne (2014) and Sudalaimani et al. (2015) among others. A challenge for prediction systems is the required trade-off between the requirement for 100% sensitivity and 0% false positive rate (P. W. Mirowski et al., 2008; Chisci et al., 2010); tuning classification algorithms to maximise either of the two measures invariably improves one at the expense of the other. Choosing which measure to optimise depends on the specific therapeutic application (Howbert et al., 2014).

If the therapy is a device that administers low-dosage medication with limited side-effects, then a higher sensitivity should be sought, with an associated higher rate of false positives. False alarms would result in unnecessary medication, but present a lower risk to the patient compared to consistent AED use. Conversely, a device designed solely for seizure warning should also demand high sensitivity, but patients will be less tolerant of false alarms. A missed seizure could put a patients' life at risk (if they are driving or swimming, for instance). Jouny et al. (2011) report that patients desire a warning just before the onset of disabling symptoms - warnings hours in advance lead to periods of anxiety and diminished quality of life.

Moghim and Corne (2014) use an *S1-score* to report their findings. The s1-score is

the harmonic mean of the sensitivity and specificity, and the authors maintain that it is a fair single-value summary that simplifies discussion of results. A harmonic mean gives a high value only if both inputs are high, and a low value otherwise.

2.8 Conclusions

This chapter has critically examined the main factors currently affecting seizure prediction research. It is clear that there are many approaches that can be taken to solve the seemingly intractable seizure prediction problem. Comparative studies that reduce this uncertainty will help focus research efforts on the algorithms, features and methodological approaches that will offer the best opportunity for predicting epileptic seizures, with the high accuracy that patients demand and deserve.

Chapter 3

Design and Methodology

This chapter presents details of the research methodology, which follows a generally accepted data mining workflow. The key stages are data understanding, preprocessing, feature extraction, feature selection, data partitioning, modelling and evaluation. Note that there are multiple experiments based on different combinations of features, selection methods, sampling methods and classifiers. An overview of the experiments is provided in Section 3.6.

3.1 Overview of Methodology

The main phases of the methodology are briefly (Figure 3.1):

1. Data understanding - The ANN and SVM classifiers are supervised learners and require a labelled EEG dataset. EEG metadata is extracted from the data files, and includes information on numbers of electrodes, sampling frequency and the recording duration.

2. Data preprocessing - The data is pre-processed to standardise the electrode counts, sampling frequency and file labelling. Each data file is segmented into multiple windows for feature extraction. Four different types of windows are used, with different durations and overlaps.

3. Feature extraction - Statistical and spectral EEG features are extracted from each EEG channel, and from each segment of EEG file. Three sets of features are used: statistical features only; spectral features only; and a combination of both. The features are merged into a matrix of feature vectors to form the analytics base table, together with the class labels or *target feature*. The features

are normalised to a standard range.

4. Feature selection - The extracted EEG features are analysed using two commonly used feature selection techniques. The features are ranked according to importance and contribution to overall accuracy. The feature set is reduced to improve computation time and reduce over-fitting.

5. Data partitioning - The data is randomly split into training and testing datasets, using the holdout method. The former is used to train the classifiers and the latter is put aside to be used for final model evaluation. Three different types of partition are constructed: a normal partition; a reduced partition (which down-samples the majority class); and an increased partition (which up-samples the minority class).

6. Model training - The training datasets are applied to the SVM and ANN classifiers. Model performance is provisionally evaluated and parameter optimisation is completed using a grid search algorithm. The output produces trained and optimised SVM and ANN classification models.

7. Classification - The test data is applied to the trained classifiers, to predict if the EEG data segments are classed as interictal or preictal. A random classifier is used to baseline the performance of the ANN and SVM models. A majority voting mechanism is used to classify each EEG file.

8. Performance evaluation - The results are analysed using specificity, sensitivity and S1-scores. The results for the classifiers and different experimental options (windows, features, feature selection, sampling and classifiers) are compared.

9. Hypothesis testing - Based on the results from the performance evaluation the hypothesis may be accepted or rejected according to the empirical evidence derived from the experimental process.

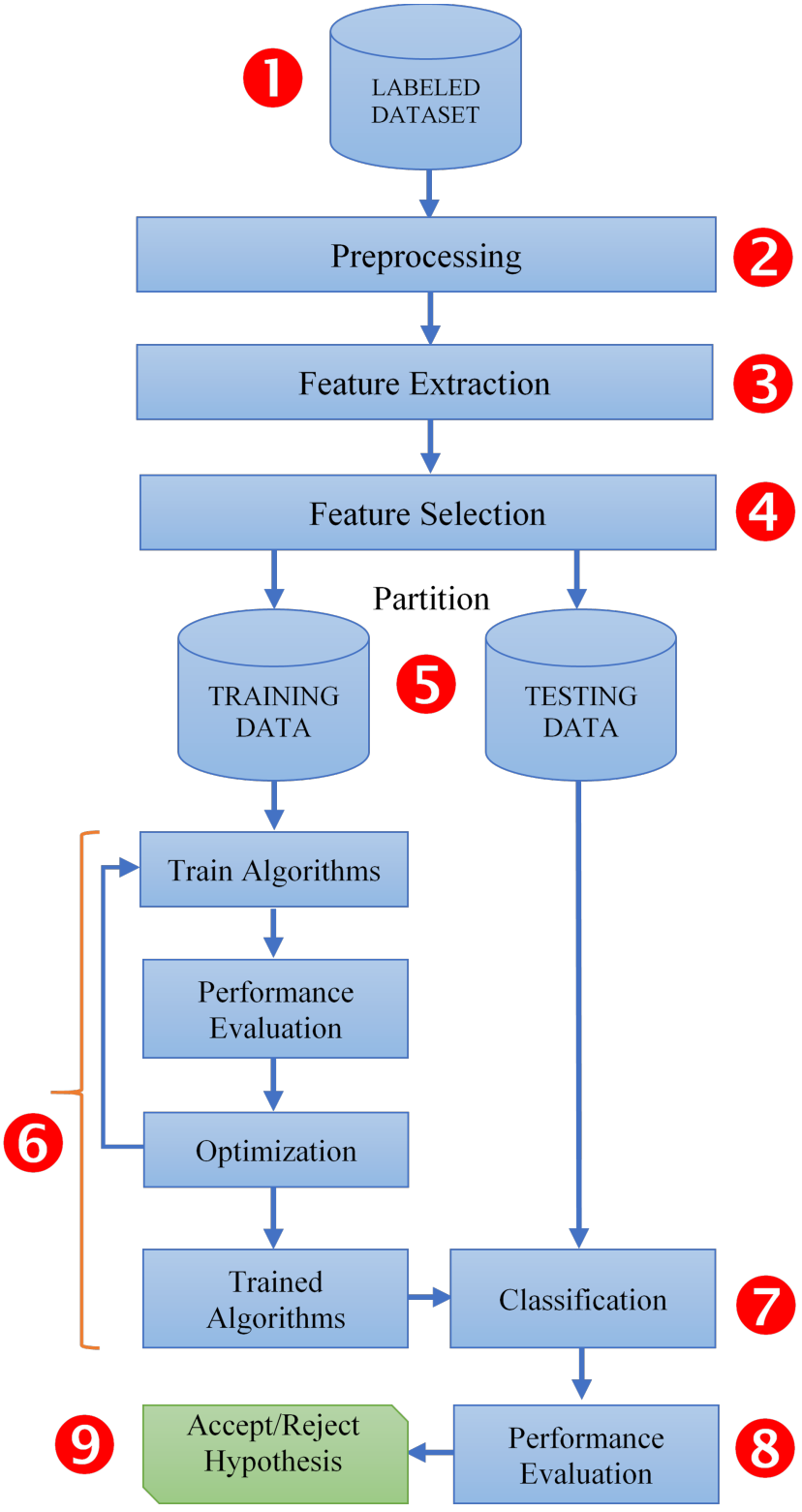

Figure 3.1. Stages of research methodology, based on supervised machine learning techniques

3.2 Data Understanding

3.2.1 EEG Dataset

This study uses the labelled EEG dataset provided by the Mayo Clinic and Mayo Systems Electrophysiology Lab (http://msel.mayo.edu/data.html), which was introduced in Section 2.3. The dataset is freely available online for download, and contains ambulatory iEEG recordings taken from five canine subjects and presurgical iEEG recordings from two human patients.

3.2.2 EEG Data Files

For the original Kaggle competition, the dataset was divided into two approximately equal sets of training and testing data. From the perspective of this study, this subdivision was considered arbitrary and was ignored thereafter. All EEG files were pooled together and new test and training partitions were constructed. The total size of the data files for both training and testing sets was roughly 120 GB.

The EEG files use the MATLAB *.mat format. Each file contains the time series iEEG data for each electrode, and metadata which describes: the number of electrodes; electrode labels; duration in seconds; sampling rate in Hz; and file sequence number indicating the file's position in the temporal sequence. The sequence number is not intended to be used as a feature for model training.

Preictal data is provided in six 10-minute segments from a 66 minute period prior to a seizure. The clips are separated by 10 second intervals and exclude the five minutes immediately prior to seizure onset, as determined by the expert reviewer. Interictal clips are also provided in sets of six 10-minute clips, with 10 second spacing between clips. The interictal files are similarly constructed, but are selected from recordings more than one week prior to any seizure event. See Figure 3.2.

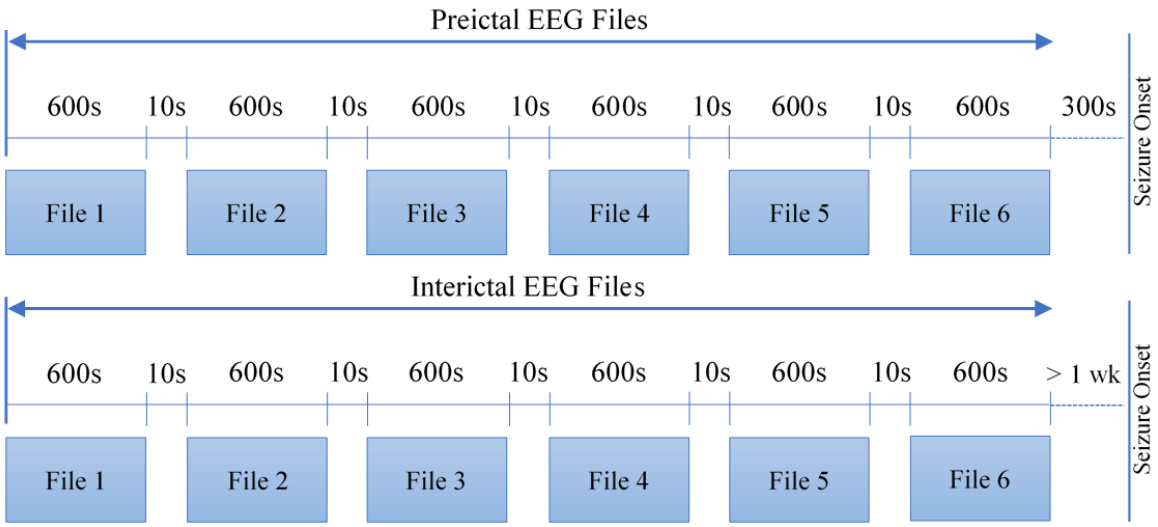

Figure 3.2. Preictal and interictal file segments

3.3 Data Preparation

3.3.1 EEG Resampling

There is some variance in the EEG sampling rate between the files in the MSEL dataset. Data files for Patient 1 and Patient 2 are sampled at 5 kHz, whereas data from the other subjects are sampled at 400 Hz. The sampling frequency affects the maximum frequency that can be extracted from the EEG signal - the *Nyquist frequency*, which is half the sampling rate - and higher sample rates increase the size of the data files substantially. EEG data sampled at 5 kHz contains 5000 discrete measurements per EEG channel, per second, compared to 400 measurements per second for a 400 Hz sample rate.

Yu et al. (2016) prepared their data by first decimating the EEG signals to 250 Hz, which they argue is a practical sample rate for potential implantable devices. For this study, the sampling rate was standardised to 400 Hz, which is sufficient to analyse frequencies up to 200 Hz. Resampling also improves data processing and computation speed in subsequent stages of the process.

3.3.2 Outlier Treatment

Outliers in EEG signals are usually caused by physiological interference, including facial movements, muscle twitches and eye blinks. Moghim and Corne (2014) consider artefacts as types of outliers that are disturbances in the signal that do not originate from the brain. Artefacts are either external (arising from exceeding a measurement signal range or disconnecting electrodes) or are internal (arising from body movements or eye blinks).

Similarly, Bandarabadi, Teixeira, et al. (2015) found that features are highly variable around a long-term average value, possibly due to noise or other artefacts such as eye blinks. To counteract this trend, they preprocessed each feature with a 1 minute moving average. Alternatively, Bandarabadi, Rasekhi, et al. (2015) treat outliers by removing values above the 98% percentile from the features. Feldwisch-Drentrup et al. (2011) used a median filter with 4-minute window to remove outliers, where each data point is replaced by the median value of the points within the window.

Due to the fragmented nature of the data files used in this study (unordered 10 minute segments), it was not possible to remove long-term temporal trends that may have been present across multiple files. However, de-trending the data may also be achieved by taking the first derivative of the raw time series data, and this study adopts this approach.

3.3.3 Electrode Selection

The EEG data files in the MSEL dataset have different numbers of electrodes for some subjects. An irregular number of electrodes would result in different numbers of features for different subjects, and would not permit a generalised study. One option is to discard the affected data files, but considering the relative sparsity of the data, it was decided to retain them. Instead, the number of electrodes is standardised to the most common value (sixteen) to resolve this issue.

As a practical matter, it is worth noting that studies that preselect electrodes must use EEG data that includes details on seizure localisation. The MSEL dataset does include this information, and consequently the full set of electrodes is analysed. Generally speaking, most studies use data from all the available EEG electrodes - the predominant view is that all channels potentially contain valuable information for prediction or detection. "Integration of all information... over spatial and temporal scales is vital to achieve accurate seizure prediction." (Karoly et al., 2016, p.1).

3.3.4 Time Series Segmentation

Following on from the discussion in Section 2.5.2, there is considerable variation in the size of the segments used in prediction studies. The duration of the window typically ranges from 5 to 60 seconds. Park et al. (2011) use 20 second overlapping windows, and Karumuri et al. (2016) divide their EEG recordings into 10 second segments, while Alexandre Teixeira et al. (2014), Moghim and Corne (2014) and Bandarabadi, Teixeira, et al. (2015) use 5 second contiguous segments. This may be contrasted with the study by Sharma (2015), which uses an EEG window of 3600 seconds and at the other extreme, studies by Kharbouch et al. (2011) and A. Shoeb et al. (2011) use windows of 1 second duration.

To address this uncertainty, the research experimented with four different windows, with varying combinations of durations and overlaps. The selected windows were: 60s contiguous; 60s 50% overlapping; 30s contiguous; and 30s 50% overlapping. These parameters were also chosen with a view to the limited time available for analysis - for instance, a 1s window generates 60 times the number of data instances of a 60s window.

3.3.5 Feature Extraction

This research uses linear univariate statistical and spectral features for seizure prediction. Whilst there is evidence that bivariate measures improve predictive performance,

the focus is on comparing classifier performance, rather than determining the most effective feature. Univariate features are used in many studies and provide a strong baseline for understanding the predictive capabilities of the classification algorithms. Additionally, univariate measures are relatively simple to derive and computationally efficient. Linear features are attractive for implantable devices because they can be calculated rapidly with low power consumption (Park et al., 2011).

A univariate time-series consists of an observation recorded sequentially and separated by equal time increments. Typical examples of this are company share prices, daily temperature changes or EEG voltage signals. With the latter, the time series is represented graphically as a sequence of voltage values plotted as a function of time over the duration of the EEG recording. Processing raw EEG data results in a set of descriptors or features, that can be analysed to determine if the data represents a non-seizure or pre-seizure state.

Statistical Features

As discussed in Section 2.5.3., statistical features are well-known and simple to calculate. The peak amplitude, mean, variance, skew and kurtosis are extracted from each EEG channel using standard statistical functions. The peak amplitude is simply the maximum EEG voltage in the time-series segment under consideration, and the *mean* is calculated from Equation 3.1, where $X_1 ... X_n$ represents the time-series electrode voltages.

$$\mu = \frac{1}{n}\sum_{i=1}^{n} X_i = \frac{X_1 + X_2 + ... + X_n}{n} \tag{3.1}$$

The *variance* is calculated from the mean using Equation 3.2.

$$\sigma^2 = \frac{\sum (X - \mu)^2}{n} \tag{3.2}$$

The mean and variance are used to derive *skew*, using Equation 3.3.

$$Skew = \frac{\mu^3}{\sigma^3} \tag{3.3}$$

Similarly, the fourth moment *kurtosis* may be calculated from Equation 3.4

$$Kurtosis = \frac{\mu^4}{\sigma^4} \tag{3.4}$$

Spectral Features

This research extracts six features based on the signal energy across six discrete frequency bands, as proposed by Howbert et al. (2014), Brinkmann et al. (2015) and Brinkmann et al. (2016). An additional six features are included based on the normalised signal energy (see Section 3.3.6). The total signal energy for the range 0-200 Hz and the frequency with the greatest signal energy are also calculated. This results in a total of 14 spectral features per EEG channel, per EEG window.

EEG frequencies are determined using a Fourier transform, which represents a signal as the sum of multiple sine and cosine functions, with varying phase and amplitude. The transform decomposes a time series signal into its individual frequency components, and results in a complex-valued function of frequency. The real component represents the magnitude of the frequencies present in the signal and the complex component represents the phase offset of the underlying sinusoidal signals.

The frequency spectrum for each EEG channel is calculated by applying a FFT to each window of time domain data. The complex DFT takes in N data points from signal $x[n]$ to produce an array $X[k]$ of length N, shown in Equation 3.5.

$$X[k] = \frac{1}{N} \sum_{n=0}^{N-1} x[n] e^{-j2\pi kn/N} \tag{3.5}$$

Each point in the array is spaced according to the *frequency resolution* Δf. To determine the frequencies, the points in $X[k]$ are converted using Equation 3.6, where f_s is the sampling frequency.

$$\Delta f = \frac{f_s}{N} \tag{3.6}$$

The resulting output of Equation 3.5 may be interpreted as follows:

- $X[1]$ = the DC frequency component
- $X[2]$ to $X[N/2 - 1]$ = the positive frequencies
- $X[N/2]$ = the Nyquist frequency (half the sample rate)
- $X[N/2 + 1]$ to $X[N]$ = the negative frequencies

For this application, the negative frequencies are a mirror image of the positive frequencies about the Nyquist frequency and can be discarded, along with the DC component. Finally, the frequency domain signal is filtered into the six frequency ranges, and the energy in each is calculated using Equation 3.7 (Moghim & Corne, 2014).

$$Energy(t, w) = \frac{1}{w}. \sum_{i=t-w}^{i=t} X(k)^2 \tag{3.7}$$

3.3.6 Data Normalisation

Normalisation or *scaling* is used to standardise the range of features before prior to model training. Some algorithms are sensitive to the data ranges, and will be overly influenced by features with a larger data range. Scaling changes the features to fall

within a specified range, while maintaining the relative differences between the features (Kelleher et al., 2015). Scaling can be done by *range normalisation* (or *rescaling*), and the feature is typically scaled to the range [0,1] or [-1,1]. This research uses range normalisation, as proposed by Rasekhi et al. (2013) in Section 2.5.4.

The research also examines the effects of normalisation on algorithm performance by using both normalised spectral features, and non-normalised spectral features. This results in are a total of six spectral energy features and six normalised spectral energy features per EEG channel, and per EEG window. Features are scaled to the range $[0, 1]$ using Equation 3.8.

$$x_{norm} = \frac{x - min(x)}{max(x) - min(x)} \tag{3.8}$$

3.3.7 Feature Selection

The most relevant features should be selected from the feature matrix, before modelling and classification can commence. The number of features represents a trade-off between computational complexity, and the discriminant capability of the feature set. Excellent predictive performance can indeed be obtained by systems using large feature sets, complex models, and involving lengthy computations. However, such systems are not practical for clinical applications, which require real-time processing, fast computation and low power consumption. A balance must be struck between the requirement for high performance and the needs of a portable low-power device.

This study uses two feature selection techniques: recursive feature elimination (RFE), and learning vector quantisation (LVQ). RFE is a wrapper method and comprises a search algorithm that takes the features as input. The algorithm selects features by recursively adding and subtracting features to search for the optimal combination that gives maximal performance. Features are then ranked according to their contribution to overall accuracy. On the downside, RFE is computationally intractable given the recursive nature of the search algorithm, and is prone to over-fitting.

LVQ is a type of artificial neural network that can be applied to classification problems. The output of the LVQ model can be used to rank the predictors according to relative importance, and a subset may be selected for final modelling. In practice, LVQ is used in conjunction with the accuracy scores from the RFE, since the LVQ's relative importance scores are uninformative in isolation.

In this research, the ANN and SVM performance is compared using the full dataset (with no feature selection); a reduced feature set using RFE; and a reduced feature set using LVQ. This evaluates the impact of feature selection in general and the relative performance of the two selection techniques. Importantly, it also identifies the most discriminatory features for seizure prediction.

3.3.8 Data Partitioning

Standard machine learning methodology dictates that data used to build and test predictive models is derived from separate datasets. Algorithms are initially trained or fitted on a subset of the data called a *training set*. The candidate models' performance is then evaluated on a *validation set* and the final model evaluation is done on a *test set*, which is wholly independent of the training and evaluation data.

The EEG dataset is divided using the holdout method, and is initially randomly split with a training/testing ratio of 70%/30%. EEG files are selected to ensure the same class distribution in each division. This follows the same methodology as Moghim and Corne (2014) and Costa et al. (2008), as described in Section 2.5.6. Where practical, k-fold validation with five folds is used during model optimisation. For some of the larger models with many data instances and features, this was limited to 2-fold cross-validation. Once model training is completed, the final model fit is evaluated using the test dataset. To ensure consistency during performance evaluation, the same EEG files are used to test all models.

3.3.9 Data Sampling

There are several studies that deal with the issue of class imbalance in EEG datasets, which either use minority upsampling or majority downsampling, as the case may be. This paper uniquely investigates the effects of three types of data sampling. Experiments are conducted with: an imbalanced dataset (the normal case); a balanced dataset (with majority down-sampling); and a balanced dataset (with minority up-sampling). Down-sampling is performed by randomly selecting a reduced number of the interictal training data instances, and up-sampling is performed using a SMOTE algorithm, which interpolates from the limited number of preictal samples. The net effect of both sampling techniques is to ensure that the final number of interictal and preictal samples in the training datasets are roughly equal. It should be noted that the techniques for dealing with class imbalance are only implemented in the training data partition, and not in the test data partition, which remains imbalanced.

3.4 Data Modelling and Classification

Predicting the onset of an epileptic seizure is a binary classification problem that attempts to classify EEG signals according to two classes: the interictal class (the normal brain state) and the preictal class (the state immediately prior to a seizure). It is worth restating that the primary objective of this research to compare the performance of SVM and ANN classifiers used to predict epileptic seizures in advance.

Models are trained on the training datasets constructed during data partitioning. The training set consists of the selected features and the target attribute, indicating if an instance is preictal or interictal. The algorithms map the input data instances to the target variable and output a fitted model that best fits the inputs.

The particular implementation of SVM and ANN used in this research permits optimisation of both classifiers using two tuning parameters. SVM may be optimised using γ and C, where γ is a kernel regularisation parameter and C sets the cost of a con-

straint violation. The ANN can be fine-tuned using a *size* and *decay* parameter, where the size represents the number of neurons in the hidden layer, and the decay adjusts the weights between iterations. This particular ANN implementation is a feedforward network with a single hidden layer.

The classifiers ouptut a class probability for each input test data instance, indicating the probability of the instance being either interictal or preictal. To confirm if the results are significant, they may be compared with the performance of a naïve chance predictor, that has no information from the EEG signal. This study adopts the approach taken by Bandarabadi, Teixeira, et al. (2015), who use a random predictor, where the probability of correctly predicting seizures follows a binomial distribution. If the performance of the SVM and ANN classifiers is superior to that of the random prediction, the results are considered significant (Feldwisch-Drentrup et al., 2011).

3.5 Performance Evaluation

The primary goal of this study is to produce optimally-trained classifiers that correctly predict all the samples in the test data, thus reproducing the target labels. For this application, every sample misclassified as preictal may be considered a false alarm and every sample misclassified as interictal may be considered a missed seizure. In accordance with best practice as described in Section 2.7.2, and in common with the majority of seizure prediction studies, this study evaluates performance using sensitivity and specificity measures. Sensitivity is defined as the percentage of preictal instances that are correctly predicted. Sensitivity is calculated using the standard definition in Equation 3.9.

$$Sensitivity = \frac{TP}{TP + FN}.100 \tag{3.9}$$

In Equation 3.9, TP is the *true positives*, which counts the number of correctly clas-

sified preictal samples, and FN is the *false negatives*, which counts the number of preictal instances that were incorrectly classified as interictal. Specificity is defined using Equation 3.10.

$$Specificity = \frac{TN}{TN + FP}.100 \tag{3.10}$$

In Equation 3.10, TN is the *true negatives*, which represents the number of correctly classified interictal samples, and FP is the *false positives*, which counts the number of interictal instances that were incorrectly classified as preictal.

Performance is also evaluated using the S1-score in Equation 3.11, which provides a single-value summary of sensitivity and specificity. The S1-score is calculated as the harmonic mean of the two values, and results in a high score only if both input values are high, and a low score otherwise.

$$S1 = 2 * \frac{Sensitivity * Specificity}{Sensitivity + Specificity} \tag{3.11}$$

3.6 List of Experiments

The primary objective of this research is a performance comparison of ANN and SVM classifiers when used to predict seizure onset. There are four secondary objectives, which investigate the impact on performance of four data preprocessing techniques. In order to fully explore the role of each of these variables individually, multiple models are developed based on combinations of the EEG window parameters; the types of features; the feature selection methods; the data sampling techniques; and the classification algorithms. In each case, the independent variable under test (EEG window, feature set, and so on) is changed to observe the effect on the dependent variable (the classifiers), while all other independent variables are held constant. This gives rise to a large number of classification models and corresponding sets of results.

The raw EEG data files are segmented using four windows, with overlapping and

contiguous windows of 60 second and 30 second durations. Two sets of statistical and spectral features are extracted from each of these four windows. The two feature sets are combined to form a third set for each window, giving twelve sets of features derived from the four windows.

Each set is reduced using RFE and LVQ, in addition to the null case, where all features are retained. The resulting thirty-six feature sets are each partitioned into training and testing sets, and the training sets are further processed by up-sampling, down-sampling or retained as is. The resulting 108 features sets are used to train a total of 216 models for the SVM and ANN classifiers. The random predictor is not trained, but is applied to the test data only. The three classifiers (including the random classifier) make predictions on the test data and generate a grand total of 324 sets of results.

In summary, there are 4 different EEG windows, 3 combinations of features, 3 different feature selection options, 3 methods of sampling and 3 classifiers. This translates to a total of $4 * 3 * 3 * 3 * 3 = 324$ sets of results. See Figure 3.3.

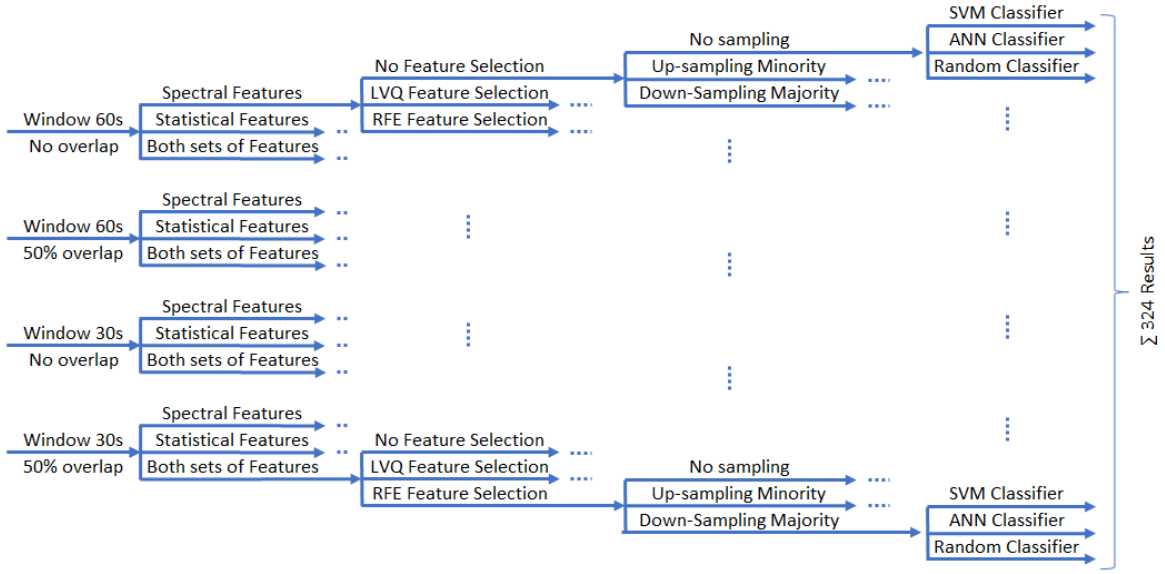

Figure 3.3. Flowchart of experimental models. There are a total of $4*3*3*3*3 = 324$ sets of results.

Chapter 4

Implementation and Results

The previous chapter gave a detailed explanation of the research methodology, and this chapter now describes how the research was implemented, based on the proposed approach. The first section describes the practical steps taken to complete the data understanding, preparation and modelling phases. Section 4.2. contains the full set of results, with a limited discussion to guide the reader.

4.1 Implementation

4.1.1 Software Tools

All data manipulation, transformation, modelling and evaluation was implemented with the R programming language. R is an open-source language licensed under the GNU Project, and is arguably the most popular programming language used in data analytics. R provides many statistical and graphical functions and is supplemented by a large number of user-created packages.

The Matlab EEG files were handled using the R.matlab package. The e1071 package was used for SVM modelling and classification. The neural network classifier was implemented with the nnet package. Optimisation with grid search was done using the caret package. Classifier performance was analysed using the ROCR package (Sing, Sander, Beerenwinkel, & Lengauer, 2005). The DMwR package was used to implement the SMOTE algorithm and all visualisations used the plotly package. All R script files are available on GitHub at https://github.com/IanTW/Epilepsy.

4.1.2 Data Understanding

EEG Metadata

The EEG files' metadata was extracted to present a summary of the EEG raw data. The results of this analysis are shown in Table 4.1. There were a total of 7,999 EEG files from seven subjects, and all files had a duration of 600 seconds. There were variances in the sampling frequency and electrode numbers between subjects. Two of the subjects' EEG recordings were sampled at 5 kHz, while the rest were sampled at 400 Hz. With respect to electrode counts, four of the subjects' data had 16 electrodes, one had 15 electrodes and two had 24 electrodes.

Refer to Table 4.2 for a breakdown of the EEG data files, per subject. The data was divided into testing and training sets for the 2014 Kaggle seizure prediction competition. The number of interictal and preictal files clearly shows the class imbalance in the dataset. In total there are 7447 interictal files and 552 preictal file, a ratio of 13.5. The preictal files comprise approximately 7.5% of the total.

Table 4.1

Metadata from Mayo Systems Electrophysiology Lab EEG Files

Subject	Sample Rate (Hz)	Channels	Seizures	File Duration (s)	Tot. Files
Dog 1	400	16	8	600	1006
Dog 2	400	16	40	600	1542
Dog 3	400	16	18	600	2416
Dog 4	400	16	27	600	1891
Dog 5	400	15	8	600	671
Patient 1	5000	15	4	600	263
Patient 2	5000	24	6	600	210

Table 4.2

Summary of Mayo Systems Electrophysiology Lab EEG Data Files

Subject	Training		Testing		Total
	Interictal Files	Preictal Files	Interictal Files	Preictal Files	
Dog 1	480	24	478	24	1006
Dog 2	500	42	910	90	1542
Dog 3	1440	72	862	42	2416
Dog 4	804	97	933	57	1891
Dog 5	450	30	179	12	671
Patient 1	50	18	183	12	263
Patient 2	42	18	136	14	210

4.1.3 Data Preparation

Resampling

The 5 kHz data files were resampled to 400 Hz using standard signal processing techniques. This standardised the data and reduced computation time during feature extraction. The files from Patient 1 and Patient 2 were converted to 400 Hz, and saved for subsequent processing and feature extraction.

Electrode Selection

The number of EEG channels was standardised to 16 channels per file. A sixteenth channel was imputed for each file from Dog 5 and Patient 1, using the mean of the other fifteen channels in the file. In the case of Patient 2, which had 24 channels, 16 channels were randomly selected for inclusion in the study. Since it was not possible to determine the most discriminatory channels in advance, a random channel selection process was considered suitable.

EEG File Segmentation

The experiments used four different windows, which varied the size of the base tables thus generated. A 600 second EEG file segmented into contiguous 60 second windows resulted in a total of 10 windows, and similarly a contiguous 30 second window resulted in 20 windows. For a specific duration, if X is the number of contiguous windows, then the number of overlapping windows Y may be determined from $Y = 2X - 1$. Features from one window of data formed one row in the base table. See Table 4.3 for details of the window configurations and the base table row counts, using a total of 7,999 EEG files.

Table 4.3

EEG File Segments Used in Experiments

Segment Duration (s)	Segment Overlap (%)	Segments per File	Total Rows
60	0	10	79,990
60	50	19	151,981
30	0	20	159,980
30	50	39	311,961

The segmentation process is shown in Figure 4.1. EEG files are contained in a folder for each subject, and an outer code loop steps through each folder. An inner loop opens each EEG file, and extracts the time-series data into a matrix object. The matrix is sliced into multiple windows, using simple indexing.

Feature Extraction

Each window of data was processed individually in turn, and an inner loop iterated through all the channels for feature extraction. Using the equations discussed in Section 3.3.5, statistical and spectral features were calculated for each channel. Once all channels for a particular segment had been parsed, the features were bound into a single list which was written to a feature vector and the next window was processed.

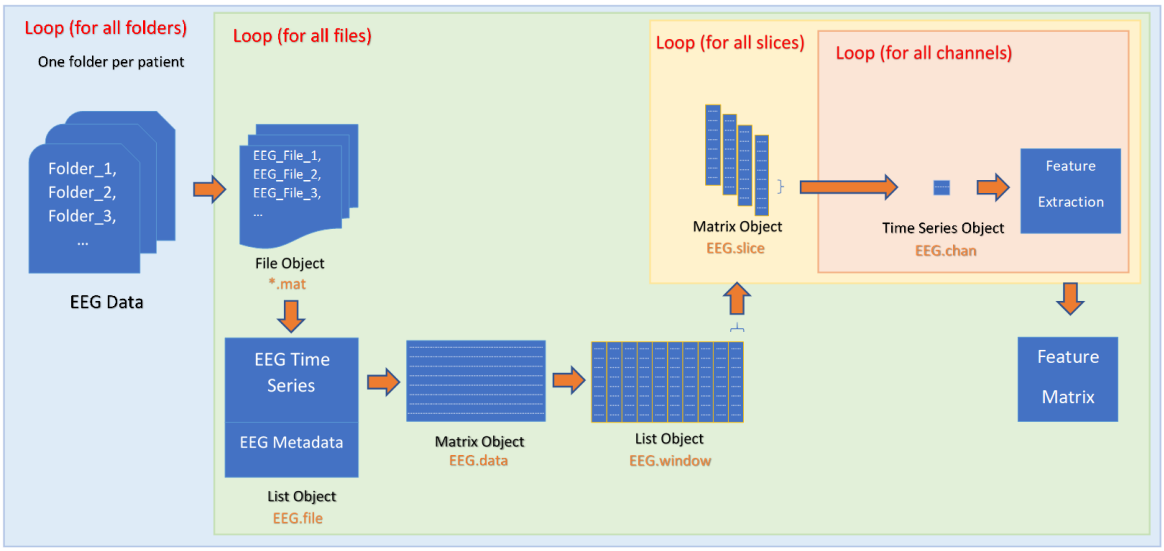

Figure 4.1. EEG file segmentation process. Four loops are used to process each folder, file, segment and EEG channel in order to extract the features. The features from one file segment comprise one row in the analytics base table.

A feature matrix was formed from the individual feature vectors, where each row represented the features taken from one window of 16 channels.

The experiments used three sets of features; 5 univariate statistical features; 14 univariate spectral features; and a combination of both the preceding statistical and spectral features. Features were extracted from 16 EEG channels: the statistics feature set thus had a total of 80 features ($16 * 5$); the spectral feature set had a total of 224 features ($16 * 14$); and the combined feature set had a grand total of 304 features ($16 * 19$) per EEG window.

The analytics base table included a file identifier, the target attribute (preictal or interictal), a window identifier and the features. The base table dimensions were determined by the window type (rows) and feature set (columns) - the largest table had 304 features and 311,961 rows. A total of twelve base tables were constructed, based on the combinations of four windows and three sets of features. Refer again to Figure 3.3 in Section 3.6.

Feature Selection

The features were initially explored using a collinearity study to determine any closely correlated predictors. The study showed large numbers of highly correlated features, which were removed and the remaining features were used for model training. However, this approach led to very poor predictive performance and was therefore discontinued. Multi-channel EEG signals are highly correlated by their nature, and removing so many features potentially compromised important spatial information.

Each base table was reduced using two feature selection techniques (RFE and LVQ). For comparison purposes, the complete base table was also retained. The classification accuracy of the features was determined by plotting the outputs of the RFE models. Refer to Figure 4.2, which shows the accuracy of the features using a 60 second, 50% overlap window. The statistics feature set has optimal accuracy with 21 features, shown by the flattened curve. Similarly, the spectral and combined feature sets show the best performance at 132 and 270 features respectively, which clearly shows the predictive value of features derived from multiple EEG electrodes. A similar analysis of the features generated by the other windows produced the similarly shaped curves, with differing optimal feature counts.

For all window options, the accuracy curves roll-off and plateau at a feature count of approximately 20 (for statistical features) and 50 (for spectral and combined features) respectively. Additional features do not appreciably add to overall accuracy, but do significantly add to computation time and model complexity. Accordingly, the 20 top-ranked features from the statistics set and the 50 top-ranked features from the other two were selected from the relevant base tables. See Figure 4.3.

The LVQ selection method uses a neural network classifier to rank features by their relative importance. See Figure 4.4. Adopting a similar approach to the RFE method, the 20 top-ranked and 50 top-ranked features were selected from the feature sets. The output of the feature selection stage produced a total of 36 base tables based on the experimental combinations of windows, features and selection methods.

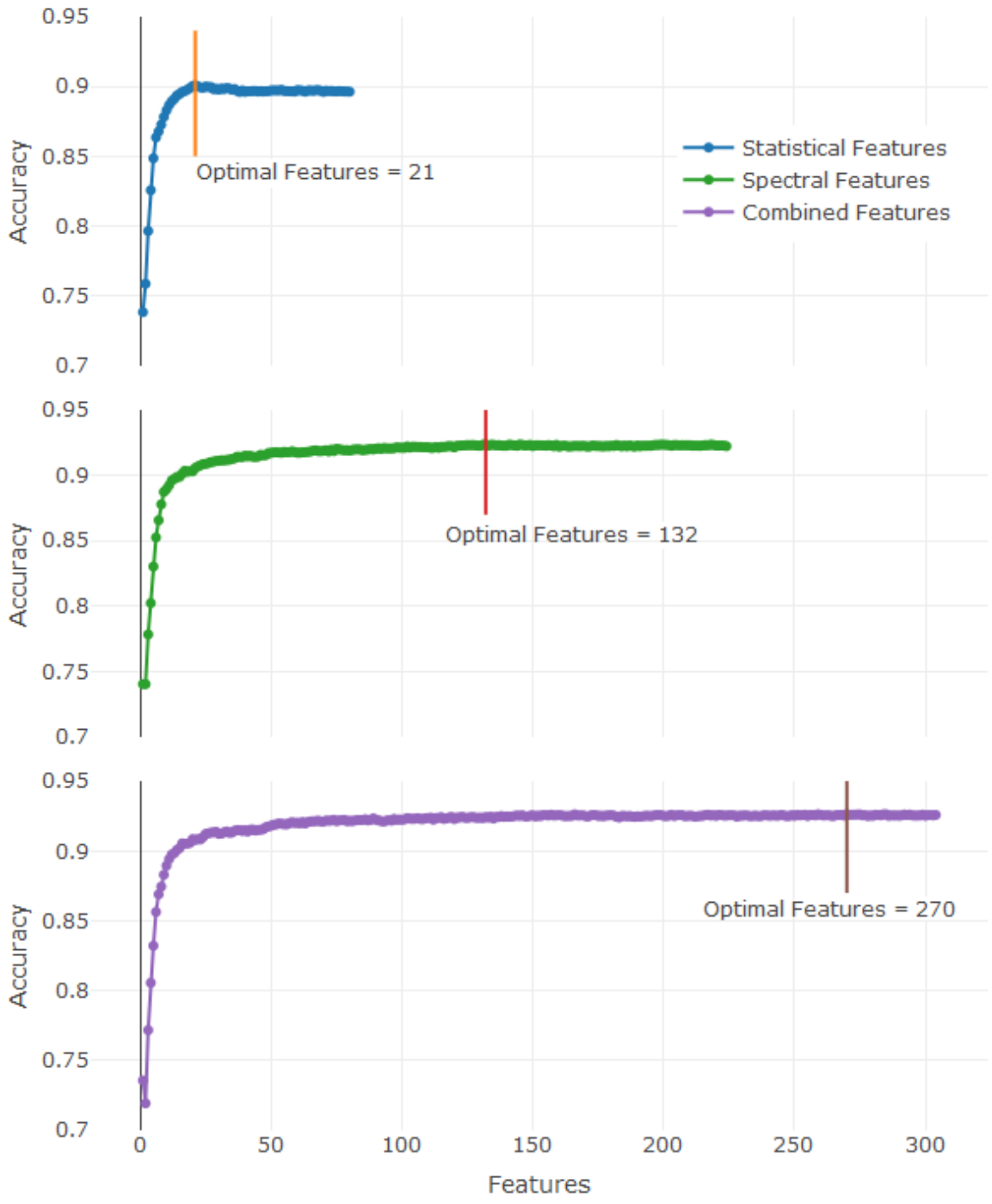

Figure 4.2. RFE models showing classification accuracy versus feature count, for features extracted using the 60s window with 50% overlap. The optimal number of features giving maximum accuracy is indicated by the vertical lines.

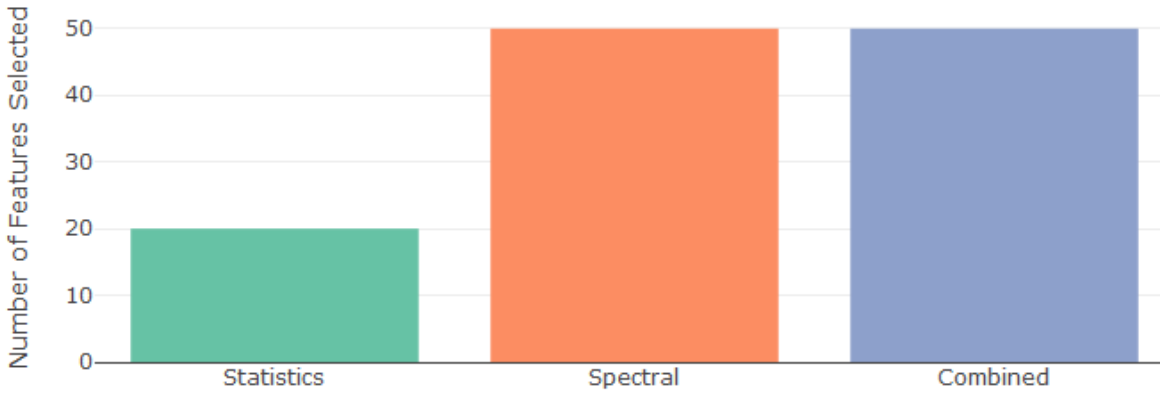

Figure 4.3. Number of features selected from each feature set

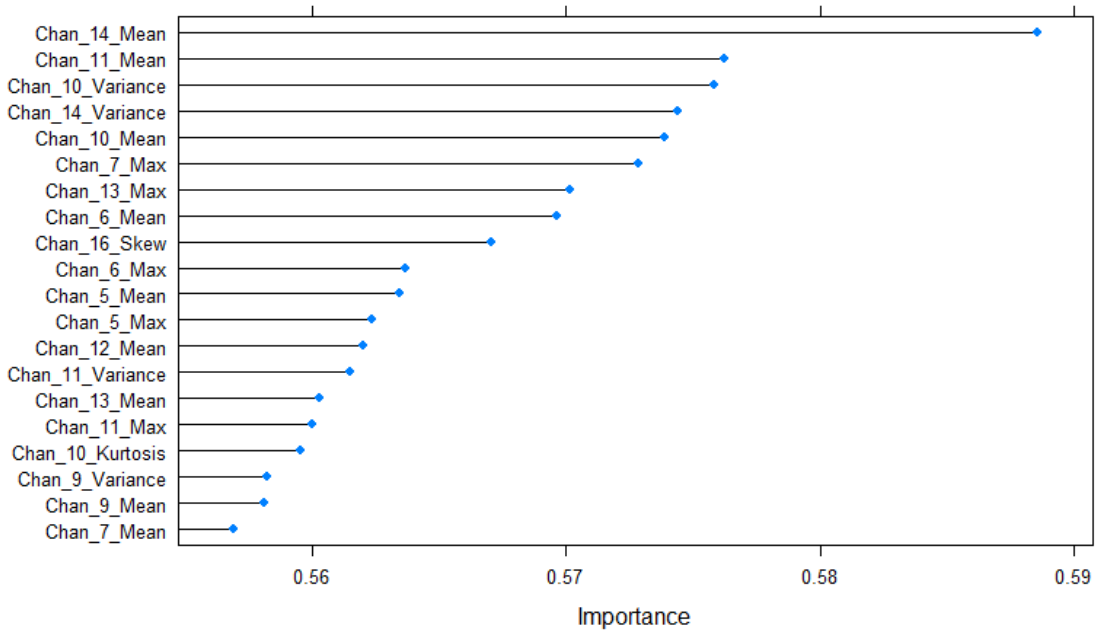

Figure 4.4. Output of LVQ model showing relative importance of features. Displays the top 20 *statistics* features extracted using a window of 60s and 50% overlap.

Data Sampling and Partitioning

Following feature selection, the base tables were randomly split 70/30 into training and testing partitions, using a set random seed to ensure consistency. The data in the training partitions were highly imbalanced, as shown in Table 4.2. The interictal samples were thus considered the *majority class* and the preictal samples the *minority class*. The 36 training sets from the previous stages were retained without any sampling treatment; processed using random down-sampling of the majority class; and processed by up-sampling the minority class with a SMOTE algorithm. Data sampling thus produced a grand total of 108 base tables for model training.

4.1.4 Optimisation and Modelling

The algorithm's tuning parameters were selected using a grid search, which used k-fold cross validation and combinations of model parameters to search for the optimal values. However, initial optimisation attempts showed that it was not practical to optimise all 216 models individually. As an example, a 10-fold cross validation grid search with two parameters, each with 4 values, will iterate $10 * 4 * 4 = 160$ times. During optimisation some of the larger models took multiple days to complete a single iteration, using a compute-optimised Amazon Web Service (AWS) virtual machine. Accordingly, the SVM models used the default values of $\gamma = 1$ and $C = 1$. With respect to the ANN models, *size* was set to half the number of features in the training set; *decay* was set to 0.1; and iterations were limited to 1000.

4.2 Results

This section discusses the key results from the experiments, and compares the classifiers' performance with respect to the various processing techniques under review. A total of 216 predictions were made on the test data and the results were ranked by sensitivity, and then by specificity. The results are listed in Appendix A.

4.2.1 Comparison of Classifier Performance

The top 28 models, and 38 out of the top 50 models were ANNs. The maximum sensitivity was 0.98 and 0.87 for ANN and SVM respectively, and the mean sensitivity was 0.6843 and 0.6271. As anticipated, the random predictor showed a tight distribution around a score of 0.5. See Table 4.4 and Figure 4.5. The differences between the ANN and SVM sensitivity scores were assessed using Wilcoxon's rank-sum test, which is a non-parametric version of the independent t-test. This test was selected because the results are not normally distributed. Sensitivity scores for the ANN classifier ($Mdn = 0.8$) differed significantly from the SVM classifier ($Mdn = 0.8$), with $W = 6800$, $p = 0.035$, $r = -0.143$. See Tables 4.5 and 4.6.

Table 4.4

Statistics Showing ANN and SVM Classifer Performance

Class.	Measure	Minimum	1st Qu.	Med.	Mean	3rd Qu.	Maximum
ANN	Sensitivity	0.0100	0.5275	0.8000	0.6843	0.8725	0.9800
SVM	Sensitivity	0.0100	0.4600	0.8000	0.6271	0.8300	0.8700
ANN	Specificity	0.6100	0.7975	0.9400	0.8869	1.0000	1.0000
SVM	Specificity	0.5600	0.7250	0.9000	0.8419	0.9900	1.0000
ANN	S1-Score	0.0200	0.6875	0.8100	0.7107	0.8700	0.9000
SVM	S1-Score	0.0200	0.6300	0.7400	0.6450	0.8200	0.9100

The ANN and SVM sensitivity scores were also compared to those from the random predictor, using a Kruskal-Wallis test to check for statistical significance. The results showed that sensitivity scores differed significantly according to classifier, $H(2) = 52.92$, $p < 0.001$. *Post hoc* tests further showed that sensitivity scores were significantly different when using an ANN classifier (difference = 89.73) compared to the random predictor. Similarly, sensitivity scores were significantly different for the SVM classifier (difference= 64.97) compared to the random predictor. In both cases, the critical difference ($\alpha = 0.05$, corrected for the number of tests) was 28.57. Thus, the SVM and ANN classifiers have sensitivity scores that are significantly higher compared to the random predictor. In conclusion, the ANN classifier clearly demonstrated superior sensitivity compared with the SVM. Furthermore, both ANN and SVM demonstrated performance well above that of a random classifier.

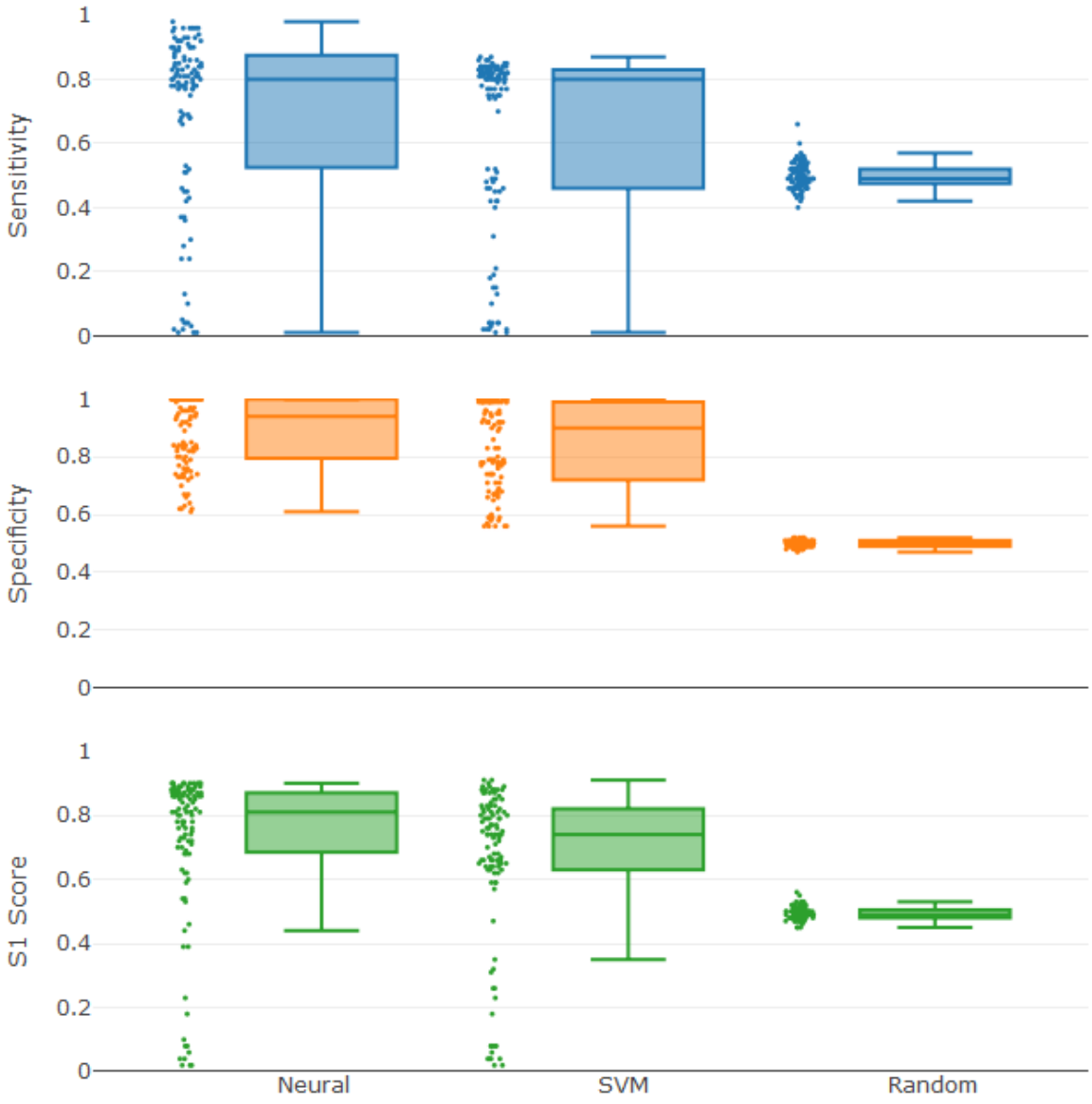

Figure 4.5. Distribution of sensitivity, specificity and S1-score by classifier. On aggregate, the ANN outperforms the SVM classifier on every measure. The results for both ANN and SVM classifiers are superior to the random predictor and may be considered significant.

Table 4.5

Wilcoxon Rank-Sum Tests for Classifier Performance

Measure	Median (ANN)	Median (SVM)	W Score	Significance	Size
Sensitivity	0.80	0.80	6800	0.035	-0.143
Specificity	0.94	0.90	7057	0.0069	-0.184
S1 Score	0.81	0.74	7326	0.0011	-0.221

Table 4.6

Kruskal-Wallis Tests for Classifier Performance

Measure	Comparison	H Score	df	Sig.	Difference	Crit. Difference
Sensitivity	Random-Neural	52.30	2	<0.001	89.73	28.57
Sensitivity	Random-SVM	52.30	2	<0.001	64.97	28.57
Specificity	Random-Neural	220.8	2	<0.001	173.3	28.57
Specificity	Random-SVM	220.8	2	<0.001	150.3	28.57
S1 Score	Random-Neural	102.1	2	<0.001	123.8	28.57
S1 Score	Random-SVM	102.1	2	<0.001	92.31	28.57

Specificity measures the classifiers' ability to correctly predict the non-seizure state. Ranking the results by specificity and then by sensitivity showed that the top 16 models and 36 of the top 50 models are ANNs. Both classifiers had a maximum specificity of 1.0000 and a mean specificity of 0.8869 and 0.8419 for ANN and SVM respectively. Again, the random predictor showed a tight clustering around a value of 0.5. The results of a Wilcoxon's rank-sum test showed that specificity scores for the ANN classifier ($Mdn = 0.94$) differed significantly from the SVM classifier ($Mdn = 0.90$), with $W = 7057$, $p = 0.0069$, $r = -0.184$.

A Kruskal-Wallis test compared the ANN and SVM specificity scores to those from the random predictor, and the results showed that specificity scores differed significantly according to classifier, $H(2) = 220.8$, $p < 0.001$. *Post hoc* tests further showed that specificity scores were significantly different for the ANN classifier (difference = 173.3) compared to the random predictor. Similarly, the specificity scores for the SVM classifier were significantly different (difference = 150.7) compared to the random predictor. For both cases, the critical difference (α=0.05, corrected for the number of tests) was 28.57. Therefore, the ANN and SVM classifiers have specificity scores that are significantly higher compared to the random predictor. In conclusion, the ANN

demonstrated superior specificity compared to the SVM and furthermore, both ANN and SVM performed well above the levels of a random predictor.

Ranking the S1-scores reiterated the dominance of the ANN classifier. Although the top 2 models were SVM, 34 of the top 50 models were ANNs. The maximum S1-score was 0.9100 and 0.9000 for the SVM and ANN classifiers respectively and the mean was 0.6540 and 0.7107. A Wilcoxon's rank-sum test compared the S1 scores between ANN and SVM, and a Kruskal-Wallis test compared their performance against the random predictor. The results show the S1-score for the ANN are significantly different from SVM and that both ANN and SVM had superior performance relative to the random predictor.

Finally, the feasibility of combining the best ANN and SVM models into an ensemble model was explored. Ensemble models can improve predictive performance by taking the output of multiple models and using a majority voting mechanism for final classification. However, the underlying base learners should be both accurate and diverse in order to improve ensemble accuracy. Calculating the Pearson's correlation coefficient between the sensitivity scores for the ANN and SVM models showed that the two measures are almost perfectly correlated ($\rho = 0.97$). As a result, the ensemble model was not pursued any further.

4.2.2 Influence of Processing Methods on Performance

Time Series Segmentation

Assessment of the results according to window type showed very little variance in performance. See Table 4.7 and Figure 4.6. Models using the 60s 50% overlapping window had marginally better maximum sensitivity of 0.98, while the others had a maximum of 0.96. The 60s 50% window also had the highest mean sensitivity of 0.6088, while the 30s 0% window had the lowest score of 0.5957.

Similarly, the specificity and S1 scores show marginal differences in performance. The

highest mean specificity score was shared by both the 60s 50% and 30s 50% windows. The former had the highest mean S1 score and the latter had the second-highest. Kruskal-Wallis tests showed that the differences in sensitivity, specificity and S1 scores were not significant. See Table 4.8. The windowing parameters chosen for segmenting the EEG time-series thus showed no clear performance advantage for any particular configuration.

Table 4.7

Statistics Showing Influence of Segmentation on Performance

Window	Measure	Min.	1st Qu.	Median	Mean	3rd Qu.	Maximum
30s 0%	Sensitivity	0.0100	0.4600	0.5300	0.5957	0.8200	0.9600
30s 50%	Sensitivity	0.0100	0.4700	0.5500	0.6073	0.8400	0.9600
60s 0%	Sensitivity	0.0100	0.4900	0.5600	0.6004	0.8100	0.9600
60s 50%	Sensitivity	0.0200	0.4800	0.5500	0.6088	0.8100	0.9800
30s 0%	Specificity	0.4800	0.5100	0.7700	0.7423	0.9600	1.0000
30s 50%	Specificity	0.4700	0.5100	0.7400	0.7444	0.9600	1.0000
60s 0%	Specificity	0.4800	0.5000	0.7500	0.7377	0.9700	1.0000
60s 50%	Specificity	0.4800	0.5100	0.7600	0.7444	0.9600	1.0000
30s 0%	S1 Score	0.0200	0.4800	0.6300	0.6116	0.8100	0.9000
30s 50%	S1 Score	0.0200	0.5000	0.6300	0.6195	0.8100	0.9100
60s 0%	S1 Score	0.0200	0.4900	0.6400	0.6117	0.8000	0.9000
60s 50%	S1 Score	0.0400	0.4900	0.6600	0.6247	0.8200	0.9000

Table 4.8

Kruskal-Wallis Tests for Segmentation Performance

Measure	Comparison	H Score	df	Sig.	Difference	Crit. Difference
Sensitivity	60s 50%-60s 0%	0.737	3	0.865	7.028	28.79
Sensitivity	60s 50%-30s 50%	0.737	3	0.865	3.019	28.79
Sensitivity	60s 50%-30s 0%	0.737	3	0.865	1.620	28.79
Specificity	60s 50%-60s 0%	0.186	3	0.980	4.241	28.79
Specificity	60s 50%-30s 50%	0.186	3	0.980	0.3704	28.79
Specificity	60s 50%-30s 0%	0.186	3	0.980	1.167	28.79
S1 Score	60s 50%-60s 0%	0.665	3	0.881	8.796	28.79
S1 Score	60s 50%-30s 50%	0.665	3	0.881	0.8889	28.79
S1 Score	60s 50%-30s 0%	0.665	3	0.881	4.463	28.79

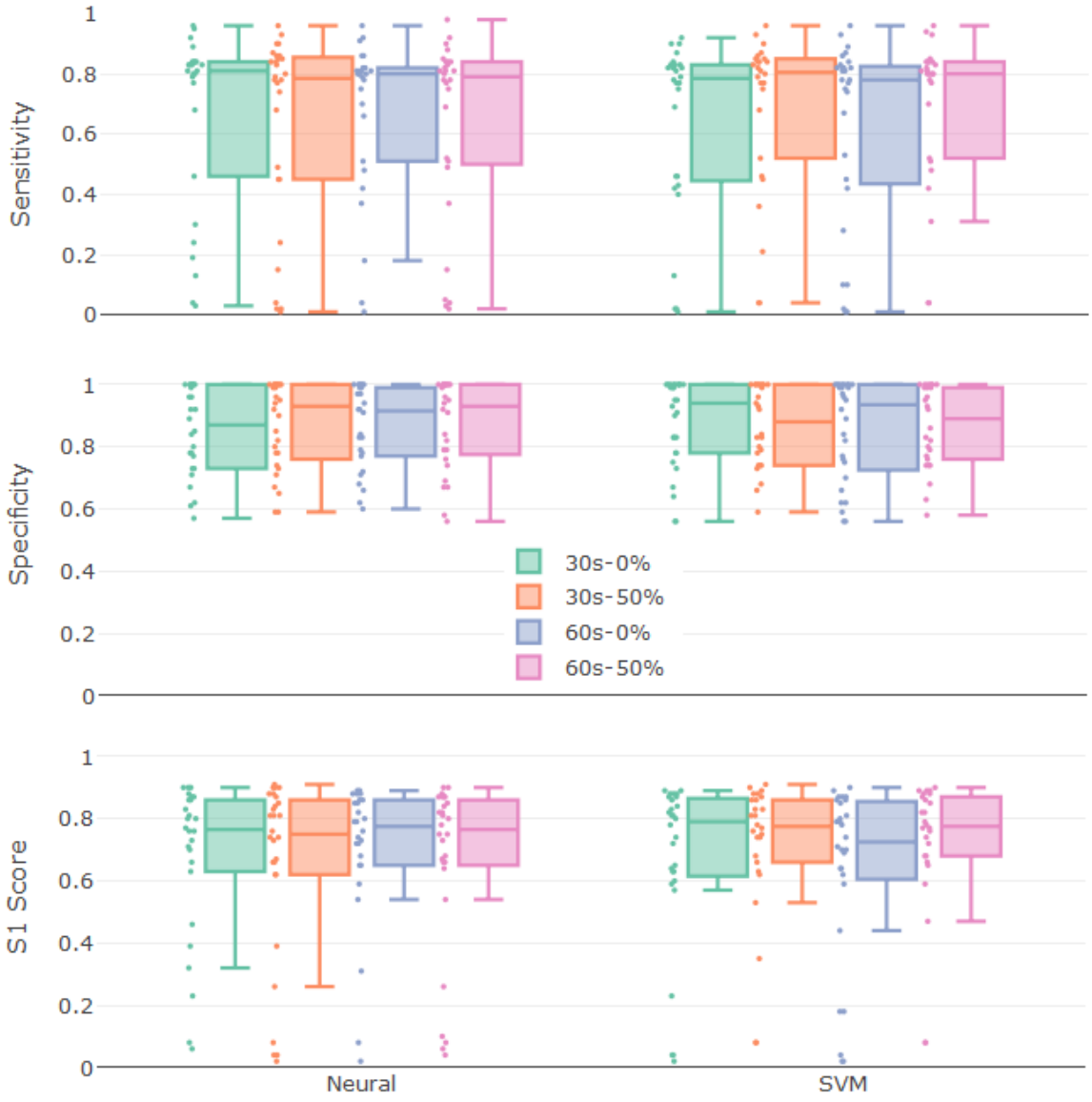

Figure 4.6. Distribution of sensitivity, specificity and S1-score by window method. Differences in performance for all the segmentation options are not statistically significant.

EEG Feature Set

The top 3 models used the combined feature sets, which had a maximum sensitivity score of 0.98. The spectral features performed well and had a maximum sensitivity of 0.96, while the statistics features had the lowest score of 0.9. The spectral features had a mean sensitivity of 0.6439, marginally better than the combined set at 0.6368. The statistics set performed poorly with a mean sensitivity of 0.5284, which scored just above chance level. See Table 4.9 and Figure 4.7.

The distribution of sensitivity scores for the statistics features showed clusters of points at around 0.8 and zero, which accounted for the poor mean value. However, the statistics features' 3rd quartile score is 0.8, compared to 0.83 and 0.84 for the spectral and combined sets respectively. The worst results for the statistics features were all coincident with models that have unbalanced data sets. See Figure 4.10.

Table 4.9

Statistics Showing Influence of Feature Set on Performance

Feature	Measure	Min.	1st Qu.	Median	Mean	3rd Qu.	Maximum
Statistics	Sensitivity	0.0100	0.4475	0.5300	0.5284	0.8000	0.9000
Spectral	Sensitivity	0.1800	0.4900	0.5550	0.6439	0.8300	0.9600
Both	Sensitivity	0.1000	0.4800	0.5850	0.6368	0.8400	0.9800
Statistics	Specificity	0.4800	0.5075	0.6500	0.6876	0.7600	1.0000
Spectral	Specificity	0.4800	0.5100	0.8200	0.7736	0.9900	1.0000
Both	Specificity	0.4700	0.5000	0.8300	0.7655	0.9750	1.0000
Statistics	S1 Score	0.0200	0.4700	0.5150	0.5039	0.7300	0.8600
Spectral	S1 Score	0.3100	0.5000	0.6800	0.6801	0.8600	0.9100
Both	S1 Score	0.1800	0.4900	0.6750	0.6667	0.8525	0.9100

The statistical significance of each feature set's performance was further examined using Kruskal-Wallis tests. See Table 4.10. The difference in sensitivities of the spectral and combined feature sets compared to the statistics sets were significant ($p < 0.001$). However, *post hoc* tests showed the difference between the spectral and combined sets were not statistically significant (difference = 3.632, critical difference = 24.94).

In respect of specificity, the spectral set had the highest mean score of 0.7736, followed by the combined features at 0.7655, and the statistics set at 0.6876. The difference

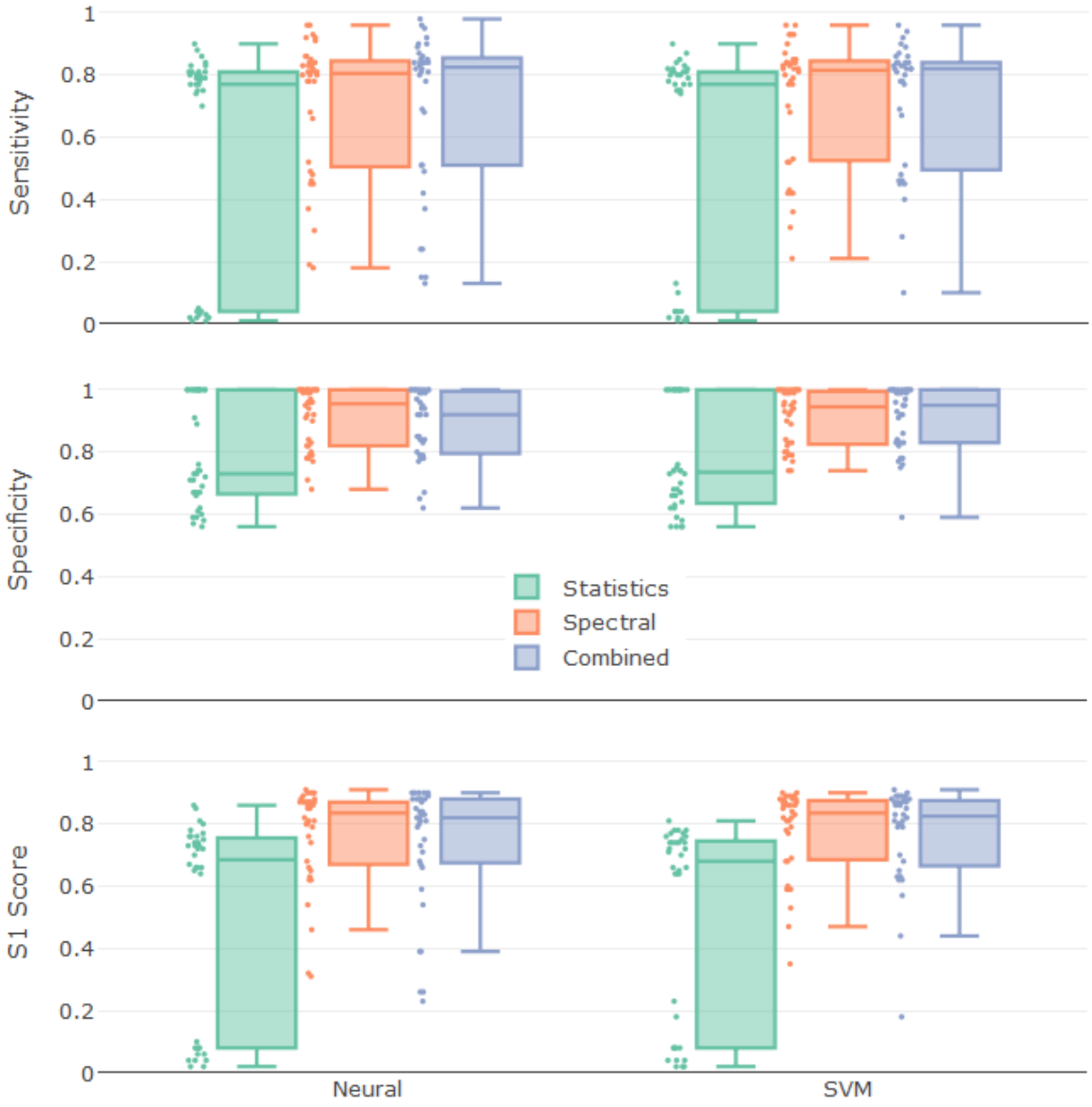

Figure 4.7. Distribution of sensitivity, specificity and S1-score by feature set. The statistics set is particularly affected by the class imbalance in the dataset. The difference in performance between the spectral and combined datasets is not statistically significant.

Table 4.10

Kruskal-Wallis Tests for Feature Set Performance

Measure	Comparison	H Score	df	Sig.	Difference	Crit. Diff.
Sensitivity	Statistic-Spectral	18.2	2	< 0.001	36.51	24.94
Sensitivity	Statistic-Both	18.2	2	< 0.001	40.14	24.94
Sensitivity	Spectral-Both	18.2	2	< 0.001	3.632	24.94
Specificity	Statistic-Spectral	20.3	2	< 0.001	41.89	24.94
Specificity	Statistic-Both	20.3	2	< 0.001	37.99	24.94
Specificity	Spectral-Both	20.3	2	< 0.001	3.903	24.94
S1 Score	Statistic-Spectral	51.13	2	< 0.001	65.69	24.94
S1 Score	Statistic-Both	51.13	2	< 0.001	63.17	24.94
S1 Score	Spectral-Both	51.13	2	< 0.001	2.521	24.94

in specificities of the spectral and combined feature sets, compared to the statistics sets, were significant ($p < 0.001$). The differences in specificity between spectral and combined sets were not significant (difference = 3.903 and critical difference = 24.94).

The differences in the S1 scores for both the spectral and combined sets were significant ($p < 0.001$), compared to the statistics set. However, the difference in the scores between the spectral and combined sets were not significant (difference = 2.521 and critical difference = 24.94). To conclude, the statistics feature set performed the worst, and was severely impacted by any imbalance in the classes. Whilst the spectral and combined sets demonstrated statistically significant performance relative to the statistics set, the differences between them were not significant.

Feature Selection Method

The top 8 models ranked by sensitivity used all the features, with a maximum sensitivity of 0.98 and a mean sensitivity of 0.6320, compared to a mean of 0.6072 and 0.5698 for RFE and LVQ respectively. See Table 4.11 and Figure 4.8. The differences in mean specificity were marginal and the best score of 0.7851 was obtained using the full feature set, closely followed by RFE and LVQ with 0.7413 and 0.7273 respectively.

A Kruskal-Wallis test established that the differences in sensitivity ($p = 0.414$) and specificity ($p = 0.238$) were not significant. However, the differences in S1 scores ($p = 0.00263$) were statistically significant. *Post hoc* testing showed this was accounted for by the difference between the results from the full data set and the LVQ method (difference = 35.76 and critical difference = 24.94). The full data set without selection thus showed a statistically significant performance improvement compared to the LVQ method. See Table 4.12.

Table 4.11

Statistics Showing Influence of Selection Method on Performance

Selection	Measure	Min.	1st Qu.	Median	Mean	3rd Qu.	Maximum.
None	Sensitivity	0.0200	0.4900	0.6750	0.6320	0.8200	0.9800
LVQ	Sensitivity	0.0100	0.4400	0.5350	0.5698	0.8100	0.9300
RFE	Sensitivity	0.0100	0.4800	0.5300	0.6072	0.8200	0.9400
None	Specificity	0.4700	0.5100	0.7850	0.7581	0.9900	1.0000
LVQ	Specificity	0.4800	0.5100	0.7150	0.7273	0.9325	1.0000
RFE	Specificity	0.4800	0.5000	0.7750	0.7413	0.9600	1.0000
None	S1 Score	0.0400	0.5000	0.6750	0.6515	0.8300	0.9100
LVQ	S1 Score	0.0200	0.4800	0.5300	0.5758	0.7625	0.9000
RFE	S1 Score	0.0200	0.5000	0.6400	0.6233	0.8100	0.9000

Table 4.12

Kruskal-Wallis Tests for Feature Selection Performance

Measure	Comparison	H Score	df	Sig.	Difference	Crit. Diff.
Sensitivity	None-LVQ	1.77	2	0.414	12.95	24.94
Sensitivity	None-RFE	1.77	2	0.414	2.264	24.94
Sensitivity	LVQ-RFE	1.77	2	0.414	10.68	24.94
Specificity	None-LVQ	2.87	2	0.238	17.03	24.94
Specificity	None-RFE	2.87	2	0.238	11.74	24.94
Specificity	LVQ-RFE	2.87	2	0.238	5.285	24.94
S1 Score	None-LVQ	11.9	2	0.00263	35.76	24.94
S1 Score	None-RFE	11.9	2	0.00263	20.38	24.94
S1 Score	LVQ-RFE	11.9	2	0.00263	15.38	24.94

A frequency count of the features selected from each set was used to rank the features according to their discriminatory power. See Figure 4.9. The most commonly selected feature from the statistics set was the mean. The 2nd, 3rd and 4th moments were selected less frequently and showed a diminishing relevance the higher the mo-

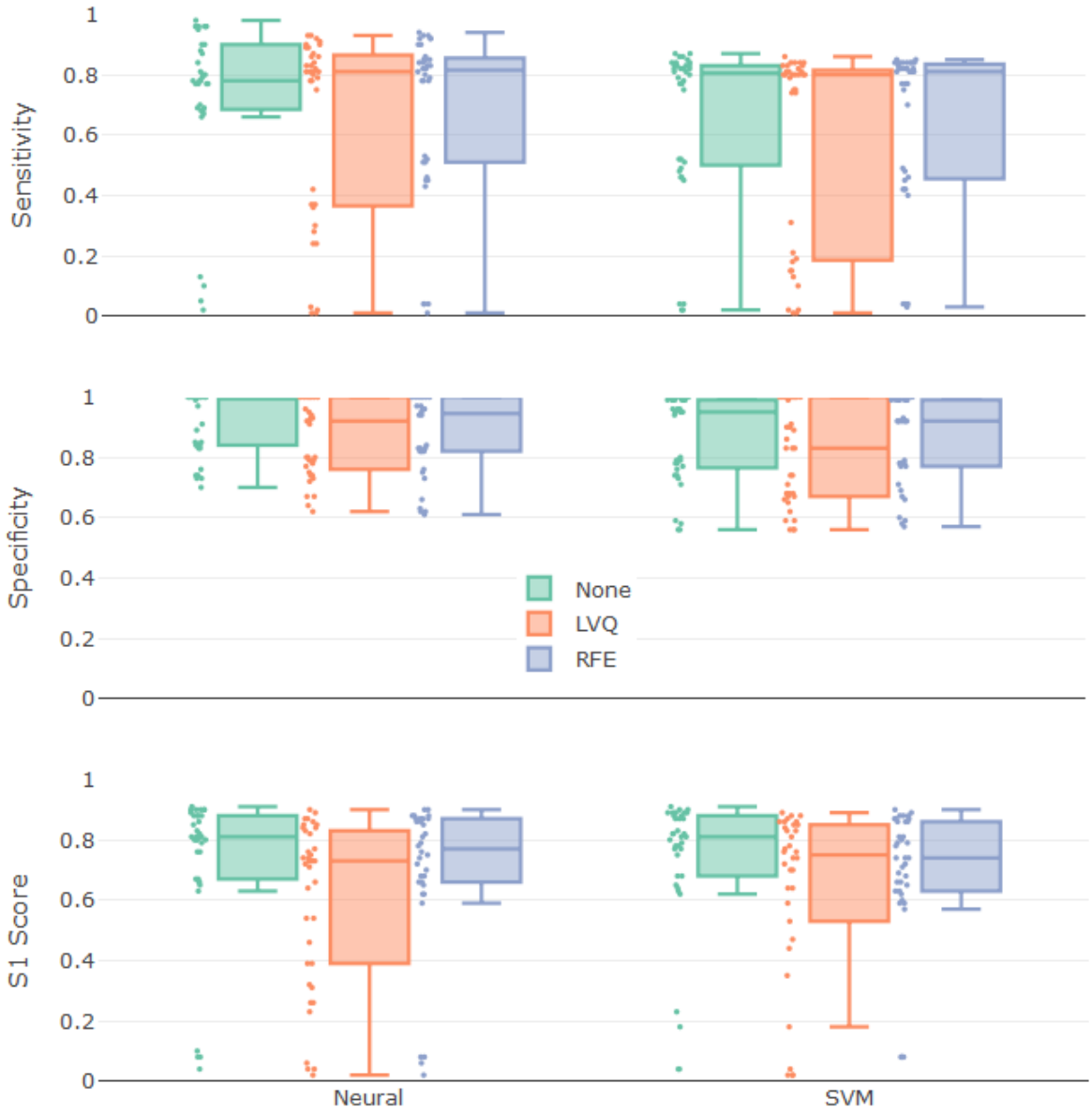

Figure 4.8. Distribution of sensitivity, specificity and S1-score by selection method. Differences in sensitivity and specificity scores are not statistically significant, but the differences in S1 scores between the full set and the LVQ method is.

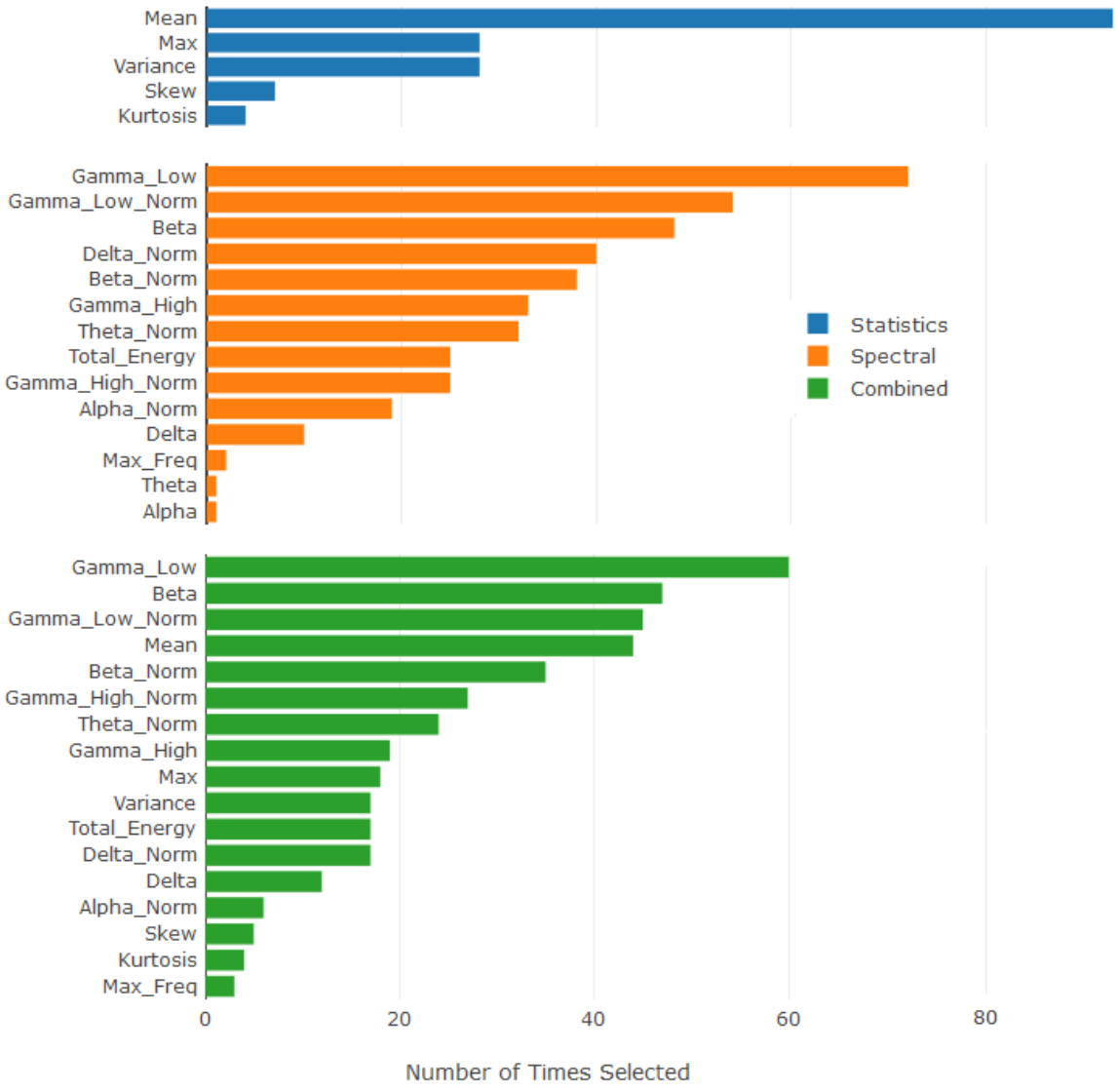

Figure 4.9. Frequency count of features selected using RFE and LVQ. The statistical mean and higher frequency bands (beta, gamma-low and gamma-high) show the greatest discriminatory power.

ment. The spectral features showed a clear dominance of the higher EEG frequencies (gamma-low, gamma-high and beta) relative to the lower frequencies (delta, theta and alpha). The normalised delta, theta and alpha bands were ranked higher than their non-normalised counterparts and the converse was true for the beta, gamma-low and gamma-high bands. The combined set reflected the same mix of preferred features as the constituent sets did. The statistical mean and higher frequency bands again dominated the ranking, and the 2nd, 3rd and 4th moments and the lower frequency bands were less relevant.

To summarise, the influence of the feature selection methods did not have a statistically significant impact on the sensitivity or specificity scores. However, the differences in S1 scores between the models without any feature selection and those using LVQ was significant. Lastly, the mean, gamma and beta frequency bands were the EEG features that were most frequently selected for their discriminative abilities.

Data Sampling Method

The top 27 models all used feature sets with a reduced majority class to balance the class distribution. The maximum sensitivity of models using a reduced set was 0.9800, compared to 0.8700 and 0.7000 for the increased minority method and the imbalanced datasets. The mean sensitivity was 0.8725 for the reduced majority and 0.8200 for the increased minority sets, while the imbalanced sets performed very poorly at 0.3672.

As expected, the imbalanced dataset had the highest mean specificity score of 0.8311, followed by the increased minority method at 0.7347. The reduced sets had the lowest mean specificity at 0.6513. The increased minority models had a mean S1 score of 0.7223, and the reduced majority models had a mean S1 score of 0.6513. The imbalanced data set had the worst mean S1 score, which again relates to the weak sensitivity. See Table 4.13 and Figure 4.10.

Kruskal-Wallis tests for significance showed that the differences in sensitivity, speci-

ficity and S1 scores between the normal, increased minority and reduced majority models were all statistically significant ($p < 0.001$). Models using the reduced sampling method had the best sensitivities, whereas those using the increased minority had the best S1 scores (difference =110.3, critical difference = 24.94). See Table 4.14.

Table 4.13

Statistics Showing Influence of Sample Method on Performance.

Sampling	Measure	Min.	1st Qu.	Median	Mean	3rd Qu.	Maximum
None	Sensitivity	0.0100	0.1500	0.4600	0.3672	0.5100	0.7000
Increased	Sensitivity	0.4300	0.5175	0.7900	0.7071	0.8200	0.8700
Reduced	Sensitivity	0.4200	0.5375	0.8100	0.7347	0.8725	0.9800
None	Specificity	0.4800	0.5000	1.0000	0.8311	1.0000	1.0000
Increased	Specificity	0.4200	0.5375	0.8100	0.7347	0.8725	0.9800
Reduced	Specificity	0.4700	0.5000	0.6350	0.6513	0.7800	0.8500
None	S1 Score	0.0200	0.2600	0.4900	0.4405	0.6050	0.8200
Increased	S1 Score	0.4500	0.5000	0.7700	0.7223	0.8700	0.9100
Reduced	S1 Score	0.4500	0.5100	0.7200	0.6879	0.8125	0.9000

Table 4.14

Kruskal-Wallis Tests for Sampling Performance

Measure	Comparison	H Score	df	Sig.	Difference	Crit. Diff.
Sensitivity	None-Increased	151	2	< 0.001	93.69	24.94
Sensitivity	None-Reduced	151	2	< 0.001	122.3	24.94
Sensitivity	Increased-Reduced	151	2	< 0.001	28.59	24.94
Specificity	None-Increased	153	2	< 0.001	78.81	24.94
Specificity	None-Reduced	153	2	< 0.001	126.0	24.94
Specificity	Increased-Reduced	153	2	< 0.001	47.19	24.94
S1 Score	None-Increased	120	2	< 0.001	110.3	24.94
S1 Score	None-Reduced	120	2	< 0.001	79.73	24.94
S1 Score	Increased-Reduced	120	2	< 0.001	30.59	24.94

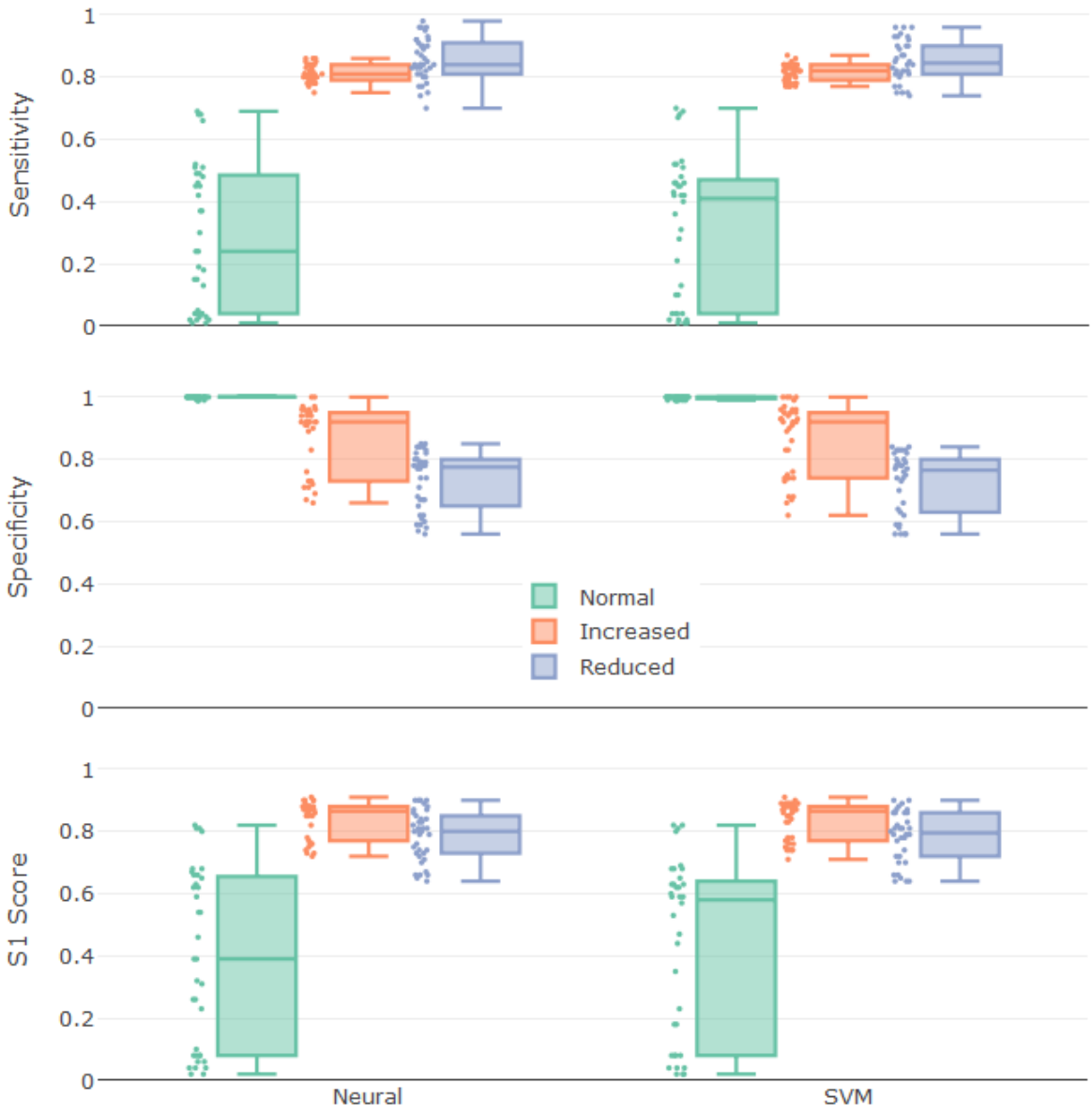

Figure 4.10. Distribution of sensitivity, specificity and S1-score by classifier and sampling technique. The best sensitivity is attained by models using the reduced majority class. Models trained on the normal datasets are severely impacted, particularly those using the statistics features.

Chapter 5

Analysis and Discussion

In the last chapter, the results from the experiments were presented in full. This chapter summarises the key findings, and then comprehensively discusses the results and the implications for seizure prediction research. The results are then compared to some selected studies, and the strengths and limitations of the research are considered.

5.1 Summary of Key Findings

The results prove that ANN classifiers have superior classification performance compared to SVM, when used to predict epileptic seizures using EEG spectral features. Of the four EEG preprocessing methods that were investigated, the EEG segmentation and feature selection methods show no clear statistically significant differences in performance. Models using spectral features demonstrate superior performance relative to statistics-based features. In particular, the gamma and beta bands are the most important indicators of seizure onset. Furthermore, models which use a dataset with a balanced class distribution show statistically significant performance gains.

5.2 Discussion

The primary goal of this research is to compare the performance of SVM and ANN classifiers when used to predict the onset of epileptic seizures. The results show that on average, the neural network models developed during the study have attained higher sensitivity scores relative to the support vector models. The best performing neural network model successfully predicted 98% of epileptic seizures, while the best performing support vector machine predicted 87% of seizures correctly. The findings

are in agreement with those of Alexandre Teixeira et al. (2014), but lend further weight by confirming that the differences in performance are statistically significant ($p = 0.035$), although the effect size is small ($r = -0.143$).

Classifier performance is generally a trade-off between sensitivity and specificity, where the latter measures the ability of the classifier to predict the non-seizure state. In addition to better sensitivity scores, the results clearly demonstrate that the ANN achieves increased specificity scores on average; models using the ANN classifier have higher mean and median scores relative to the SVM models and these differences in performance are statistically significant ($p = 0.0069$).

According to the research goal, the null hypothesis (H_0) is:

> "An artificial neural network does not demonstrate superior classification performance compared to a support vector machine classifier, when using EEG spectral features to predict epileptic seizures".

The alternative hypothesis (H_A) is stated as:

> "An artificial neural network demonstrates superior classification performance compared to a support vector machine classifier, when using EEG spectral features to predict epileptic seizures".

This research clearly shows that an ANN classifier has superior performance compared to a SVM, and furthermore proves that the results are statistically significant. Thus the null hypothesis may be rejected and the alternative hypothesis is accepted.

The secondary goal of this paper is to explore the roles of four preprocessing techniques to, determine their influence on classification performance. Preprocessing raw EEG signals is an important part of the analysis process and contributes towards the development of robust classification models suitable for practical seizure prediction systems.

There is considerable variation in the EEG segmentation parameters used in seizure prediction research, and broadly speaking, none of the studies provide any justification

for their choice of window. There is clearly no agreement on the optimal EEG segment size and scant evidence to support the various research methodologies. The size of the window is a trade-off between information content and computational cost. Features extracted from a longer window will potentially lose valuable temporal information but are faster to compute. Conversely, a narrow window will capture short-term EEG dynamics but is computationally expensive. Each EEG window is a data instance in the final feature matrix - choosing a smaller window can increase the size of the feature matrix by an order of magnitude. This is particularly relevant during model training, where computational complexity is a function of the number of data instances.

The four window sizes had little impact on the performance of the classifiers, and none of these effects were statistically significant. The results suggest that the information content does not vary between EEG segments with 30 second and 60 second durations, and irrespective of whether they are overlapping or contiguous. However, it is certainly feasible that a shorter or longer duration segment may provide some performance improvements. EEG signals may contain important short-span dynamics that require windows shorter than 30 seconds; alternatively, segments longer than 60 seconds may potentially contain long-term trends that would otherwise be overlooked. Evidence for the existence of the preictal state suggests that seizures gradually develop over minutes and hours prior to seizure onset, and longer duration EEG segments may be more useful for detecting these dynamical effects. On a practical note, the research demonstrates that there is no rationale for using overlapping windows. Omitting these in future studies will vastly improve model training times, and reduce data storage requirements. Further research will be needed to determine the optimal EEG segment duration.

The number of different EEG features used in the literature is considerable. Considering the primary goal of this research, this paper does not attempt to determine the optimal type or combination of features, but rather focuses on constructing robust models, based on two types of feature that are well known and relatively simple to extract. Spectral EEG features are known for their discriminatory capabilities, while

statistical features are commonly used in conjunction with other types of features. It is useful to understand the impact of these features on classifier performance, and whether combining the features has any advantages.

The results clearly show that models using the statistics feature set perform the worst relative to the spectral and combined sets. Ranked by sensitivity and then specificity, the bottom 22 models all used statistics features and the normal imbalanced dataset. These models perfectly correctly predict the interictal class (specificity = 1), and conversely incorrectly predict the preictal class (sensitivity < 0.05). The models are thus highly over-trained on the majority interictal class, and the combination of the statistics features and class imbalance makes them ineffective for seizure prediction in this configuration. Nonetheless, a model using statistics features with a reduced majority data set achieved a sensitivity of 0.9, making statistics-based features an attractive, computationally efficient alternative for baselining other more complex models.

Whilst the spectral and combined feature sets show statistically significant differences ($p < 0.01$) compared to the statistics set, the same can not be said for the spectral set compared the combined set. Indeed, including the statistics features does not improve classification performance and needlessly increases training time and model complexity. The best results are obtained using the spectral features, and there is no statistical evidence to support the use of the statistics or combined feature sets.

Seizure prediction research is generally not focused on producing practical seizure forecasting algorithms, and is instead dedicated to maximising classification performance, regardless of any practical implications. Whilst this approach produces good results, it can lead to highly complex ensemble models with thousands of features and intractable training times. Feature selection certainly impacts classification performance, but is an essential consideration when developing algorithms intended for use in embedded devices.

The results show that feature selection reduced the maximum sensitivity score by 0.04, relative to the full feature set. However, the differences in performance between models using the full set of features, and those using a subset of features are not statistically

significant. This is a positive finding, since it means that feature selection does not unduly degrade the classification performance, thus increasing its utility during model exploration. Models using features selected by the RFE method have marginally better mean and median results relative to those using LVQ, but these differences are not statistically significant. Practically speaking, the RFE method is extremely slow to search for the optimal features compared to the LVQ method, and future research efforts will explore PCA and LDA as an alternative.

More importantly, the feature selection methods also show that the most commonly selected spectral features are extracted from the the beta and gamma frequency bands. This provides further support for the findings of Netoff et al. (2009), Park et al. (2011), Bandarabadi, Rasekhi, et al. (2015) and Sharma (2015), who propose that the higher frequency bands are the most useful predictors of seizure activity. Future efforts will focus on exploring higher frequency bands and other potentially useful EEG features, such as entropy or measures of channel synchronisation.

The sampling method has a marked effect on the classifiers' performance. Models using the reduced majority class and the increased minority class base tables show statistically significant improvements in sensitivity, compared to those trained with imbalanced data. Conversely, the imbalanced datasets almost perfectly predict the non-seizure state due to over-training of the majority class, but at the expense of poor sensitivity. The bottom 71 models all used imbalanced data, which clearly shows the negative impact of uneven class distributions on classifier performance. An advantage of using the reduced majority class is that it gives the best sensitivity scores and has the smallest base table, which greatly improves model training times.

The superior sensitivity scores attained by the balanced models are offset by weaker specificity scores. Reducing the over-training effects of the majority class improved classification of the minority preictal class, but diminished the classifiers' ability to predict the non-seizure state. The S1 scores showed that the increased minority sampling was the most balanced, with a good compromise between sensitivity and specificity. Therefore, this method is potentially more suitable for clinical applications that require

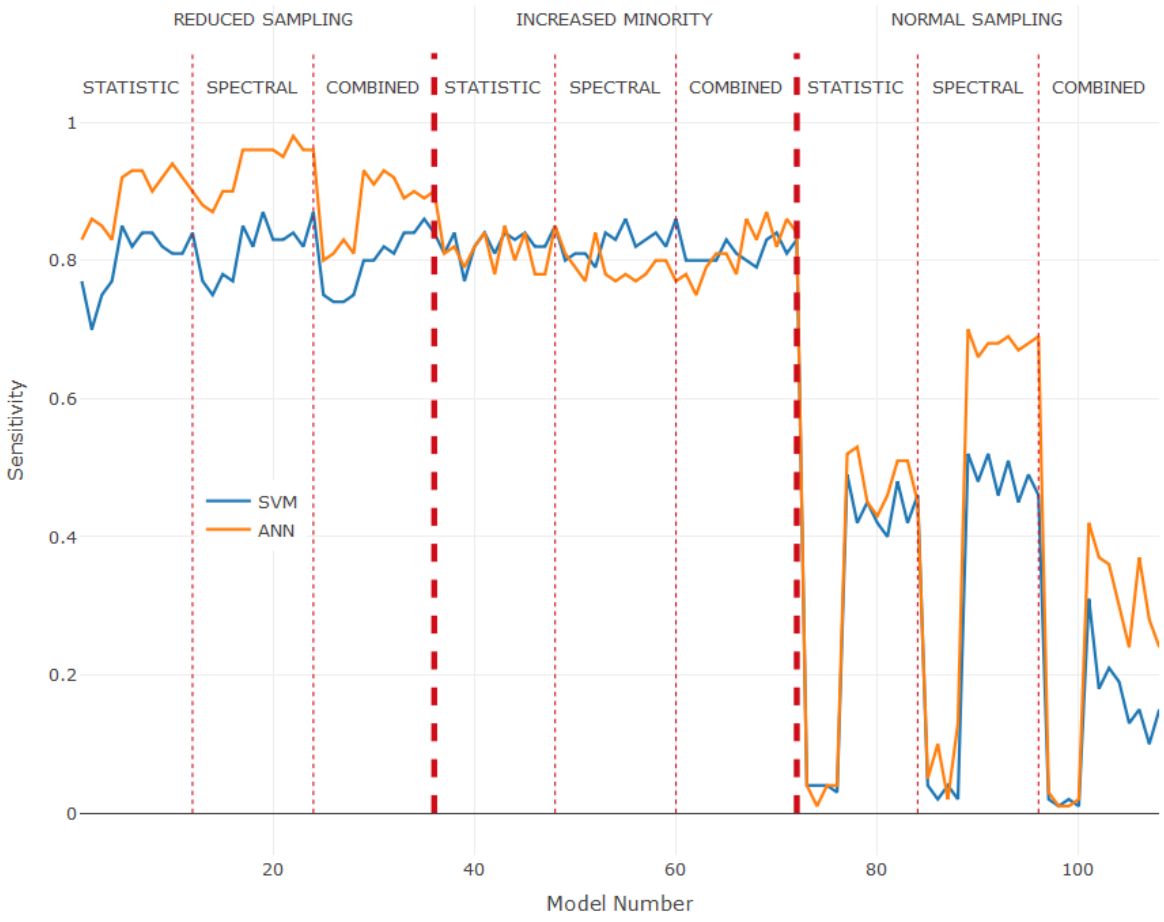

Figure 5.1. Sensitivity scores sorted by sampling method and feature type. Models using the reduced majority and the spectral features are the most promising.

a balance between accurate seizure prediction and low false positive rates.

Finally, assessing the sensitivity scores in their entirety demonstrates the relative impact of each preprocessing method. The results have shown that the feature type and the sampling method are the greatest determinants of performance. Figure 5.1 shows the importance of these two factors in overall classification performance, and again highlights the gap in performance between ANN and SVM classifiers. Furthermore, it shows where future studies may most profitably focus.

5.3 Comparison with Other Studies

One of the principle challenges in seizure prediction research is comparing the results of studies; variations in datasets, classifiers, features, methodologies and evaluation measures all confound attempts at meaningful analysis. Notwithstanding, this study produced a generalised ANN classifier with maximum sensitivity = 0.98 and specificity = 0.84, which compares favourably with the results from the patient-specific studies listed in Table 5.1. It is worth repeating that generalised prediction algorithms are at a performance disadvantage, due to the heterogeneity of seizures between patients.

Table 5.1

Results from Related Seizure Prediction Studies

Study	Database	Classifier	Measure	Result
Costa et al. (2008)	FSPEEG	Various ANN	Sens.	97.8%
			Spec.	99.2%
Chisci et al. (2010)	FSPEEG	SVM & AR	Sens.	100%
Park et al. (2011)	FSPEEG	SVM	Sens.	97.5%
			FPR	0.27/h
Rasekhi et al. (2013)	EPILEPSIAE	SVM	Sens.	73.9%
			FPR	0.15/h
Moghim and Corne (2014)	FSPEEG	SVM	Sens.	91.14%
			Spec.	99.55%
Alexandre Teixeira et al. (2014)	EPILEPSIAE	SVM	Sens.	73.73%
		MLP ANN	Sens.	74.17%
		RBF ANN	Sens.	69.14%
Sudalaimani et al. (2015)	SCTIMST *	SVM	Ratio	26/27
Bandarabadi, Teixeira, et al. (2015)	EPILEPSIAE	SVM	Sens.	75.8%
			FPR	0.1/h

All studies are patient-specific. Number of patients and seizures varies.
Results are averages, with the exception of Sudalaimani et al. (2015).
Refer to Section 3.2 for more information on selected EEG databases.
* See https://www.sctimst.ac.in/

5.4 Research Limitations

Whilst this study used a reputable dataset that contains long-term continuous EEG recordings with mutliple seizures, it could benefit from extending the research to a larger commercial EEG database, such as the EPILEPSIAE repository. There is generally a negative correlation between performance and the number of seizures under review (Schulze-Bonhage et al., 2011) and a larger dataset will further validate the findings of this paper.

The EEG preprocessing techniques did not include any outlier treatment prior to range normalisation, and it is not clear whether removing or retaining the outliers would improve or degrade performance. Most studies retain the raw signal and do not remove outliers, since this may potentially exclude the most discriminatory values. Without further exploration and specific domain knowledge in respect of EEG interpretation, it is difficult to comment definitively.

This paper only uses univariate EEG features for model training, and predictive performance may be restricted by excluding bivariate or multivariate features from consideration. Including multivariate features may improve performance, since they consider information from spatial interactions in the brain (Jouny et al., 2011).

The ANN and SVM algorithms were not individually optimised, due to the extended training times and the number of models under investigation. It may be the case that one of the classifiers is disproportionally affected by this lack of optimisation. Optimisation would potentially improve classification performance, but it is not anticipated that this effect would be large.

The comparison of the SVM and ANN classifiers is not exhaustive, since it only considers a single architecture for each classifier. This paper has clearly shown the impact of imbalanced class distributions, and including a cost sensitive SVM model may improve performance. Similarly, testing with additional neural network architectures would add further weight to the findings.

5.5 Strength of Findings

It is reasonable to state that there is no consensus on the optimal classifier and EEG features that will deliver the predictive performance required for clinical seizure warning applications. This research has investigated some of the problems faced by the field of epileptic seizure research, and tried to avoid the commonly cited criticisms.

Whereas the majority of seizure prediction studies are patient-specific, this research has developed generalised seizure prediction models, which are viewed as more appropriate for off-the-shelf seizure warning systems. It is one of the few papers to compare the performance of ANN classifiers to SVM, and provides fresh evidence for the usefulness of neural networks in seizure prediction research. The results show that the performance of ANN classifiers are superior to SVM, and furthermore that the results are statistically significant. Additionally, the results showed that spectral features are superior to statistics features, and that combining the two does not lead to improved performance. This reinforces the utility of spectral features for seizure prediction research. Related to this, the study has also shown the importance of the beta and gamma frequency bands as predictors of seizure onset. This adds to the existing body of evidence, and points to the potential offered by EEG recordings with high-frequency sampling rates.

This paper also contributes to the limited literature on EEG preprocessing techniques. It is the only research to compare the effects of different EEG windows on classification performance. The results showed no significant difference in performance for the segmentation settings that were explored, and that overlapping windows are unnecessary. The impact of feature selection was also assessed, and showed that reducing the feature count did not significantly reduce classification performance. Finally, the results also highlight the importance of treating the inherent class imbalance present in EEG datasets. Reducing the majority class results in substantial performance gains, and should be a de facto requirement for any future studies.

Chapter 6

Conclusion

This final chapter gives a brief summary of the important aspects of the study. It includes an overview of the original research objectives, the research process and the key results. It concludes by examining the impact of the findings, and offers some final thoughts on potential directions for future seizure prediction research.

6.1 Research Overview

The primary objective was to conduct a quantitative comparison of ANN and SVM supervised machine learning classifiers, used to predict the onset of epileptic seizures. Features were extracted from the 2014 Kaggle Seizure Prediction Challenge EEG dataset, which contains long-term EEG recordings taken from seven subjects. The database contains 7,999 files with a duration of 1,300 hours, which are sampled at 400 Hz or 5 kHz. The number of EEG channels from each file was standardised to 16 during preprocessing.

There were four secondary objectives, that aimed to evaluate the influence of EEG preprocessing techniques on performance. The study principally used EEG spectral features, due their proven efficacy. Fourteen spectral features were extracted from each channel, and included EEG signal energy from six discrete frequency bands, and six features using the normalised signal energy. The total signal energy (0-200 Hz) and the frequency with the maximum signal energy were also included. A second set of features used five statistical measures and included the peak amplitude, mean, variance, skew and kurtosis. Lastly, a third feature set was created by combining both spectral and statistics sets.

Features were extracted from four different EEG segments, using durations of 30 and

60 seconds, with and without an overlap of 50%. The resulting sets of features were reduced using recursive feature elimination (RFE), and a learning vector quantisation method, which ranks the features by relative importance. Based on these results, the full spectral and combined feature sets were reduced to contain 50 features, and the statistics set was reduced to 20 features. Lastly, the impact of data sampling was tested, by balancing the class distributions using a majority downsampling and a minority upsampling process. In order to fully test the effects of each of these methods, a total of 216 support vector machine and neural network models were developed. This large number of models was required in order to hold the control variables constant, while testing the influence of the independent variable in question.

Performance was evaluated using three well-known and widely used measures; sensitivity, specificity and S1 scores. The ANN and SVM performance was compared to that of a random naive classifier, and all results were tested for statistical significance using non-parametric Kruskal-Wallis tests. The classifier performance was assessed against each of the preprocessing techniques, in order to determine the best combination of features and methods that yielded the best results.

There were several difficulties encountered during the study. The R language has limited support for EEG analysis and visualisation, which restricted the scope for exploring different features. In the future it would be preferable to use MATLAB, which has a powerful toolbox called EEGLAB that is widely used and well supported. Additionally, the computing time required for model training and feature selection (particularly RFE) was excessive. The study used multiple virtual machines and took approximately 3,000 hours of computation to complete. This limited the opportunity for model optimisation, which conventionally uses cross-fold validation in a grid search to select optimised parameters.

6.2 Design, Evaluation and Results

A critical review of the literature was completed to provide an overview of the current research on epileptic seizure prediction. This provided the necessary background material and information on the current state-of-the art methods. It also highlighted criticisms and challenges faced by prediction studies, and gaps in the existing body of knowledge. The subsequent quantitative research followed a structured approach, broadly based on the well-known CRISP-DM process. This included data understanding, data preprocessing, modelling, evaluation and analysis. See Figure 6.1.

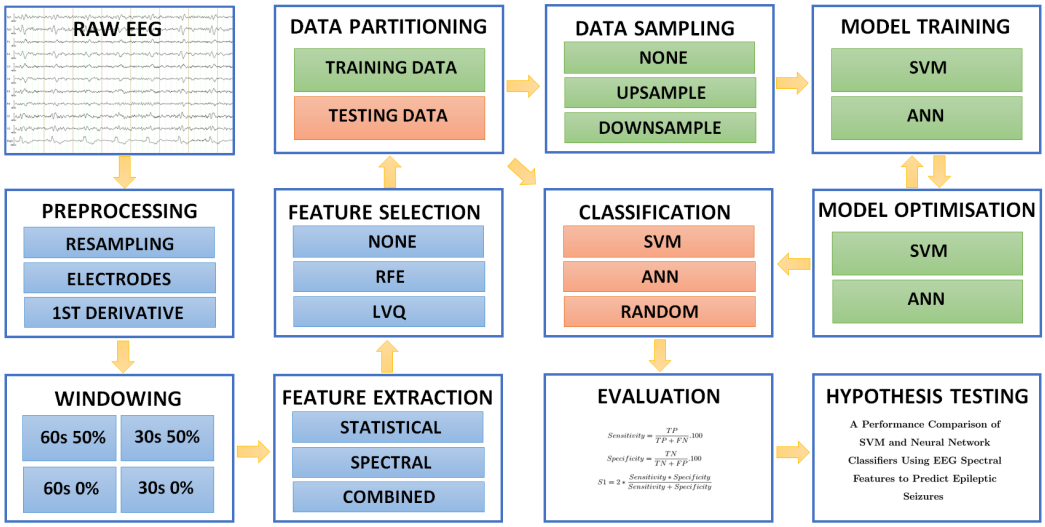

Figure 6.1. Overview of research design

Performance was evaluated using sensitivity, specificity and S1 scores. The results were checked for statistical significance using Kruskal-Wallis tests. The experiments proved that the models using a neural network classifier had superior predictive capabilities compared to those using SVM ($p = 0.035$). Additionally, the spectral EEG features outperformed the statistical features ($p < 0.001$), but the differences between the spectral and combined sets were not significant. The beta and gamma bands were also shown to be the most important spectral features for predicting seizure onset.

In respect of the preprocessing methods, four different EEG windows were tested, but had no significant influence on overall classification performance. Similarly, models

using feature selection showed no significant differences in performance, compared to those using the full set of features. Finally, the data sampling method had a significant effect on performance ($p < 0.001$) and showed that addressing the class imbalance in EEG datasets can bring considerable performance benefits.

6.3 Contributions and Impact

A critical assessment of the literature showed that there is limited research to compare the performance of ANN and SVM, which are two important classifiers used in seizure forecasting. It also highlighted the relative scarcity of information concerning the most effective techniques for preprocessing EEG signals, prior to data modelling. The uncertainty surrounding feature types, segmentation, feature selection and data sampling thus naturally led to the secondary objectives of this research.

- This dissertation provides one of the few explorations of generalised epileptic seizure prediction models, which are viewed as a necessary precursor to practical seizure warning devices. Patient-specific algorithms potentially limit the accessibility of medical devices, and places an additional burden on the limited neurological healthcare resources dedicated to epilepsy treatment.

- This is also one of the few papers to comprehensively compare ANN and SVM algorithms. The results are statistically significant, and are based on a total of 216 models. The findings provide additional evidence for the utility of neural networks in the prediction of epileptic seizure onset.

- The findings reinforce the efficacy of spectral features as predictors of seizure activity. Whilst simple statistical features can provide adequate performance, their usefulness is more limited, particularly if the dataset is imbalanced. Also, combining the spectral and statistical features did not enhance performance, as some studies have hypothesised.

- The output of the RFE and LVQ feature selection show that the beta and gamma bands are the most discriminative frequencies in an EEG. This concurs with

previous research that the higher frequency bands are predictors of seizure onset, and points to the potential utility of analysing EEG with high sampling rates.

- This is the first paper to evaluate the effects of using different EEG window sizes. Whilst further research is needed, this is a valuable first step in determining the optimal window size, which has a material impact on model training times and power consumption in portable devices.
- This paper is also unique in comparing the performance of models using different methods of data sampling. The EEG dataset was sampled using majority downsampling and minority upsampling, and findings show that the best results are achieved with the former. This highlights the necessity for correct data preparation when using EEG recordings, and conveniently shows that optimal performance is obtained from the smallest training sets.

6.4 Future Work and Recommendations

Extending the analysis to a commercial EEG database will test the algorithms against a more heterogeneous set of seizures, from a larger number of patients. If the database contains recordings with high sampling rates (>1 kHz), it will also offer opportunities for exploring the role of high frequency spikes as a seizure precursor. This study was limited to a maximum frequency of 200 Hz and Alexandre Teixeira et al. (2014) showed a positive correlation between higher sampling rate and predictive performance.

The evaluation of the optimal EEG window size should be extended to consider a wider range of values. If it can be proven that longer duration EEG windows provide the best performance, then this will have practical implications for low-power portable medical devices. In principle, longer intervals should result in smaller feature vectors, lower computation times and extended battery life.

This study was never intended to determine the best type of EEG feature, but nonetheless performance could be improved by including additional univariate features, and bivariate features that integrate spatial and temporal information from different re-

gions of the brain. In particular, spectral entropy (Brinkmann et al., 2016), mean phase coherence (Karumuri et al., 2016) and the Hjorth parameters (Rasekhi et al., 2013) are potentially useful features that are worth investigating.

Future efforts should also consider the impact of using cost-sensitive SVM. It is generally more important to classify preictal samples correctly, at the expense of more false positives. With CSVM, higher misclassification penalties can be set for the preictal cases to weight the results accordingly (Park et al., 2011). It would also be informative to develop and compare models using other neural network architectures, such as RBF or recurrent networks, for instance.

Final Thoughts

The unforseen nature of epileptic seizures represents one of the most debilitating aspects of the disease, and instils a feeling of helplessness that has a powerful impact on the patients' quality of life (Andrzejak et al., 2009). The goal for researchers is to develop seizure-triggered diagnostic systems, based on reliable and accurate seizure prediction algorithms. Despite considerable advances, Osorio and Schachter (2011) question the emphasis on implantable intracranial devices to address the needs of pharmacoresistant epileptic patients. They believe that the medical and psychosocial impact of epilepsy, the high cost of care and the substantial technological and human resources required to address the issue, constitutes a health care problem. Scientific advances in the future will have little impact on epilepsy care if devices are not widely accessible.

Seizure prediction remains an active research area with many unanswered questions that must be resolved before commercial seizure prediction devices can be successfully deployed (Gadhoumi et al., 2016). The provision of a practical, widely available and accessible device could have a considerable impact on the quality of life of millions of patients afflicted by this debilitating disorder. Generalised seizure prediction algorithms are one small part of this complex medical and social issue and could help "usher in a new era of epilepsy treatment" (Carney et al., 2011, p.S100).

References

Aarabi, A., Fazel-Rezai, R., & Aghakhani, Y. (2009). EEG seizure prediction: measures and challenges. In *2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (pp. 1864–1867). IEEE. doi: 10.1109/IEMBS.2009.5332620

Aarabi, A., & He, B. (2012, June). A rule-based seizure prediction method for focal neocortical epilepsy. *Clinical Neurophysiology*, *123*(6), 1111–1122. doi: 10.1016/j.clinph.2012.01.014

Alexandre Teixeira, C., Direito, B., Bandarabadi, M., Le Van Quyen, M., Valderrama, M., Schelter, B., ... Dourado, A. (2014, May). Epileptic seizure predictors based on computational intelligence techniques: A comparative study with 278 patients. *Computer Methods and Programs in Biomedicine*, *114*(3), 324–336. doi: 10.1016/j.cmpb.2014.02.007

Andrzejak, R. G., Chicharro, D., Elger, C. E., & Mormann, F. (2009, August). Seizure prediction: Any better than chance? *Clinical Neurophysiology*, *120*(8), 1465–1478. doi: 10.1016/j.clinph.2009.05.019

Arthurs, S., Zaveri, H. P., Frei, M. G., & Osorio, I. (2010, November). Patient and caregiver perspectives on seizure prediction. *Epilepsy & Behavior*, *19*(3), 474–477. doi: 10.1016/j.yebeh.2010.08.010

Asha, S. A., Sudalaimani, C., Devanand, P., Thomas, T. E., & Sudhamony, S. (2013). Automated seizure detection from multichannel eeg signals using support vector machine and artificial neural networks. In *Automation, Computing, Communication, Control and Compressed Sensing (iMac4s), 2013 International Multi-Conference on* (pp. 558–563). IEEE. doi: 10.1109/iMac4s.2013.6526473

Badawy, R., Macdonell, R., Jackson, G., & Berkovic, S. (2009, April). The peri-ictal state: cortical excitability changes within 24 h of a seizure. *Brain*, *132*(4), 1013–1021.

doi: 10.1093/brain/awp017

Bandarabadi, M., Rasekhi, J., Teixeira, C. A., Karami, M. R., & Dourado, A. (2015, May). On the proper selection of preictal period for seizure prediction. *Epilepsy & Behavior*, *46*, 158–166. doi: 10.1016/j.yebeh.2015.03.010

Bandarabadi, M., Teixeira, C. A., Rasekhi, J., & Dourado, A. (2015, February). Epileptic seizure prediction using relative spectral power features. *Clinical Neurophysiology*, *126*(2), 237–248. doi: 10.1016/j.clinph.2014.05.022

Berendt, M., & Gram, L. (1999). Epilepsy and seizure classification in 63 dogs: a reappraisal of veterinary epilepsy terminology. *Journal of Veterinary Internal Medicine*, *13*(1), 14–20. doi: 10.1111/j.1939-1676.1999.tb02159.x

Brinkmann, B. H., Patterson, E. E., Vite, C., Vasoli, V. M., Crepeau, D., Stead, M., ... Worrell, G. A. (2015, August). Forecasting Seizures Using Intracranial EEG Measures and SVM in Naturally Occurring Canine Epilepsy. *PLOS ONE*, *10*(8), e0133900. doi: 10.1371/journal.pone.0133900

Brinkmann, B. H., Wagenaar, J., Abbot, D., Adkins, P., Bosshard, S. C., Chen, M., ... Worrell, G. A. (2016, June). Crowdsourcing reproducible seizure forecasting in human and canine epilepsy. *Brain*, *139*(6), 1713–1722. doi: 10.1093/brain/aww045

Carney, P. R., Myers, S., & Geyer, J. D. (2011, December). Seizure prediction: Methods. *Epilepsy & Behavior*, *22*, S94–S101. doi: 10.1016/j.yebeh.2011.09.001

Chisci, L., Mavino, A., Perferi, G., Sciandrone, M., Anile, C., Colicchio, G., & Fuggetta, F. (2010, May). Real-Time Epileptic Seizure Prediction Using AR Models and Support Vector Machines. *IEEE Transactions on Biomedical Engineering*, *57*(5), 1124–1132. doi: 10.1109/TBME.2009.2038990

Cook, M. J., O'Brien, T. J., Berkovic, S. F., Murphy, M., Morokoff, A., Fabinyi, G., ... others (2013). Prediction of seizure likelihood with a long-term, implanted seizure advisory system in patients with drug-resistant epilepsy: a first-in-man study. *The Lancet Neurology*, *12*(6), 563–571. doi: 10.1016/S1474-4422(13)70075-9

Costa, R. P., Oliveira, P., Rodrigues, G., Leitao, B., & Dourado, A. (2008). Epileptic seizure classification using neural networks with 14 features. In *International Conference on Knowledge-Based and Intelligent Information and Engineering Systems* (pp. 281–288). Springer. doi: 10.1007/978-3-540-85565-1_35

Da Silva, F. L., Blanes, W., Kalitzin, S. N., Parra, J., Suffczynski, P., & Velis, D. N. (2003). Epilepsies as dynamical diseases of brain systems: basic models of the transition between normal and epileptic activity. *Epilepsia*, *44*(s12), 72–83. doi: 10.1111/j.0013-9580.2003.12005.x

Davis, K. A., Sturges, B. K., Vite, C. H., Ruedebusch, V., Worrell, G., Gardner, A. B., ... Litt, B. (2011, September). A novel implanted device to wirelessly record and analyze continuous intracranial canine EEG. *Epilepsy Research*, *96*(1-2), 116–122. doi: 10.1016/j.eplepsyres.2011.05.011

Direito, B. (2011). Feature selection in high dimensional EEG features spaces for epileptic seizure prediction.. doi: 10.3182/20110828-6-IT-1002.03331

Feldwisch-Drentrup, H., Ihle, M., Le Van Quyen, M., Teixeira, C., Dourado, A., Timmer, J., ... Schelter, B. (2011, December). Anticipating the unobserved: Prediction of subclinical seizures. *Epilepsy & Behavior*, *22*, S119–S126. doi: 10.1016/j.yebeh.2011.08.023

Feldwisch-Drentrup, H., Schelter, B., Jachan, M., Nawrath, J., Timmer, J., & Schulze-Bonhage, A. (2010, January). Joining the benefits: Combining epileptic seizure prediction methods: Combining Epileptic Seizure Prediction Methods. *Epilepsia*, *51*(8), 1598–1606. doi: 10.1111/j.1528-1167.2009.02497.x

Fergus, P., Hussain, A., Hignett, D., Al-Jumeily, D., Abdel-Aziz, K., & Hamdan, H. (2016, January). A machine learning system for automated whole-brain seizure detection. *Applied Computing and Informatics*, *12*(1), 70–89. doi: 10.1016/j.aci.2015.01.001

Gadhoumi, K., Lina, J.-M., & Gotman, J. (2013, September). Seizure prediction in patients with mesial temporal lobe epilepsy using EEG measures of state similarity.

Clinical Neurophysiology, *124*(9), 1745–1754. doi: 10.1016/j.clinph.2013.04.006

Gadhoumi, K., Lina, J.-M., Mormann, F., & Gotman, J. (2016, February). Seizure prediction for therapeutic devices: A review. *Journal of Neuroscience Methods*, *260*, 270–282. doi: 10.1016/j.jneumeth.2015.06.010

Gotman, J. (2011, December). A few thoughts on What is a seizure?. *Epilepsy & Behavior*, *22*, S2–S3. doi: 10.1016/j.yebeh.2011.08.025

Harati, A., Lopez, S., Obeid, I., Picone, J., Jacobson, M. P., & Tobochnik, S. (2014). The TUH EEG CORPUS: A big data resource for automated eeg interpretation. In *Signal Processing in Medicine and Biology Symposium (SPMB), 2014 IEEE* (pp. 1–5). IEEE. doi: 10.1109/SPMB.2014.7002953

Howbert, J. J., Patterson, E. E., Stead, S. M., Brinkmann, B., Vasoli, V., Crepeau, D., ... Worrell, G. A. (2014, January). Forecasting Seizures in Dogs with Naturally Occurring Epilepsy. *PLoS ONE*, *9*(1), e81920. doi: 10.1371/journal.pone.0081920

Jouny, C. C., Franaszczuk, P. J., & Bergey, G. K. (2011, December). Improving early seizure detection. *Epilepsy & Behavior*, *22*, S44–S48. doi: 10.1016/j.yebeh.2011.08.029

Karoly, P. J., Freestone, D. R., Boston, R., Grayden, D. B., Himes, D., Leyde, K., ... Cook, M. J. (2016, April). Interictal spikes and epileptic seizures: their relationship and underlying rhythmicity. *Brain*, *139*(4), 1066–1078. doi: 10.1093/brain/aww019

Karumuri, B. K., Vlachos, I., Liu, R., Adkinson, J. A., & Iasemidis, L. (2016, March). Classification of Pre-ictal and Interictal Periods Based on EEG Frequency Features in Epilepsy. In (pp. 9–10). IEEE. doi: 10.1109/SBEC.2016.9

Kelleher, J. D., Namee, B. M., & D'Arcy, A. (2015). *Fundamentals of Machine Learning for Predictive Data Analytics: Algorithms, Worked Examples, and Case Studies*. The MIT Press.

Kharbouch, A., Shoeb, A., Guttag, J., & Cash, S. S. (2011, December). An algorithm for seizure onset detection using intracranial EEG. *Epilepsy & Behavior*, *22*, S29–

S35. doi: 10.1016/j.yebeh.2011.08.031

Klatt, J., Feldwisch-Drentrup, H., Ihle, M., Navarro, V., Neufang, M., Teixeira, C., ... Schelter, B. (2012, September). The EPILEPSIAE database: An extensive electroencephalography database of epilepsy patients: European EEG Database EPILEPSIAE. *Epilepsia*, *53*(9), 1669–1676. doi: 10.1111/j.1528-1167.2012.03564.x

Kumar, S. P., Sriraam, N., & Benakop, P. G. (2008). Automated detection of epileptic seizures using wavelet entropy feature with recurrent neural network classifier. *TENCON IEEE*, 1–5. doi: 10.1109/TENCON.2008.4766836

Lehnertz, K., Andrzejak, R. G., Arnhold, J., Kreuz, T., Mormann, F., Rieke, C., ... Elger, C. E. (2001). Nonlinear EEG Analysis in Epilepsy: Its Possible Use for Interictal Focus Localization, Seizure Anticipation, and. *Journal of Clinical Neurophysiology*, *18*(3), 209–222.

Maiwald, T., Winterhalder, M., Aschenbrenner-Scheibe, R., Voss, H. U., Schulze-Bonhage, A., & Timmer, J. (2004, July). Comparison of three nonlinear seizure prediction methods by means of the seizure prediction characteristic. *Physica D: Nonlinear Phenomena*, *194*(3-4), 357–368. doi: 10.1016/j.physd.2004.02.013

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

Mirowski, P. W., LeCun, Y., Madhavan, D., & Kuzniecky, R. (2008). Comparing SVM and convolutional networks for epileptic seizure prediction from intracranial EEG. In *2008 IEEE Workshop on Machine Learning for Signal Processing* (pp. 244–249). IEEE. doi: 10.1109/iMac4s.2013.6526473

Moghim, N., & Corne, D. W. (2014, June). Predicting Epileptic Seizures in Advance. *PLoS ONE*, *9*(6), e99334. doi: 10.1371/journal.pone.0099334

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

brain/awl241

Mormann, F., & Elger, C. E. (2013, June). Seizure Prediction and Documentation Two Important Problems. *The Lancet Neurology*, *12*(6), 531–532. doi: 10.1016/S1474-4422(13)70092-9

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

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

Niknazar, M., Mousavi, S., Motaghi, S., Dehghani, A., Vosoughi Vahdat, B., Shamsollahi, M., ... Noorbakhsh, S. (2013, May). A unified approach for detection of induced epileptic seizures in rats using ECoG signals. *Epilepsy & Behavior*, *27*(2), 355–364. doi: 10.1016/j.yebeh.2013.01.028

Osorio, I., & Schachter, S. (2011, December). Extracerebral detection of seizures: A new era in epileptology? *Epilepsy & Behavior*, *22*, S82–S87. doi: 10.1016/j.yebeh.2011.09.012

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

Qiao, X., & Liu, Y. (2009, March). Adaptive Weighted Learning for Unbalanced Multicategory Classification. *Biometrics*, *65*(1), 159–168. doi: 10.1111/j.1541-0420.2008.01017.x

Ramgopal, S., Thome-Souza, S., Jackson, M., Kadish, N. E., Snchez Fernndez, I., Klehm, J., ... Loddenkemper, T. (2014, August). Seizure detection, seizure predic-

tion, and closed-loop warning systems in epilepsy. *Epilepsy & Behavior*, *37*, 291–307. doi: 10.1016/j.yebeh.2014.06.023

Rasekhi, J., Mollaei, M. R. K., Bandarabadi, M., Teixeira, C. A., & Dourado, A. (2013, July). Preprocessing effects of 22 linear univariate features on the performance of seizure prediction methods. *Journal of Neuroscience Methods*, *217*(1-2), 9–16. doi: 10.1016/j.jneumeth.2013.03.019

Salami, P., Lvesque, M., Gotman, J., & Avoli, M. (2012, November). A comparison between automated detection methods of high-frequency oscillations (80500hz) during seizures. *Journal of Neuroscience Methods*, *211*(2), 265–271. doi: 10.1016/j.jneumeth.2012.09.003

Schelter, B., Winterhalder, M., Maiwald, T., Brandt, A., Schad, A., Schulze-Bonhage, A., & Timmer, J. (2006, March). Testing statistical significance of multivariate time series analysis techniques for epileptic seizure prediction. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, *16*(1), 013108. doi: 10.1063/1.2137623

Schulze-Bonhage, A., Feldwisch-Drentrup, H., & Ihle, M. (2011, December). The role of high-quality EEG databases in the improvement and assessment of seizure prediction methods. *Epilepsy & Behavior*, *22*, S88–S93. doi: 10.1016/j.yebeh.2011.08.030

Shakirullah, S., Niaz, A., Khan, A., & Nabi, M. (2014). The Prevalence, Incidence and Etiology of Epilepsy. *International Journal of Clinical and Experimental Neurology*, *2*(2), 29–39. doi: 10.12691/ijcen-2-2-3

Sharma, A. (2015). Epileptic seizure prediction using power analysis in beta band of EEG signals. In *Soft Computing Techniques and Implementations (ICSCTI), 2015 International Conference on* (pp. 117–121). IEEE. doi: 10.1109/ICSCTI.2015.7489552

Shoeb, A., Kharbouch, A., Soegaard, J., Schachter, S., & Guttag, J. (2011, December). A machine-learning algorithm for detecting seizure termination in scalp EEG. *Epilepsy & Behavior*, *22*, S36–S43. doi: 10.1016/j.yebeh.2011.08.040

Shoeb, A. H. (2009). *Application of machine learning to epileptic seizure onset detection and treatment* (Doctoral dissertation, Massachusetts Institute of Technology). Retrieved 2016-11-12, from http://dspace.mit.edu/handle/1721.1/54669

Shoeb, A. H., & Guttag, J. V. (2010). Application of machine learning to epileptic seizure detection. In *Proceedings of the 27th International Conference on Machine Learning (ICML-10)* (pp. 975–982). Retrieved 2017-03-14, from http://physionet.mit.edu/physiobank/database/chbmit/shoeb-icml-2010.pdf

Sing, T., Sander, O., Beerenwinkel, N., & Lengauer, T. (2005, October). ROCR: visualizing classifier performance in R. *Bioinformatics*, *21*(20), 3940–3941. doi: 10.1093/bioinformatics/bti623

Smith, S. J. M. (2005, June). EEG in the diagnosis, classification, and management of patients with epilepsy. *Journal of Neurology, Neurosurgery & Psychiatry*, *76*(suppl_2), ii2–ii7. doi: 10.1136/jnnp.2005.069245

Snyder, D. E., Echauz, J., Grimes, D. B., & Litt, B. (2008, December). The statistics of a practical seizure warning system. *Journal of Neural Engineering*, *5*(4), 392–401. doi: 10.1088/1741-2560/5/4/004

Stacey, W., Le Van Quyen, M., Mormann, F., & Schulze-Bonhage, A. (2011, December). What is the present-day EEG evidence for a preictal state? *Epilepsy Research*, *97*(3), 243–251. doi: 10.1016/j.eplepsyres.2011.07.012

Sudalaimani, C., Asha, S. A., Parvathy, K., Thomas, T. E., Devanand, P., Sasi, P. M., ... Thomas, S. V. (2015). Use of electrographic seizures and interictal epileptiform discharges for improving performance in seizure prediction. In *Intelligent Computational Systems (RAICS), 2015 IEEE Recent Advances in* (pp. 229–234). IEEE. doi: 10.1109/RAICS.2015.7488419

Sudalaimani, C., Thomas, T. E., Sivakumaran, N., Valsalam, S. R., Meethal, S. P., & Sathish, E. (2017). Seizure prediction using general regression neural network. In *Trends in Industrial Measurement and Automation (TIMA), 2017* (pp. 1–7). IEEE. doi: 10.1109/TIMA.2017.8064810

Temko, A., Sarkar, A., & Lightbody, G. (2015). Detection of Seizures in Intracranial EEG: UPenn and Mayo Clinic's Seizure Detection Challenge. In *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (pp. 6582–6585). IEEE. doi: 10.1109/EMBC.2015.7319901

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

Yu, P.-N., Naiini, S. A., Heck, C. N., Liu, C. Y., Song, D., & Berger, T. W. (2016). A sparse Laguerre-Volterra autoregressive model for seizure prediction in temporal lobe epilepsy. In *Engineering in Medicine and Biology Society (EMBC), 2016 IEEE 38th Annual International Conference of the* (pp. 1664–1667). IEEE. doi: 10.1109/EMBC.2016.7591034

Zheng, Y., Wang, G., Li, K., Bao, G., & Wang, J. (2014, June). Epileptic seizure prediction using phase synchronization based on bivariate empirical mode decomposition. *Clinical Neurophysiology*, *125*(6), 1104–1111. doi: 10.1016/j.clinph.2013.09.047

Appendix A

Performance Results

Table A.1

ANN and SVM performance ranked by sensitivity and specificity (descending)

Classifier	Window	Feature	Selection	Sampling	Sens.	Spec.	S1 Score
Neural	60s 50%	Both	None	Reduced	0.98	0.84	0.9
Neural	30s 50%	Both	None	Reduced	0.96	0.85	0.9
Neural	60s 0%	Both	None	Reduced	0.96	0.84	0.9
Neural	30s 50%	Spectral	None	Reduced	0.96	0.84	0.9
Neural	30s 0%	Spectral	None	Reduced	0.96	0.84	0.9
Neural	60s 50%	Spectral	None	Reduced	0.96	0.83	0.89
Neural	60s 0%	Spectral	None	Reduced	0.96	0.83	0.89
Neural	30s 0%	Both	None	Reduced	0.95	0.85	0.9
Neural	60s 50%	Both	RFE	Reduced	0.94	0.82	0.88
Neural	60s 0%	Spectral	RFE	Reduced	0.93	0.82	0.87
Neural	30s 50%	Spectral	RFE	Reduced	0.93	0.82	0.87
Neural	60s 50%	Spectral	LVQ	Reduced	0.93	0.8	0.86
Neural	30s 50%	Spectral	LVQ	Reduced	0.93	0.8	0.86
Neural	60s 0%	Both	RFE	Reduced	0.92	0.84	0.88
Neural	30s 0%	Both	RFE	Reduced	0.92	0.83	0.87
Neural	60s 50%	Spectral	RFE	Reduced	0.92	0.82	0.87
Neural	30s 0%	Spectral	LVQ	Reduced	0.92	0.8	0.86
Neural	60s 0%	Spectral	LVQ	Reduced	0.91	0.79	0.85
Neural	30s 50%	Both	RFE	Reduced	0.9	0.83	0.86
Neural	30s 0%	Spectral	RFE	Reduced	0.9	0.83	0.86
Neural	60s 50%	Both	LVQ	Reduced	0.9	0.79	0.84
Neural	30s 50%	Both	LVQ	Reduced	0.9	0.78	0.84
Neural	30s 50%	Statistics	None	Reduced	0.9	0.74	0.81
Neural	30s 0%	Statistics	None	Reduced	0.9	0.73	0.81
Neural	30s 0%	Both	LVQ	Reduced	0.89	0.77	0.83
Neural	60s 0%	Both	LVQ	Reduced	0.89	0.75	0.81
Neural	60s 50%	Statistics	None	Reduced	0.88	0.74	0.8
Neural	30s 0%	Both	LVQ	Increased	0.87	0.91	0.89
SVM	30s 50%	Both	None	Reduced	0.87	0.8	0.83
SVM	30s 50%	Spectral	None	Reduced	0.87	0.79	0.83
Neural	60s 0%	Statistics	None	Reduced	0.87	0.7	0.78
SVM	30s 50%	Spectral	None	Increased	0.86	0.96	0.91
SVM	30s 50%	Both	None	Increased	0.86	0.96	0.91
Neural	30s 50%	Spectral	LVQ	Increased	0.86	0.94	0.9
Neural	60s 0%	Both	LVQ	Increased	0.86	0.92	0.89
Neural	60s 0%	Statistics	RFE	Reduced	0.86	0.62	0.72
SVM	60s 0%	Both	LVQ	Reduced	0.86	0.59	0.7
Neural	30s 50%	Both	RFE	Increased	0.85	0.95	0.9

Continued on next page

Table A.1 – continued from previous page

Classifier	Window	Feature	Selection	Sampling	Sens.	Spec.	S1 Score
Neural	30s 50%	Spectral	RFE	Increased	0.85	0.94	0.89
SVM	30s 50%	Both	RFE	Increased	0.85	0.92	0.88
SVM	60s 50%	Spectral	None	Reduced	0.85	0.79	0.82
SVM	60s 50%	Spectral	RFE	Reduced	0.85	0.79	0.82
Neural	30s 50%	Statistics	RFE	Reduced	0.85	0.66	0.74
Neural	60s 50%	Spectral	RFE	Increased	0.84	0.96	0.9
Neural	30s 0%	Both	RFE	Increased	0.84	0.96	0.9
SVM	60s 50%	Spectral	None	Increased	0.84	0.96	0.9
SVM	60s 50%	Both	None	Increased	0.84	0.95	0.89
SVM	30s 50%	Spectral	RFE	Increased	0.84	0.93	0.88
Neural	30s 50%	Both	LVQ	Increased	0.84	0.92	0.88
SVM	60s 50%	Spectral	RFE	Increased	0.84	0.92	0.88
SVM	30s 0%	Both	RFE	Increased	0.84	0.92	0.88
Neural	30s 0%	Statistics	None	Increased	0.84	0.89	0.86
SVM	60s 50%	Both	LVQ	Increased	0.84	0.86	0.85
SVM	60s 50%	Both	None	Reduced	0.84	0.79	0.81
SVM	30s 50%	Spectral	RFE	Reduced	0.84	0.78	0.81
SVM	30s 50%	Both	RFE	Reduced	0.84	0.78	0.81
SVM	30s 0%	Spectral	RFE	Reduced	0.84	0.78	0.81
SVM	60s 50%	Both	LVQ	Reduced	0.84	0.67	0.75
SVM	60s 0%	Statistics	RFE	Increased	0.84	0.66	0.74
SVM	30s 50%	Both	LVQ	Reduced	0.84	0.65	0.73
SVM	30s 0%	Both	LVQ	Reduced	0.84	0.62	0.71
SVM	60s 0%	Spectral	None	Increased	0.83	0.95	0.89
SVM	30s 0%	Both	None	Increased	0.83	0.95	0.89
Neural	30s 0%	Spectral	LVQ	Increased	0.83	0.93	0.88
SVM	30s 0%	Spectral	RFE	Increased	0.83	0.92	0.87
SVM	60s 50%	Spectral	LVQ	Increased	0.83	0.91	0.87
SVM	30s 50%	Both	LVQ	Increased	0.83	0.83	0.83
SVM	30s 0%	Both	LVQ	Increased	0.83	0.83	0.83
SVM	30s 0%	Spectral	None	Reduced	0.83	0.78	0.8
SVM	30s 0%	Both	None	Reduced	0.83	0.78	0.8
Neural	30s 50%	Statistics	LVQ	Reduced	0.83	0.67	0.74
Neural	60s 50%	Statistics	RFE	Reduced	0.83	0.63	0.72
Neural	30s 0%	Statistics	RFE	Reduced	0.83	0.61	0.7
SVM	30s 0%	Spectral	None	Increased	0.82	0.95	0.88
SVM	60s 0%	Both	None	Increased	0.82	0.94	0.88
Neural	60s 50%	Both	LVQ	Increased	0.82	0.93	0.87
SVM	60s 50%	Both	RFE	Increased	0.82	0.92	0.87
SVM	60s 0%	Both	RFE	Increased	0.82	0.92	0.87
SVM	30s 0%	Both	RFE	Reduced	0.82	0.78	0.8
SVM	60s 0%	Spectral	None	Reduced	0.82	0.77	0.79
SVM	60s 0%	Spectral	RFE	Reduced	0.82	0.77	0.79
SVM	60s 0%	Both	None	Reduced	0.82	0.76	0.79
Neural	30s 0%	Statistics	RFE	Increased	0.82	0.75	0.78
SVM	30s 50%	Spectral	LVQ	Reduced	0.82	0.74	0.78
SVM	30s 0%	Statistics	RFE	Increased	0.82	0.67	0.74
Neural	60s 0%	Statistics	RFE	Increased	0.82	0.62	0.71
Neural	60s 50%	Spectral	LVQ	Increased	0.81	0.95	0.87
SVM	60s 0%	Spectral	RFE	Increased	0.81	0.92	0.86

Continued on next page

Table A.1 – continued from previous page

Classifier	Window	Feature	Selection	Sampling	Sens.	Spec.	S1 Score
SVM	60s 0%	Spectral	LVQ	Increased	0.81	0.89	0.85
SVM	60s 0%	Both	LVQ	Increased	0.81	0.83	0.82
SVM	60s 0%	Both	RFE	Reduced	0.81	0.78	0.79
SVM	60s 50%	Both	RFE	Reduced	0.81	0.77	0.79
Neural	60s 50%	Statistics	None	Increased	0.81	0.76	0.78
Neural	60s 50%	Statistics	RFE	Increased	0.81	0.76	0.78
SVM	30s 50%	Statistics	None	Increased	0.81	0.74	0.77
Neural	30s 0%	Statistics	LVQ	Increased	0.81	0.73	0.77
SVM	60s 0%	Statistics	None	Increased	0.81	0.71	0.76
SVM	30s 0%	Spectral	LVQ	Reduced	0.81	0.71	0.76
SVM	60s 50%	Statistics	RFE	Increased	0.81	0.69	0.75
Neural	30s 0%	Statistics	LVQ	Reduced	0.81	0.64	0.72
Neural	60s 0%	Statistics	LVQ	Reduced	0.81	0.62	0.7
Neural	60s 50%	Both	None	Increased	0.8	1	0.89
Neural	60s 0%	Both	None	Increased	0.8	0.97	0.88
Neural	30s 0%	Spectral	RFE	Increased	0.8	0.96	0.87
SVM	30s 50%	Spectral	LVQ	Increased	0.8	0.9	0.85
SVM	60s 50%	Statistics	None	Increased	0.8	0.74	0.77
SVM	60s 50%	Spectral	LVQ	Reduced	0.8	0.74	0.77
SVM	60s 50%	Statistics	LVQ	Increased	0.8	0.68	0.74
SVM	60s 0%	Spectral	LVQ	Reduced	0.8	0.68	0.74
SVM	30s 50%	Statistics	LVQ	Increased	0.8	0.68	0.74
Neural	60s 50%	Statistics	LVQ	Reduced	0.8	0.67	0.73
SVM	30s 0%	Statistics	LVQ	Increased	0.8	0.67	0.73
SVM	60s 0%	Statistics	LVQ	Increased	0.8	0.66	0.72
Neural	60s 0%	Statistics	None	Increased	0.79	0.91	0.85
SVM	30s 0%	Spectral	LVQ	Increased	0.79	0.9	0.84
Neural	30s 50%	Statistics	LVQ	Increased	0.79	0.74	0.76
Neural	30s 50%	Statistics	RFE	Increased	0.79	0.73	0.76
SVM	30s 0%	Statistics	None	Increased	0.79	0.73	0.76
Neural	60s 50%	Spectral	None	Increased	0.78	1	0.88
Neural	30s 50%	Spectral	None	Increased	0.78	1	0.88
Neural	30s 0%	Both	None	Increased	0.78	1	0.88
Neural	60s 0%	Spectral	RFE	Increased	0.78	0.97	0.86
Neural	60s 0%	Both	RFE	Increased	0.78	0.97	0.86
Neural	60s 0%	Spectral	LVQ	Increased	0.78	0.96	0.86
Neural	60s 50%	Both	RFE	Increased	0.78	0.94	0.85
Neural	60s 50%	Statistics	LVQ	Increased	0.78	0.74	0.76
SVM	30s 50%	Statistics	None	Reduced	0.78	0.59	0.67
Neural	30s 50%	Both	None	Increased	0.77	1	0.87
Neural	30s 0%	Spectral	None	Increased	0.77	1	0.87
Neural	60s 0%	Spectral	None	Increased	0.77	0.99	0.87
Neural	30s 50%	Statistics	None	Increased	0.77	0.73	0.75
SVM	30s 50%	Statistics	RFE	Increased	0.77	0.71	0.74
SVM	60s 50%	Statistics	None	Reduced	0.77	0.58	0.66
SVM	60s 50%	Statistics	RFE	Reduced	0.77	0.58	0.66
SVM	30s 0%	Statistics	RFE	Reduced	0.77	0.57	0.66
SVM	30s 0%	Statistics	None	Reduced	0.77	0.56	0.65
Neural	60s 0%	Statistics	LVQ	Increased	0.75	0.72	0.73
SVM	30s 50%	Statistics	RFE	Reduced	0.75	0.59	0.66

Continued on next page

Table A.1 – continued from previous page

Classifier	Window	Feature	Selection	Sampling	Sens.	Spec.	S1 Score
SVM	60s 50%	Statistics	LVQ	Reduced	0.75	0.56	0.64
SVM	60s 0%	Statistics	None	Reduced	0.75	0.56	0.64
SVM	30s 0%	Statistics	LVQ	Reduced	0.75	0.56	0.64
SVM	30s 50%	Statistics	LVQ	Reduced	0.74	0.59	0.66
SVM	60s 0%	Statistics	LVQ	Reduced	0.74	0.56	0.64
Neural	60s 50%	Spectral	None	Normal	0.7	1	0.82
SVM	60s 0%	Statistics	RFE	Reduced	0.7	0.6	0.65
Neural	60s 50%	Both	None	Normal	0.69	1	0.82
Neural	30s 0%	Both	None	Normal	0.69	1	0.82
Neural	30s 50%	Spectral	None	Normal	0.68	1	0.81
Neural	30s 50%	Both	None	Normal	0.68	1	0.81
Neural	30s 0%	Spectral	None	Normal	0.68	1	0.81
Neural	60s 0%	Both	None	Normal	0.67	1	0.8
Neural	60s 0%	Spectral	None	Normal	0.66	1	0.8
Neural	60s 0%	Spectral	RFE	Normal	0.53	1	0.69
Neural	60s 50%	Spectral	RFE	Normal	0.52	1	0.68
SVM	30s 50%	Spectral	None	Normal	0.52	1	0.68
SVM	60s 50%	Spectral	None	Normal	0.52	0.99	0.68
Neural	60s 50%	Both	RFE	Normal	0.51	1	0.68
Neural	60s 0%	Both	RFE	Normal	0.51	1	0.68
SVM	60s 50%	Both	None	Normal	0.51	0.99	0.67
SVM	60s 50%	Spectral	RFE	Normal	0.49	1	0.66
SVM	30s 50%	Both	None	Normal	0.49	0.99	0.66
SVM	60s 50%	Both	RFE	Normal	0.48	0.99	0.65
SVM	60s 0%	Spectral	None	Normal	0.48	0.99	0.65
Neural	30s 0%	Both	RFE	Normal	0.46	1	0.63
SVM	30s 50%	Both	RFE	Normal	0.46	0.99	0.63
SVM	30s 0%	Spectral	None	Normal	0.46	0.99	0.63
SVM	30s 0%	Both	None	Normal	0.46	0.99	0.63
Neural	30s 50%	Spectral	RFE	Normal	0.45	1	0.62
Neural	30s 50%	Both	RFE	Normal	0.45	1	0.62
SVM	60s 0%	Both	None	Normal	0.45	0.99	0.62
SVM	30s 50%	Spectral	RFE	Normal	0.45	0.99	0.62
Neural	30s 0%	Spectral	RFE	Normal	0.43	1	0.6
Neural	60s 50%	Spectral	LVQ	Normal	0.42	1	0.59
SVM	60s 0%	Spectral	RFE	Normal	0.42	0.99	0.59
SVM	60s 0%	Both	RFE	Normal	0.42	0.99	0.59
SVM	30s 0%	Spectral	RFE	Normal	0.42	0.99	0.59
SVM	30s 0%	Both	RFE	Normal	0.4	0.99	0.57
Neural	60s 50%	Both	LVQ	Normal	0.37	1	0.54
Neural	60s 0%	Spectral	LVQ	Normal	0.37	1	0.54
Neural	30s 50%	Spectral	LVQ	Normal	0.36	1	0.53
SVM	60s 50%	Spectral	LVQ	Normal	0.31	0.99	0.47
Neural	30s 0%	Spectral	LVQ	Normal	0.3	1	0.46
Neural	60s 0%	Both	LVQ	Normal	0.28	1	0.44
Neural	30s 50%	Both	LVQ	Normal	0.24	1	0.39
Neural	30s 0%	Both	LVQ	Normal	0.24	1	0.39
SVM	30s 50%	Spectral	LVQ	Normal	0.21	1	0.35
SVM	30s 0%	Spectral	LVQ	Normal	0.19	1	0.32
SVM	60s 0%	Spectral	LVQ	Normal	0.18	1	0.31

Continued on next page

Table A.1 – continued from previous page

Classifier	Window	Feature	Selection	Sampling	Sens.	Spec.	S1 Score
SVM	60s 50%	Both	LVQ	Normal	0.15	1	0.26
SVM	30s 50%	Both	LVQ	Normal	0.15	1	0.26
Neural	30s 0%	Statistics	None	Normal	0.13	1	0.23
SVM	30s 0%	Both	LVQ	Normal	0.13	1	0.23
Neural	60s 0%	Statistics	None	Normal	0.1	1	0.18
SVM	60s 0%	Both	LVQ	Normal	0.1	1	0.18
Neural	60s 50%	Statistics	None	Normal	0.05	1	0.1
Neural	60s 50%	Statistics	RFE	Normal	0.04	1	0.08
Neural	30s 50%	Statistics	RFE	Normal	0.04	1	0.08
Neural	30s 0%	Statistics	RFE	Normal	0.04	1	0.08
SVM	60s 50%	Statistics	None	Normal	0.04	1	0.08
SVM	60s 50%	Statistics	RFE	Normal	0.04	1	0.08
SVM	60s 0%	Statistics	RFE	Normal	0.04	1	0.08
SVM	30s 50%	Statistics	None	Normal	0.04	1	0.08
SVM	30s 50%	Statistics	RFE	Normal	0.04	1	0.08
Neural	60s 50%	Statistics	LVQ	Normal	0.03	1	0.06
SVM	30s 0%	Statistics	RFE	Normal	0.03	1	0.06
Neural	30s 50%	Statistics	None	Normal	0.02	1	0.04
Neural	30s 0%	Statistics	LVQ	Normal	0.02	1	0.04
SVM	60s 50%	Statistics	LVQ	Normal	0.02	1	0.04
SVM	60s 0%	Statistics	None	Normal	0.02	1	0.04
SVM	30s 50%	Statistics	LVQ	Normal	0.02	1	0.04
SVM	30s 0%	Statistics	None	Normal	0.02	1	0.04
Neural	60s 0%	Statistics	RFE	Normal	0.01	1	0.02
Neural	60s 0%	Statistics	LVQ	Normal	0.01	1	0.02
Neural	30s 50%	Statistics	LVQ	Normal	0.01	1	0.02
SVM	60s 0%	Statistics	LVQ	Normal	0.01	1	0.02
SVM	30s 0%	Statistics	LVQ	Normal	0.01	1	0.02