MONASH University

The Analysis of Machine Learning Algorithm Performance on Long Term Data using Short Time Scale Analysis

Maurice Ntahobari

Masters of Philosophy

A Thesis Submitted for the Degree of Masters of Philosophy at
Monash University in 2023
School of Information Technology

Abstract

Epilepsy is a neurological disease that has serious impacts on a patient's daily routine, and involves more than two unprovoked seizures. The prediction of seizures can improve patient quality of life and alleviate the stress caused by the unpredictability of seizures. Seizure prediction is a challenging problem that has been studied for many years through the development of machine learning algorithms. However, the performance of current seizure prediction methods is not good enough for clinical usage and there is no clear stand out machine learning algorithm that is suitable for seizure prediction, especially for short-time scale feature analysis (sliding analysis windows < 60 minutes in duration). Again, most of the algorithm in the past are more complex and this makes them infeasible practically. This is due to that seizure prediction need run in small and wearable device for energy saving with less computational cost. In this thesis, we compare short-time scale analysis based seizure prediction algorithms evaluated on long-term intracranial EEG data (greater than 6 months in duration per patient). In the past the majority of seizure prediction studies involved short-time scale analysis, but nearly all of these studies looked at features and algorithms on short-term datasets, most not longer than 2 weeks and containing a paucity of seizures. This thesis looks at what the expected performance might be for these short-time scale analysis studies on long-term data. Short-time scale analysis is in contrast to the recent promising results of seizure forecasting based on long-term seizure cycles where features are calculated on time-scales longer than 60 minutes and even over periods or days, weeks and months. As such this thesis which focuses on short-time scale features analysis from 10 minutes window segment evaluated on long-term data and provides a measure of the reliability and performance levels of such analysis because the long-term data contains adequate seizure and interictal data for reliable evaluation. This thesis thus marks a reference point for what can be achieved with short-time scale analysis, from which new directions of analysis can be compared and contrasted to in order to advance the field of seizure prediction. As a starting point we implemented the second place algorithm from the 2016 Melbourne University AES/MathWorks/NIH Seizure Prediction Kaggle contest as a benchmark. We performed features reduction through the use of Hapley Additive exPlanation(ShAP) algorithm to reduce computational cost for both model training and evaluation time. The second place algorithm was chosen because it was the simplest and most computationally efficient algorithm of the top 3 contest algorithms. Using this algorithm and its very large set of short-time scale EEG analysis features from 10 minutes window segment as a starting point, we have explored the performance of 9 classifiers including the Extra Tree classifier used in the second place algorithm, on the long-term intracrianial EEG NeuroVista dataset, to statistically assess whether any of the algorithms is better than the others at the individual and group level. The

same features used for the second place algorithm were used by all classifiers considered. At the group level, the algorithms were evaluated for area under the receiver-operator curve (AUC) where the obtained average performance across all 15 patients was 0.569, 0.621, 0.557, 0.666, 0.689, 0.652, 0.657, 0.587, 0.621 for Support Vector Machine (SVM), Random Forest, Linear Regression, Extra Tree, Linear Discriminant Analysis (LDA), Adaptive Boosting (AB), Extreme Gradient Boosting, Nearest Neighbour and Gaussian Naive Bayes classifier, respectively. On average, all classifiers have achieved better results than chance prediction based on one sample t–testing (P-values$<0.05/9$). LDA achieved the highest average AUC, however, the statistical analysis at the group level based on Bonferroni correction shows that there is no statistically significant difference between the benchmark algorithm from the Kaggle contest and the other algorithms. A similar comparison was done at the individual subject level and the results show that different classifiers perform differently for each patient and the performance is better than chance prediction. This analysis suggests that a one size fits all short-time scale analysis seizure prediction algorithm is not likely to yield stand out results. This is largely due to the variability of a given algorithms performance across subjects. An additional clinically relevant evaluation on the extra tree classifier shows that seizures can be predicted in 13 out of 15 patients with minimal time in warning. As a whole this thesis makes it clear that short-time scale analysis seizure prediction methods can perform above chance, there is no consistently best algorithm across subjects and the best model needs to be chosen in a patient-specific manner. Moreover, the performance of short-time scale analysis seizure prediction does not appear to be at a clinically useful level. Further analysis of this kind may yield improvements but it is not currently clear how such improvements can be achieved without considering long-time scale analysis.

Declaration

I, Maurice Ntahobari, declare that this thesis is an original work of my research and contains no material which has been accepted for the award of any other degree or diploma at any university or equivalent institution and that, to the best of my knowledge and belief, this thesis contains no material previously published or written by another person, except where due reference is made in the text of the thesis.

Signature: _____

Print Name: Maurice Ntahobari

Date: 20/01/2023

Acknowledgements

I would like to thank Almighty God for the love and protection from my childhood and in all my previous studies.

Many thanks to my lovely family for their love and encouragement and their uncomplaining to allow me to leave them for three years and stay alone in my home town Rwanda.

So far so good. I would like to thank my main supervisor Dr Levin Kuhlmann for his unlimited support, guidance for three years which have helped me to clearly understand my research area, problem in the state of the arts seizure prediction and find the topic for this thesis. I would like also to thank my second supervisor Dr Mario Boley for his guidance and teaching initially in the field of mathematics for statistical analysis and programming in python. This has enabled me to implement the studied algorithms in python as well as performing statically analysis to find if the obtained results are statistically significant. The last not the least is my thanks to Dr Zhinoos Razavi Hesabi for her assistance in the algorithm implementation and preparation of dataset for model evaluation. She has made it possible to access the performance of the designed algorithm for more than one testing set and have helped me to draw a better conclusion.

Contents

List of Figures

List of Tables

Abbreviations

SVM	**S**upport **V**ector **M**achine
EEG	**E**lectroencephalography
RF	**R**andom **F**orest
ET	**E**xtra **T**ree
LR	**L**ogistic **R**egression
KNN	**K**-Nearest **N**eigbour
EGB	**E**xtreme **G**radient **B**oosting
AB	**A**daptive **B**oosting
LDA	**L**inear **D**iscriminant **A**nalisysis
AUC	**A**rea **U**nder **C**urve
ROC	**R**eceiver **O**perating **C** urve
AV	**A**verage
AED	**A**nti- **E**pileptic **d**ruges
ANN	**C**onvolution **N**eural **N**etwork
DFT	**D**iscret **F**ourier **T**ransiform
FFT	**F**ast **F**ourier **T**ransiform
FN	**F**alse **N**egative
FP	**F**alse **P**ositive
TP	**T**rue **P**ositive
TN	**T**rue **N**egative
IWSP	**I**nternational **W**orkshop **S**eizure **P**rediction
PSD	**P**ower **S**pectrum **D**ensity
RMS	**R**oot **M**ean **S**quare
TiW	**T**ime in **W**arning
WPF	**W**avelet **P**acket **F**eature

PCDM **P**reictal **C**hange **D**etection **M**ethod

RBCS **R**etrospective **B**est **C**hannel **S**election

Constants

The significant level alpha (α) = 0.05

Chapter 1

Introduction

1.1 Background

Epilepsy is one of the most common neurological diseases. It affects more than 50 million around the world according to the world health organization [4]. It is caused by brain abnormalities which result in electric discharges in the brain known as seizures. The unpredictability of seizure affects patients daily activities like not safely driving vehicles and not performing various sport activities like swimming independently. Seizure prediction systems are one of the approaches to manage seizures. This is done by predicting the future occurrence of seizures so that preventive mechanisms can be taken to save a patient's life from danger that can happen during the seizure state. These measures may include but not limited to taking medication, brain stimulation and stopping dangerous activities until the seizure period disappears. Besides saving patient lives, this can also alleviate side effects that would be caused by unnecessarily taking of anti-epileptic drugs even when they are not needed which is normally done regularly by epileptic patients.

Seizure prediction system development consist of the use of machine learning techniques to learn the patterns in electroencephalogram data so that the future seizure onset can be predicted and warn the patients and caregivers to take preventive measures before a seizure occurs.

Machine learning is a part of artificial intelligence which generally means giving the capacity a system to learn to find a solution to a problem by recognizing the patterns in a new data set based on the existing algorithms. In machine learning, a system is trained

to distinguish new objects in the future based on the information it already trained for. Machine learning algorithms are very useful and performs two main tasks of identifying pattern in data sets and generating solutions [5].

From the seizure prediction perspective, machine learning techniques are used for a binary classification problem to distinguish the main phases of seizures such as preictal and interictal phases [6]. The research has proven that the spatial temporal dynamic of brain activities is made of four main phases which are the phase between two seizures(interictal), the phase that precede seizure(preictal), seizure(ictal) and the phase that is directly after seizure known as postictal.

The main goal of a seizure prediction system is to predict seizure onset by discriminating the interictal from preictal phases [7]. This algorithm is developed following four main phases of data preprocessing, features extraction/selection, model training and evaluation [1].

In time series where noise and outliers affect the quality of a signal, a data preprocessing phase is important. This includes data filtering to eliminate unwanted signals as well as two more phases of feature extraction and selection. Feature extraction, is a process of extracting the possible characteristic of brain signal. Brain characteristic features can be extracted from long or short-time scale analysis windows. Here we define long and short time-scale features as features extracted from continuous electroencephalography (EEG) recordings of brain activity using analysis windows of more or less than an hour, respectively. Short time-scale features are more typical of machine learning applications to EEG because long-term data (i.e. days, weeks, months or years) is not so readily available. Recent evidence has, however, shown the importance of long time-scale features in seizure risk and prediction [8], [9]. These long and short time-scale features can be extracted before, during and after seizures. Even though it has been shown to be useful in seizure prediction, long-time-scale feature analysis is not concerned in this thesis since most of the seizure prediction model have been implemented in short-time-scale feature analysis for short-term data (less than 6 months in duration/usually less than two weeks in duration) as the predominant dataset.

Feature selection is a process of choosing relevant features to train a classification algorithm. These may improve the model performance, interpretability, computational

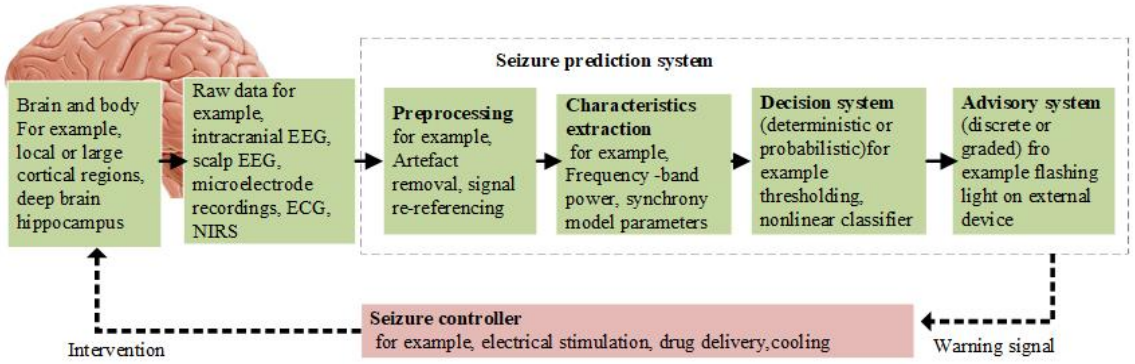


FIGURE 1.1: The overview of a seizure prediction system. This is the same as it has been presented in[1]

efficiency and reduce model complexity [10] which is useful for small and wearable devices such as seizure prediction devices.

A complete seizure prediction system is presented in figure 1.1. The figure illustrates a seizure prediction system in a closed loop fashion where EEG data are recorded from the brain, processed and characteristic features are extracted then a decision system is trained. Once a seizure is found, a warning alert is issued, then some intervention measures can be taken such as taking medicine and avoiding dangerous activities. In the next subsection1.1.1, we introduce the relation between seizure and epilepsy.

1.1.1 Seizure and Epilepsy

Epilepsy is a neurological disease that is characterised by recurrence of two or more unprovoked seizures. A seizure is characterised by an unexpected change in the electrical and chemical activity in the brain known as electric discharge. Epileptic seizures are the phases where brain activity is augmented and turns out to be abnormally harmonised. Seizures can be linked with various physical or psychological signs, subject to which side of the brain is affected. The main clinical symptoms of epileptic seizures are loss of consciousness, awareness and uncontrollable muscle movements where the person seems to fall asleep [11].

The common method to monitor the abnormal brain activity is the use of electroencephalograph (EEG) where two approaches are currently in use. The first is non- invasive and the second is invasive brain activity monitoring approaches. In non-invasive

approach, electrodes are placed on the surface of the head known as the scalp and record average electric activity from various predefined region of the cortex. This kind of EEG recording causes minimum risk to the patient and is mostly used for the diagnosis and treatment of epilepsy.

In the invasive brain activity monitoring approach, the recording is performed by introducing the electrodes in the brain region for brain activity monitoring. This approach has a higher signal to noise ratio because it is placed directly on few focal points of the brain surface. Therefore, it provides an adequate representation of brain activity and has been identified to be suitable in the study of seizure prediction [12]. In this thesis, we use long-term intracranial electroencephalogram (iEEG) NeuroVista trial data from [13] for both the model training and evaluation. In the subsection 1.1.2 we discuss on the existing techniques to control epileptic seizure(to treat epilepsy).

1.1.2 Treatment of Epilepsy

Different approaches are used for epilepsy treatment, including special diets, where the most common used method is the use of anti-epileptic drug medication (AEDs) that are able to eradicate or drastically reduce seizures[14]. However, researches show that for some patients, these medications are inefficient and are the source of serious side effects and develop drug resistant mechanisms. Therefore, the refinement of AEDs is necessary. The researches have shown that 30% of patients that use AEDs have no control of their seizures." It is accurate those patients are refractory to AEDs but this means that they still have some seizures at levels for which the AED side effects are livable [15]. Therefore, the effectiveness of the use of AEDs is questioned and other therapeutic mechanisms are needed. Currently, the use of brain surgery is an option to eliminated seizure recurrence. However, this can only be considered in the case where the source of the seizure can be located and removed but the localization of seizure focus is a challenging problem. It has been reported that only 58% of cases can only be treated by means of brain resection surgery [16]. This indicates that not all patients are suitable for brain surgery for epilepsy treatment. Deep brain stimulation has been shown to be useful in the treatment of epilepsy[17]. However, it also presents the same limitation as brain resection surgery to reduce seizures. With the current technology and a need for alternatives, scientists are pushing to find solutions to develop small, implantable and

wearable devices that can predict future seizures for epileptic patient treatment. This approach involves predicting the occurrence of a seizure before its onset. This requires the development of an implantable low computational cost device that is able to record and analyse brain data in a form of wave signal.

A number of closed-loop devices have been developed but in order to used this device in an effective way, the development of seizure prediction algorithms are required. This development requires analysis of the brain activities recorded using EEG and extracting the relevant brain characteristic features to train a machine learning model.

Currently, there are many projects for seizure prediction with the purpose to remind the patients about the occurrence of seizures based on the monitored EEG using implantable devices (e.g. subscalp EEG). This includes Epiminder[18] which is a project for home-based EEG monitoring. However, the applicability of this project in clinics is still under trial and the performance is poor compared to the expected performance. In the next subsection 1.1.3, we talk about the approach of EEG recording for brain data signal collection.

1.1.3 EEG Dataset signal Recording

The EEG data recording is done by placing the electrode on the patient head. This is know as scalp electroencephalogram and it is a non-invasive approach. The brain signal recording is also recorded using invasive approach where after clinical surgery, the electrode are placed directly on the patient's brain. The main difference between scalp EEG and Intracranial EEG is that scalp EEG is prone to most of the artefacts such as muscle movement artefacts but the intracranial EEG has less artefacts which makes this data suitable for seizure prediction model development. The figure 1.2 demonstrates the difference between scalp and intracranial EEG brain activity recording. Brain signal recording is challenging and previously patients needed to be hospitalized for at least 7 days for brain monitoring. However, with technology advancement, brain recording can be made at home and the patient may continue some of his or her activities such as taking a shower and sleeping comfortably which was not possible before [19].

The seizure prediction problem can be regarded as a classification problem which is concerned with the separation of states of preictal and interictal brain states before and

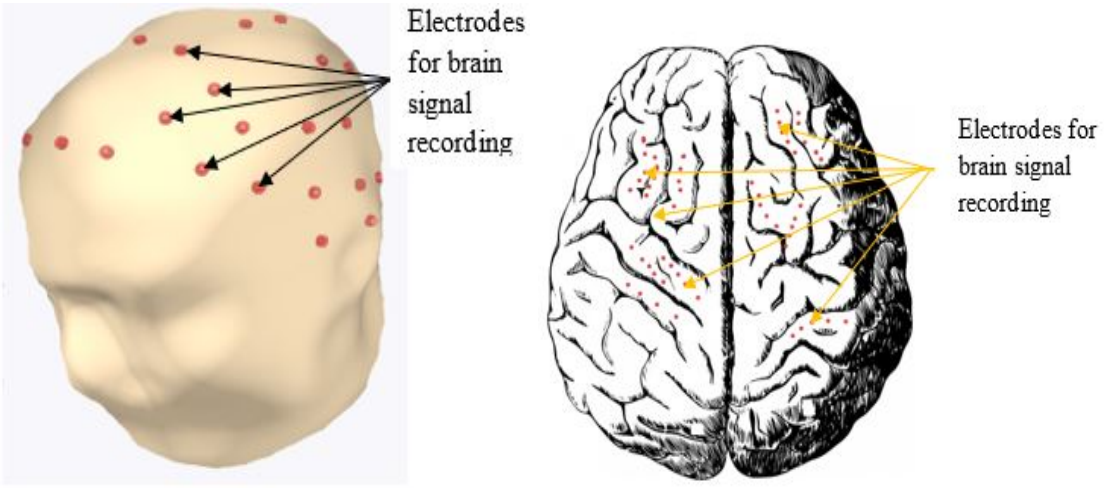


FIGURE 1.2: The brain activity signal recording process. The scalp EEG recording on the left side and intracranial brain signal recording on the right side. The concept is adapted from [2]

after ictal onsets. The development of such an algorithm requires one to distinguish the classification methods of supervised learning and unsupervised learning (clustering).

1.1.4 Supervised vs unsupervised learning

Machine learning is made of two main categories of supervised learning and unsupervised learning. These two categories differ from one another based on the objective and the problem at hand to be solved.

In supervised learning, a model is trained with the data and corresponding labels. It is mainly used in classification problems where a model is trained to distinguish objects based on the labels. This is the same case for seizure prediction model development where the algorithm is fed with the data and the corresponding label to indicate whether, the data is a preictal or interictal window. Thereafter, once a model is presented with new data, based on the data it was trained on, it makes a decision if it is a preictal or interictal brain state and issues a warning to the patient. This is different from unsupervised based model learning where, an algorithm is only fed with the data and once new data is found, the separation is done by considering different characteristics of the trained data. Let us give an example, suppose we are asked to develop a model that is designed to separate mangoes and oranges. The characteristics to be looked at are are: the shape, color and the skin. This type of learning is known as a clustering where objects are clustered or grouped based on those having the same characteristics. In this

thesis we are dealing with the supervised classification problem in the seizure prediction context to classify preictal window segments from interictal ones using machine learning techniques.

Machine learning algorithms and techniques have been applied in many fields especially in healthcare for diagnosis purposes and have been shown to be useful. Nowadays, the techniques are being upgraded to be used in the prognosis and prediction of diseases. Various studies have been carried out by different researchers to improve our understanding of applicability of machine learning in various fields especially in epileptic seizure prediction. However, most of these studies have not output the suitable results for real life application by clinicians due to the performance of the designed algorithms and their prediction time complexity for clinical evaluation (i.e. computational efficiency requirements and how that could impact monitoring device longevity).

1.2 Problem Statement

The unpredictability of seizures can be a source of serious issues for an epileptic patient such as head injury, sudden death, source of discrimination, anxiety and reduced productivity. These aspects decrease a patient's independence and safety. Again, the frequent occurrence of seizures causes school dropout and unemployment. Hence, generating a low annual income especially for the people in rural areas and in the country with low and middle income.

The treatment of epilepsy, uses either Anti Epileptic Drugs, brain resection surgery or brain stimulation. However, none of these approaches is suitable across all patients.

In state of the art seizure prediction, various models have been developed with the objective to produce a warning system to assist patients with being aware of the occurrence of impending seizures. Nevertheless, most of the existing models are inefficient due to their computational cost at prediction time and poor predictive performance where even the top models would have an unsatisfactory predictive performance in the final application context [20],[21]. Even though, different seizure prediction algorithms based on machine learning haven been developed and some of them achieve good performance, there is no stand out machine learning classifier that has been proven to perform better than others for seizure prediction.

1.3 Research objective

Considering the above problem statement in subsection 1.2, our objective in this thesis is to analyse the performance of various machine learning classifiers for seizure prediction based on short-time-scale window analysis features for long term data recording and statistically access whether any of the examined classifiers performs better than others or chance prediction for both the group and individual level. This has been done by using a large common set of short-time scale analysis features and comparing the performance of the algorithm in the second place in the 2016 Kaggle contest for seizure prediction with 8 other classifiers that have been used in the past for seizure prediction based on short-time-scale window analysis for short-term-data recordings. The second place algorithm is chosen because it has been proven to be the most computationally efficient algorithm among the first top three algorithms in the competition. This thesis puts a focus on computationally efficiency with the view to creating battery-saving algorithms for monitoring devices.

1.4 Research Questions

In this section, we introduce the research questions that we will address to meet our objective. Based on the problem statement presented in subsection 1.2 and our objective in subsection 1.3, the following is our research question:

What is the appropriate machine learning algorithm for seizure prediction modeling that can be implemented in a seizure advisory system for real time application? This question is important because currently there is no individual algorithm that has been identified to perform better for seizure prediction.

Different algorithms have been developed as an ensemble of more than one classifier and still the performance is low for its usability for a seizure prediction system. For this, we will evaluate the performance of various classifiers in terms of AUC and make comparison to find the classifier that outperforms other classifiers based on a short time scale window analysis of window segments that are less than 60 minutes. More specifically we use 10 minute analysis windows. These algorithm methods may not perform well but they

will at least form the benchmarks for future methods aiming to be both efficient and accurate.

Chapter 2

Literature Review

Machine learning techniques have been applied in different fields to make human activities easy. This includes the development of machines known as robots to perform heavy work assisted by human beings. The technology has expanded to the medical environment where it is used in patient disease diagnosis. In this case models are trained for recognizing future events based on the training information and integrate the model in devices to be used in performing specific activities. In this thesis, given the strong focus on seizure prediction performance here brief definitions of the most relevant performance metrics are given. For the model evaluation, we mainly focus on the model sensitivity and specificity (Equation 2.1). In the equations, TP and FP stands for True Positive and False Positive respectively while FN, TN stands for False Negative and true negative respectively while SENS and SPEC stands for sensitivity and specificity accordingly.

$$SENS = \frac{TP}{TP + FN} \ , \ SPEC = \frac{TN}{TN + FP} \tag{2.1}$$

The sensitivity of a seizure prediction model is very important. It shows the probability that a positive class, in this case (preictal state) is correctly classified while specificity indicates the probability that a negative class (interictal) is correctly classified. The goal in machine learning is to have a trained model with high true positive rate for low false negative rate.

Specific values for sensitivity and specificity arise from specific decision threshold values applied to the classifier output that determine if a window is preictal or not. By considering the full range of possible thresholds one can plot a true positive rate vs false

positive rate curve also known as the receiver-operator-characteristic (ROC) curve. The area under this curve (AUC) acts as a metric of overall performance of a classifier. It takes on values of 0 for poor performance, 0.5 for chance performance and 1 for excellent performance. AUC is the main performance evaluation metric used in this thesis as it takes into account both sensitivty and specificity.

AUC is one of the most commonly used evaluation metrics in machine learning, but in terms of clinically-relevant evaluation metrics for seizure prediction there are two other important measures we consider here. These are seizure sensitivity (evaluated as proportion of seizures correctly predicted, as a opposed to proportion of precictal windows correctly classified) and proportion of time in warning (TiW - which is the proportion of time that a patient is under warning relative to the duration of the whole recording).

This thesis, focuses on the application of machine learning in medical environments particularly in the prediction of epileptic seizures. The various machine learning based seizure prediction models have been developed for the purpose of producing a model that can be used in predicting seizures a time before its onset. However, finding appropriate features that characterize the brain activity before seizure occurrence has been a challenging problem. This was due to a lack of long term and uniform EEG datasets with enough seizures and appropriate statistical validation of each method[22].

Seizure prediction studies trace back to 1970 where univariate non-linear measures such as the lyapunov exponent have been used as a brain EEG characteristic [23]. After this first study, in 1981, researchers demonstrated that there are changes in brain activities before seizure occurrence[24] and these changes characteristics can be used as a sign of future seizure occurrence. Since then, different characteristic features, both linear and non linear measures, have been extracted and used to train a learning model for seizure prediction.

The number of researchers for seizure prediction using machine learning techniques rose in the year 2000 and after. These researchers have shown promising results and concluded that there is a dynamic change in brain activities before seizures and can be used as features to develop a real time seizure prediction system [25],[26].

The studies in [27],[28] have shown that the spatiotemporal brain dynamics during, and also likely before, epileptic seizures are patient specific and the EEG characteristic features vary from patient to patient. Therefore, patient-specific seizure prediction models have been shown to work better than general models [21].

In the literature, seizure prediction has dealt with significant controversy, as the researchers found that models were being developed using an inadequate amount of data (both in terms of seizure counts and EEG data duration) and it was not possible to reproduce the results of various algorithms when using different data sets [1],[29]. This is why the idea of organising seizure prediction competitions emerged out of the International Workshops for Seizure Prediction (IWSP). The Various IWSP meetings have been organized including IWSP3, IWSP4 in 2007 and 2009 respectively with the objective of providing a standardised process to compare the performance of algorithms developed on the same dataset. The first contests also involved inadequate amounts of data per patient, however, over time contests have included more data per patient.

In particular, the two most recent seizure prediction competitions were based on long-term continuous human or animal data. The first is the American Epilepsy Society Seizure Prediction Challenge that took place in 2014 and used short-term human and long-term animal intracranially recorded EEG (iEEG) [20]. The results for these competitions have shown the possibility of seizure prediction in canines. Another competition was organized by Melbourne University in 2016 and is known as Melbourne University AES/MathWorks/NIH Seizure Prediction Challenge [21]. In particular, this latter competition used long-term iEEG human recordings and contained many seizures per patient. The results of the contest was that seizure prediction appears possible in humans. The different algorithms have shown better results than chance prediction in terms of AUCs but their application is not ready for the clinical environment due to both the model performance and the model complexity based on the many features involved that make the model less interpretable and also potentially less computationally efficient.

In the aforementioned competitions, the participants were given the opportunity to use any machine learning algorithm and many simple features were extracted from 10 minute data window segments. Most of the algorithms were too complex and inefficient to be used practically given that they were computationally expensive at prediction

time. This renders them infeasible for applications with wearable/implantable devices, which is where they ultimately have to run. They also lack interpretability where it is not clear how models work, or what algorithm and features work the best. Therefore, finding relevant features to train a classifier as well as the best classification algorithm have been a challenging problem [30]. Partly due to this issue, the performance of the existing models is low and not suitable for clinical utility in all patients.

There are a multitude of features that have been used to develop seizure prediction systems such as those from univariate and bivariate measures using linear or non-linear approaches [31]. The research has shown that univariate measures performs better and are good features for seizure prediction but a combination of both univariate and bivariate measures gives promising results for seizure prediction model development. The research was not concerned with computational run-time for the features and as far as is known, all univariate and bivariate features are still relevant for seizure prediction model development.

To avoid hand crafted features, studies have applied neural networks for seizure prediction. Neural networks are a type of deep learning technique which learn features automatically from EEG to perform classification. There are different types such as Convolution Neural Network (CNN), Recurrent Neural Network (RNN) and Multi Layer Perception (MLP). In [32], a CNN has been used to classify preictal and interictal after extracting features of frequency and time domain using short-time Fourier transform on 30s EEG windows. The method was evaluated on three data sets: Freiburg Hospital intracranial EEG dataset, the Boston Children's Hospital–MIT scalp EEG data set, and the American Epilepsy Society Seizure Prediction Challenge data. The high performance achieved was sensitivity of 81.4%, 81.2%, and 75% with a false prediction rate of 0.06/h, 0.16/h, and 0.21/h accordingly. In general, the performance in terms of sensitivity is good for all subjects in all data set. However, these results rely on small data sets with few numbers of seizures per patient which puts into question the reliability of the performance measures. Nevertheless, the researchers argue that CNN performs well for seizure prediction and can identify relevant EEG features for preictal and interictal classification. We notice that the performance decreases as the total number of seizure get small. For instance, with Boston Children's Hospital–MIT scalp data set with 64 seizures, sensitivity is 81.2% while for Freiburg Hospital interictal EEG data set with 59 seizure, the sensitivity of 82.4% is obtained and for American Epilepsy

Society Seizure Prediction Challenge data set with 48 seizure, the sensitivity of 75% is achieved. Therefore, with CNN, high performance is achieved for big datasets with high number of seizure. In another study [33], 6 types of neural networks have been trained with 14 features to separate preictal from interictal brain states. The methods have been evaluated on single and multiple patients' EEG and accuracy of 99% and 98% was achieved respectively. However, since the proportion of preictal segments are not reported and accuracy is used as performance metric, the model can be overoptimistic and the reported accuracy can be for the majority class which in this case is the interictal class.

In general deep learning based algorithms are known to perform better for big dataset with enough seizures. They are also known to be computationally expensive due to number of layers from input to the output[34]. Therefore, given the focus on computational efficiency in this thesis deep learning approaches are not appropriate and were not chosen as one of the classifiers analysed in this thesis. This is because a seizure prediction algorithm need to be efficient with low computational cost model to run in a small and affordable device. Though it might still be efficient if the algorithm can be trained offline.

Most of the seizure prediction features have been studied on short time-scales and on short-term recordings as the predominately available EEG datasets are less than 2 weeks in recording duration. It was not until the world's first-in-man long-term trial of a seizure prediction device that recordings longer than two weeks were obtained. If a recording only lasts up to two weeks only a handful of seizures can be obtained per patient and it is difficult to train a reliable seizure prediction algorithm and this led to a lot of controversy in the field. The trial data provided recordings between 6 months and 3 years in duration enabling collection of a large number of seizures per patient required for development of reliable prediction algorithms. Since this trial, only a handful of studies have been performed using the trial dataset and a thorough study comparing all the features applied previously to short-term recordings has not been performed. Therefore, there is scope to explore the reliability of short time-scale features on the long-term recordings. Similarly, there is recent evidence highlighting the importance of a handful of features calculated on long time-scales for achieving state-of-the-art seizure prediction performance. Maturana and all in [9] suggests that the many features studied on short time-scales could also be studied on long-time scales to see which features and

which time-scales are the most important for seizure prediction. In the next subsection 2.1 and following subsections, we give an overview of the history of seizure prediction, recent studies on seizure prediction, seizure prediction studies involving the long-term NeuroVista trial dataset, the computationally-efficient preictal vs interictal classifiers that form the basis of the analysis in this thesis, and the large set of seizure prediction features that have been applied in the field and also are analysed in this thesis.

2.1 History of Seizure Prediction

Seizure prediction is not a new concept. The first research article have been published in 1970[23]. The paper used a small dataset. In 1980, a non-linear features in time series analysis was taken as features for seizure prediction while in 1990, the application of Lyapunov exponents was used as features for seizure prediction while correlation density and dynamical similarity index was considered in 1999 and 2001. In all the above studies, the authors were interested on the classification of pre-seizure period and did not include the evaluation of control recordings for interictal state. There is a list of researchers such as in [19],[29],[33],[35],[36],[37],[38] have been able to indicate the possibility of separating preictal (a state that precedes a seizure state) from interictal state (a state between two successive seizures) and these humans used features such as the correlation dimension [39], dynamical entertainment [40], accumulated signal energy [25] as well as phase synchronization[41].

Based on this aforementioned research, many studies have been conducted by various researchers using mostly, the extensive databases. However, surprisingly the performance of their models perform poorer than expected for the previous reports especially for the measures called correlation dimension as indicated in [42], the similarity index[43], and accumulated energy [44]. There was also a controversy among the studies on whether the previous results can be reproduced and on universal appropriateness of non-linear measures utilized to describe EEG time series. [45]. All the above studies were based on short-term EEG datasets with few number of seizures perhaps due to clinical reasons because for intracranial EEG researchers,the patients only got electrodes put in their heads temporarily for epilepsy surgery planning. In the case of scalp EEG, no one wants to wear the electrodes for too long due to stigma, discomfort or the sensors lose their fidelity/signal quality. Since, the first research on seizure prediction was done, most of

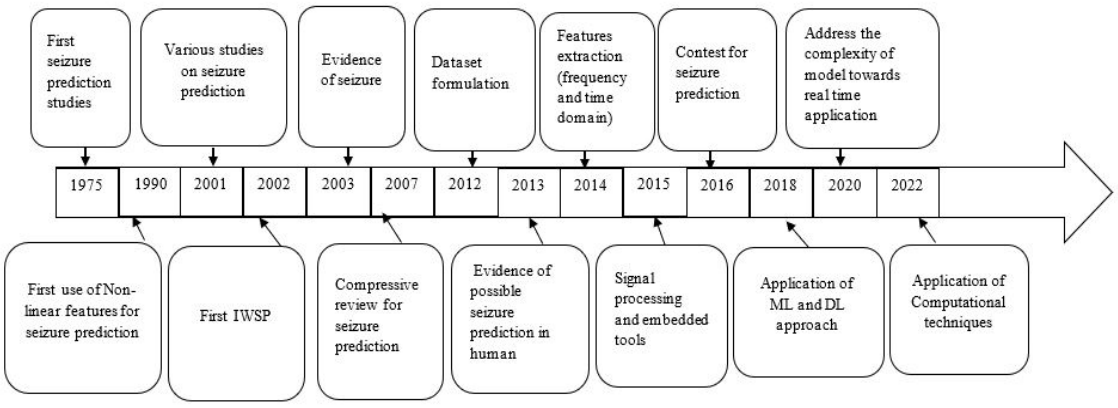


FIGURE 2.1: The chronological order of seizure prediction studies. This graph is a modified version from [3]

the following studies were based on EEG data recordings of 5-14 days that was provided by different epilepsy centers. This led to the first international workshop which took place in 2002 in Bonn University with the objective of getting various models and testing them on common agreed datasets and comparing their developed models[46]. In the conference most of the models showed low performance for the univariate features and the better performance was observed for bivariate and multivariate features[31],[47].

In the past decades, different features have been extracted from EEG data to characterise the brain behavior before seizure. Most of these features are based on statistical and spectral moments. The statistical moments include variance, mean, maximum, minimum, standard deviation, skewness and kurtosis while spectral features, are those which are based in the frequency domain, like spectral power, Fourier Transform(FT) such as Fast Fourier Transform(FFT) and Discrete Fourier Transform(DFT). The measures are computed from a time window of different sizes from EEG data channels [32].

Since 2007, many algorithms have been developed using various datasets and were based on small datasets which generate better performance with high sensitivity (true positive rate) and low false positive rates and /or small warning time portions. This is due to the small dataset, low numbers of seizures per patient or the discontinuity of data within these datasets and the limited amount of interictal segments in the dataset. The chronological order of history of seizure prediction studies is presented in the figure 2.1 .

In 2018, a complete review on seizure prediction have been presented in [1] and in its supplement material, and many databases and studies have been presented. The

research focused on those studies published from 2007 to 2018. The databases have been ordered in the decreasing order of their prospective nature, the number of seizure per patient and the amount of data per patient. However, most of the reviewed studies were based on short-term data with short time scale feature analysis. This can be seen in the table 2.1 and table 2.2. Some studies have been carried out using short-term EEG database while other have used (semi-) continuous data. Among all the presented databases, the NeuroVista study databases were the only long-term continuous EEG database. These were animals and Human databases from the 2014 and 2016 kaggle contests as well as NeuroVista intracranial EEG data recorded from 15 patients for 6 months to 3 years. The table 2.1 illustrates the databases that have been used to design and evaluate seizure prediction algorithms while table 2.2 shows the performance of the designed algorithm in terms of sensitivity and time spent in warning.

Almost all the above discussed researches perform better than chance prediction. However, the achieved performance levels are far away from being high enough to be usable in the clinical environment. This is due to model development being based on EEG signal data with small numbers of seizures. It is presented in the tables 2.1 and 2.2 that, if a dataset has a low number of seizures, the statistical performance evaluation measure increases while in the cases when the number of seizures is high in a dataset, the performance drop significantly[77]. Therefore, for the development of a trustable seizure prediction model the electroencephalogram dataset must contain enough seizure events. Moreover, preprocessing techniques for artefact removal and feature extraction and selection techniques are valuable to make conclusions regarding which model should be implemented and used in the clinical environment. In the subsection 2.2, we discuss on the recent studies for seizure prediction using big data set of long term recording of more than three months.

2.2 The Recent Studies for Seizure Prediction

The previous studies for seizure prediction were based on short-term data with recordings of less than a month. In 2013, the first-in-man trial dataset of long-term recordings in humans have been made available. This dataset is made up of recordings from 15 patients. The study in [13] using this dataset, has shown the possibility of seizure prediction in human. A hand full studies have been done with this data. In 2018,

TABLE 2.1: Seizure prediction history with the respective database, this was adapted from[1]. RFE means refractory focal epilepsy, TLE is temporal lobe epilepsy while NFE is natural focal epilepsy. Most of these studies focus on human data except for canine long-term NeuroVista device data.

Year	**Ref**	**Predictive characteristic**	**PC-DM**	**EEG type %**	**Epilepsy type %**	No.of Patients	Seizure per patient	EEG time per patient	Prospective?
Long-term canine and short-term human dataset									
2015	[48]	correlation and spectral power bands	SVM	iEEG	RFE	5(8) animals	(35.8) 22.37	(380.4)366	N
2016	[20]	Numerous features	Numerous algorithms	iEEG	RFE	5 animals and 2 humans	15.8	4033.9	N
Long-term human clinical trial									
2013	[13]	Average /Teager-Kaiser energy, line length	Decision tree,kNN, state machine	iEEG	RFE	11	81.7	6956.3	Y
2017	[49]	Spectral energy, line-length	Circadian-weighted logistic regression	iEEG	RFE	9	224.8	7677.3	N
2018	[50]	Spectro grams	Deep learning CNN	iEEG	RFE	10	281.7	14270.0	N
2018	[21]	Numerous features	Numerous algorithms	iEEG	RFE	3	211	3699	N
2020	[9]	critical slowing down features	n/a	iEEG	RFE	15	191.4	483	N
Freiburg database									
2010	[51]	Mean phase coherence, dynamical similarity index	AND operation, threshold	iEEG	RFE	8	19.1	182	N

Year	**Ref**	**Predictive characteristic**	**PC-DM**	**EEG type %**	**Epilepsy type %**	No. of Patients	Seizure per patient	EEG time h per patient	Prospective?
EPILEPSIAE database									
2012	[56]	34 algorithmic features	SVM	iEEG /EEG/ ECG	RFE	12	9	264.8	N
2013	[57]	22 univariate features	SVM	iEEG /EEG	RFE	10	8.6	138.8	N
2014	[58]	cross-frequency coupling	Threshold	iEEG	RFE	53	10.5	240.5	N
2014	[59]	22 Univariate features	SVM, RBF-NN, RBP-NN	iEEG /EEG	RFE	278	9.7	175.3	N
2015	[60]	Differences /ratios of linear univariate features	SVM	iEEG /EEG	RFE	10	8.6	138.8	N
2015	[36]	Bivariate spectral power	SVM	iEEG /EEG	RFE	24	7.6	148.5	N
2017	[61]	22 univariate features	Multi-class SVM	iEEG /EEG	RFE	216	5.6	77.5	N
Montreal database									
2015	[62]	(Scaling) cumulant, state similarity	Linear discriminant	iEEG	MTLE	17	10.3	92.1	N
Bonn database									
2016	[63]	Eigenvalue PCA, CNN/linear prediction error, crossing levels	Threshold	iEEG	RFE	20	5.2	5-575	N

studies used the data to look at seizure prediction based on deep learning convolution neural networks and circadian analysis combined with Bayesian forecasting. In parallel, this work was accompanied by seizure prediction kaggle competitions using this data as well as data obtained from animals using the same device where different learning models have been developed using various numbers of features. The most commonly used features in the competition was spectral band energy and entropy based features. The results for all those studies have demonstrated the performance above chance prediction where the highest achieved result is AUC score of 80% [20] but these results require

Year	**Ref**	**Predictive characteristic**	**PCDM**	**EEG type %**	**Epilepsy type %**	No. of Patients	Seizure per patient	EEG time h per patient	Prospective?
FSPEEG database									
2009	[64]	6 linear /nonlinear synchrony	Logistic regression, CNN, SVM	iEEG	RFE	21	4.2	27.7.0	N
2011	[65]	Univariate spectral power	Cost-sensitive SVM	iEEG	RFE	18	4.4	≈28.5	N
2012	[66]	6 nonlinear features	Rule-based	iEEG	RFE	11	4.5	≈28.7	N
2012	[67]	Univariate /multivariate correlation	SVM	iEEG	RFE	19	4.4	≈28.0	N
2013	[68]	Epileptic spike rate	Threshold	iEEG	RFE	21	4.1	27.7	N
2014	[69]	Hilbert–Huang mean phase coherence	Threshold	iEEG	RFE	10	5	≈27.1	N
2014	[38]	34 Univariate features	Voting SVM	iEEG	RFE	21	4.2	≈4.2	N
2014	[70]	N-gram EEG patterns	Fixed /dynamic threshold	iEEG	RFE	21	4.1	27.7	N
2014	[71]	Neural mass model parameters	Rule-based	iEEG	RFE	21	4.1	27.7	N
2015	[72]	44 spectral power features/ratios	Linear SVM	iEEG	RFE	18	4.4	23.7.0	N
MIT database									
2015	[73]	44 spectral power features/ratios	Linear SVM	EEG	PE	17	4.6	38.1	N
Vancouver /MIT database									
2013	[74]	Zero-crossing intervals	GMM /threshold	EEG	TLE, eTLE, PE	20	4.3	28.1	N
Pittsburgh /Charleston database									
2011	[75]	Regularity statistic	Dynamic threshold	EEG	RFE	52	2.0	36.06	N

TABLE 2.2: Performance information for canine and human seizure prediction studies. Se also in[1]

Year	**Ref**	**In-sample parameter optimization**	**RBCS**	**Out-of sample testing %**	**Sensitivity %**	False positives per hour	Warning time portion %	Assumed pre-ictal period(min)	Prediction time-(min)	Statistical validation method
Long-term canine and short-term human dataset										
2016	[20]	N	N	Y	50-70	n/a	n	n/a	n/a	None
2015	[48]	N	N	Y	40-70,30-100	3.57,4.13 /day	1.39,1.48	n/a	n/a	periodic or random Poisson
Clinical trial database										
2013	[13]	N	N	Y	54-100	n/a	3-41	n/a	114	Random poisson
2017	[49]	N	N	Y	45-78	n/a	12-41	30-60	n/a	Random Poisson
2018	[50]	N	N	Y	69	n/a	27	n/a	< 60	Random Poisson
2018	[21]	N	N	Y	50-60	n/a	10-30	60	n/a	Random Poisson
2020	[9]	N	N	Y	50-80	n/a	$6 \times 10^4 - 17$	n/a	n/a	None
Freiburg database										
2010	[51]	Y	N	Y	43.2	0.15	n/a	10-60	n/a	Random poisson
New Brunswick database										
2013	[52]	N	N	Y	73	n/a	33	180	n/a	Periodic or Random poisson
2017	[53]	N	N	Y	71-80	n/a	3-25	30	9.3-14.3	Periodic or Random Poisson
Gainsville database										
2006	[54]	Y	N	N	80	0.56-,0.12	28,32	30,180	13,90	Periodic or Random Poisson
Freiburg contest/Melbourne database										
2010	[55]	Y	Y	Y	50-88	0.64-4.69	n/a	<15	n/a	Random Poisson or alarm time surrogates

Year	**Ref**	**In-sample parameter optimization**	**RBCS**	**Out-of sample testing %**	**Sensitivity %**	False positives per hour	Warning time portion %	Assumed preictal period-(min)	Prediction-time (min)	Statistical validation method
EPILEPSIAE database										
2012	[56]	N	N	Y	2-81	n/a	n/a	20-60	n/a	Seizure time surrogates
2013	[57]	N	N	Y	73.9	0.15	n/a	10-40	n/a	Random Poisson
2014	[58]	N	N	Y	30-100	0.08-0.72	n/a	30-60	n/a	random poisson
2014	[59]	N	N	Y	> 50	< 0.15	n/a	10-40	n/a	None
2015	[60]	N	N	Y	60.9	0.11	n/a	10-40	n/a	Random Poisson
2015	[36]	N	N	Y	75.8	0.1	n/a	30	n/a	Random Poisson
2017	[61]	N	N	Y	38.5	0.2	n/a	28	n/a	Random Poisson
Montreal database										
2013	[76]	N	N	Y	81.4	0.15	30.5	n/a	50.6	Random Poisson
2015	[62]	N	N	Y	80.5	0.15	25.1	n/a	59.4	Random Poisson
Bonn database										
2016	[63]	Y	N	Y	n/a	n/a	n/a	20-120	n/a	Seizure time surrogate
FSPEEG database										
2009	[64]	N	N	Y	100	0	n/a	120	2-99	Random
2011	[65]	N	N	Y	92.5	0.2	9.5	30	n/a	Random Poisson
2012	[66]	N	N	Y	79.9-90.2	0.17,0.11	4	30,50	13,24	Periodic or Random Poisson
2012	[67]	N	N	Y	86	0.03	3	30	n/a	2-class ROC chance
2013	[68]	N	N	Y	72.7	0.11	n/a	50	49.7	Random
2014	[69]	Y	Y	Y	70	0.15	n/a	40	n/a	Random Poisson
2014	[38]	N	N	Y	89-93	n/ah	n/a	<20	n/a	Baseline or Random Poisson
2014	[70]	N	N	Y	54-94	0.01-0.22	n/a	20	n/a	Random Poisson
2014	[71]	N	N	Y	83,90	0.16,0.12	2.7	30,50	15.4,27	Periodic or Random Poisson
2015	[72]	N	N	Y	100	0.03	n/a	16-54	n/a	2-class ROC chance

Year	**Ref**	**In-sample parameter optimization**	**RBCS**	**Out-of sample testing %**	**Sensitivity %**	False positives per hour	Warning time portion %	Assumed preictal period(min)	Prediction time-(min)	Statistical validation method
MIT database										
2015	[73]	N	N	Y	98.7	0.05	n/a	3-78	n/a	2-class ROC chance
Vancouver/MIT database										
2013	[74]	N	N	Y	88.3	0.16	n/a	40	22.5	Random Poisson
Pittsburgh/Charleston database										
2011	[75]	N	N	Y	68.3	0.235	n/a	20	n/a	Random

improvement.

Even though, the highest result outcomes have been achieved in the competition, most of the contestants have used many features and ensemble of more than one classifiers which renders the model to be computationally expensive and takes lot of time during model training. This renders the model to be not suitable in real life application in seizure prediction since a model needs to run in a small implantable/wearable device with less computational requirements to save battery life. Nevertheless, the results in the competition as well as the those in the studies in long-term intracranial EEG recording are promising and gives the next path for seizure prediction studies.

Recently, the research in [9] have shown good performance compared to the results for other studies using human NeuroVista long-term EEG recordings and demonstrated signal autocorrelation and variance as the best biomarker for seizure prediction. However, for some patients, the performance is still poor. With the first-in-man seizure prediction trial, for the best subject the algorithm achieved 100% sensitivity with 3% time in warning while several other patients achieved poor performance, e.g. one patient achieved sensitivity of 17% with 41% time in warning. The full description of the data recording are described in [13]. The three patients with the lowest performance were chosen as contest data and made publicly available to be used in the algorithm development. In this sense, many algorithms have been developed to improve the original trial performance through a Kaggle competition[21]. Thereafter, epilepsyecosystem.org was created, which is a platform that offers the opportunity to researchers and specialists to

develop and submit their models for evaluation in order to achieve further improvements in seizure prediction performance using the contest dataset.

2.3 Seizure Prediction on Long-term NeuroVista Dataset

In 2013, the first-in-man NeuroVista Seizure Warning System device trial dataset was recorded using intracranial electroencephalography (iEEG) from 15 epileptic patients. Prior to this study only short-term data recordings were available. This study opened the door to long-term data recordings between 6 months and 3 years in duration, thus providing an adequate number of seizures for the development of reliable patient-specific seizure prediction algorithms. In the original trial study, 288 short-time scale analysis features were extracted and with backward elimination techniques, only 16 features were used for classification. The results in term of sensitivity showed that seizure prediction is possible in humans [13]. Since this trial various competitions and research publications have tried to push this field forward and make seizure prediction viable for all patients by attempting to gain improvements in performance with the human trial dataset and a similar canine dataset obtained using the same NeuroVista device in animals.

In 2014, various contestants have used a canine and human iEEG dataset to develop models for evaluation in a Kaggle contest [20]. The canine data were also recorded with the NeuroVista device. The highest AUC score was 83% . In [21], the data from the first-in-man NeuroVista trial was used in a second Kaggle contest and the winning algorithm achieved an AUC score of 85% on the public test set with sensitivity of 0.58. In parallel with these contests, the same human NeuroVista trial data has been used to study seizure prediction with decision tree and K-nearest neighbor (KNN) classifiers [13] where the average results for all 15 patient was around 61% for sensitivity and 27% time in warning (TiW).

In [78], the detection of probable seizure activities was done with the use of the preictal spike rate as the indicator that the brain state is approaching a seizure and found that spike rate is not a reliable bio-marker for all patients, as it was demonstrated that the spike rate increases before seizures in 9 out of 15 patients and decreases for the remain 6 patients. In another research using circadian-modelling and logistic regression [49], the average sensitivity of 62% for all 15 patients with time in warning of 21% on average.

The deep learning convolution neural network (CNN) in [50] achieved a sensitivity of 69% with the portion of time in warning of 27%, while the recent publication in [79] which is an ensemble of algorithms from the 2016 kaggle competition achieved an insignificant improvement of performance which was between 0.5% and 1.5% for patient specific seizure prediction.

The table 2.3 illustrates the comparison of the algorithms' results for the study done with the Neurovista long-term human iEEG dataset. It is clear that performance improvement is still an open question to advance the research in seizure prediction. Again, critical slowing features improved significantly the model performance to almost all patients. However, more improvement is required to make seizure prediction clinically viable.

Each brain EEG data signal is different from patient to patient. This is why the similar set of features fed in the classification algorithm may perform differently. In the past and recent studies, various algorithms have been developed considering all the features especially in the 2016 Kaggle contest with the second winner among the top teams of the contest[21]. The second place algorithm seems to be the most computationally efficient algorithm among other algorithms. In this algorithm development, the author called features, long time-scale features if they were computed on 10 minute window segments, while short-time scale features were taken as summed energy in the standard spectral bands computed from 30s window segments with overlap. However, in this thesis, all features computed in the second place algorithm in the seizure prediction competition in [21] are considered as features on short time-scales since they are computed from a window segment that is less than an hour.

In the second place algorithm, there was 2264 features in total extracted from 16 channels for each patient. The algorithm achieved on average AUC of 79.9% and was more computational efficient than other developed algorithms in the competition and therefore is of great interest when it comes to developing efficient algorithms for wearable/implantable devices. The algorithm has been enhanced in [3] in terms of performance improvement and computational efficiency by reducing the number of features to 208 features and achieved the average AUC of 74% with a computational efficiency improvement factor of 3.25 compared to the second place algorithm in the Kaggle competition.

In this thesis, the same features were applied and via stratified cross validation different sample sets were fed to the classifier and model parameters were tuned via a grid search approach to find the best classification algorithm that performs better for epileptic seizure prediction.

The Statistical analysis was done to find if on average there is significant different in performance at group level where the bench mark algorithm is compared to other classifiers and by chance prediction.

P	[13]		[50]		[49]		[49]		[49]		[21]		CW-Kaggle		[9]	
	Sens	TiW	Sens	TiW	Sens	TiW	Sens	TiW	Sens	TiW	Sens	TiW	Sens	TiW	Sens %	TiW %
1	0.77	27	0.65	21	34	27	0.54	27	0.61	27	N/A	N/A	N/A	N/A	**68**	7
2	1	31	0.74	11	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	**75**	7
3	0.45	29	0.71	53	0.36	29	0.53	29	0.55	29	0.66	29	0.60	26	N/A	N/A
4	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	**64**	2
6	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	**72**	2
7	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	**76**	21
8	0.62	28	0.77	32	0.58	28	0.71	28	0.76	28	N/A	N/A	N/A	N/A	**67**	14
9	0.17	11	0.83	43	0.28	11	0.29	11	0.45	11	0.39	11	0.52	11	**75**	10
10	0.51	17	0.68	32	0.36	17	0.38	17	0.52	17	0.48	17	0.53	17	**69**	13
11	0.39	15	0.78	18	0.43	15	0.57	15	0.58	15	N/A	N/A	N/A	N/A	**83**	17
12	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	54	0.2
13	0.50	28	0.70	21	0.61	28	0.78	28	0.76	28	N/A	N/A	N/A	N/A	**64**	14
14	1	3	0.42	2	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	**75**	60
15	0.71	21	0.59	37	0.71	21	0.51	21	0.60	21	N/A	N/A	N/A	N/A	**88**	2
Avg	0.61(0.52)	21(22)	0.69(0.71)	27(32)	(0.47)	(21)	(0.57)	21	(0.61)	(21)	N/AN/A	N/A	N/A	N/A	**72(71)**	8(9)

TABLE 2.3: The seizure prediction models that have been developed with the long-term human NeuroVista EEG dataset. The average in parenthesis ignores patients 2 and 14 and the values in bod shows the better sensitivity than the previous studies

In the following subsection 2.4, let us discuss on the most algorithms that have been implemented for seizure prediction model development.

2.4 Preictal vs Interictal Classification Algorithms

Many classification algorithms have been studied for seizure prediction using short-term data recordings. The highest performing results have been achieved for different classifiers for KNN, SVM and extra tree classifiers with tuned parameters. In this section, we will discuss the classification algorithms that we would like to study the predictive performance for long-term data using short-time scale analysis. These algorithms are discussed in the following subsections 2.4.1, 2.4.2,, 2.4.4, 2.4.5, 2.4.6, 2.4.8 and 2.4.9.

2.4.1 Support Vector Machine (SVM)

The application of SVM is popular in the field of epileptic seizure prediction and is known to give positive output results in other areas in finance and genomics [80]. Many studies have applied different types of SVM for seizure prediction. In[81], SVM was applied on a animal canine data set [82]. In the study, the predictive model was evaluated for true positive rates (sensitivity) and false negative rate. The model in this study achieved a sensitivity between 90-100% and false negative(FN) of 0% - 0.03%. For all variations of SVM, false positive rates were almost zero while FN was in the range of 10-29%. The findings show that linear SVM is the best choice for dataset with low numbers of seizures (in the range of 7-20). Furthermore, a cost sensitive SVM was applied in combination with multiple features of spectral power to achieve high sensitivity and specificity. The algorithm was evaluated on the Freiburg iEEG data set with 18 patients of 20 who had at least three or more seizure periods. The algorithm achieved a high sensitivity of 97.5% and a false positive rate of 0.27%. Even though, non-linear SVM performs well in many patients, it is very computationally expensive and can be replaced by linear SVM which is less computationally expensive with little decrease in performance. However, authors acknowledge that the proposed model approach may not be applicable for patients whose seizures change fast in time or for patients who experience many various types of seizures. In [83] a seizure prediction model using polynomial SVM of degree 2 was proposed and evaluated on intra-cranial EEG (iEEG) from

the American Epilepsy Society Seizure Prediction Challenge database. The model was patient specific, had low complexity and low power consumption. This was achieved due to the utilization of few number of features obtained by feature and electrode selection using classification and regression trees. The baseline method used RBF SVM with 275 features and achieved AUC of 0.9975 while their proposed approach used 50 features and achieved AUC of 0.9929. However, in their study a small number of data was used to train the model with few number of seizures and all features have been used to train the model. Therefore the data leakage can be an issue which may lead to model over fitting and can be computationally inefficient. An ensemble of SVM and other classifiers has been also applied by three of the top eight teams including the winner of the 2016 Kaggle competition for epileptic seizure prediction and proved to yield AUC performance between 80% and 85%[21]. Nevertheless, the algorithm relies on many features which makes the algorithm complex and in-efficient for its usability.

2.4.2 Random Forest

A Random Forest classifier is an ensemble of several individual decision trees. It works by constructing many decision trees at the training phase where each individual decision tree is considered as a classifier. The overall classification results is obtained by applying a majority voting approach. It is a supervised learning algorithm which is an ensemble of tree classifiers. The method has been proven to perform well in seizure detection [84] but few studies have applied the classifier for seizure prediction. In [85], Random Forest in combination with wavelet packet features, sub-band energy ratio and three wavelet entropy features have been used to make a seizure prediction model. The model was evaluated on the CHB-MIT EEG database. The performance of random forest outperformed SVM, K Nearest Neighbour and Linear Discriminant Analysis and its performance was evaluated in terms of accuracy [85] where an accuracy of 84% was achieved. Nevertheless, the CHB-MIT EEG database used to evaluate the model performance is a small data set with imbalanced classes. Therefore, accuracy was not a good performance measure to rely on and a class imbalance method was not applied. Hence, the study for long-term EEG data is required and imbalance techniques need to be applied to improve the performance. This has been done in the aforementioned seizure prediction competition but in combination with SVM, extra-trees and gradient

boosting and achieved AUC performance of 80% on public leaderboard of the 2016 Kaggle contest [21]. Moreover, many features were used and it is hard to identify which features or classifier is influencing the model's performance. Hence, the performance of random forest alone needs to be studied with useful selected features. The classifier is mostly used for developing a model for seizure detection and proved to perform better in terms of sensitivity [84],[86]. However, there are not many works done with the purpose of seizure prediction.

2.4.3 Logistic Regression

Logistic regression is one of the types of linear classification algorithms which outputs the value of 0 and 1. It is defined by the equation 2.2 where P stands for the probability that a class 1 event occurs while β_0 is the intercept and the β_i are the coefficients of the n features.

$$P(x) = \frac{1}{1 + e^{-\beta_0 + \sum_{i=1}^{n} \beta_i x_i}} \tag{2.2}$$

The algorithm is simple and generally said to be useful for efficient low computational requirement devices, model interpretability and performance improvement when it is regularized with lasso regression [87]. However, few studies have used logistic regression in the development of epileptic seizure prediction approaches. In the research in [88], logistic regression was applied with a discrete wavelet transformation features and obtained an AUC performance of 0.85. However, this model was evaluated on a small dataset of only 22.8 hours possibly with few seizure periods. This is why its performance was better. The comparison with their neural network model in [88] shows that logistic regression performs poorly on the same data set. In circadian logistic regression[49], logistic regression was applied on 9 patients' EEG data recorded for 320 days. The model performed better in almost all patients with the highest sensitivity of 76% in two patients.

2.4.4 Extra Tree Classifier

The extra tree classifiers are a tree based classifier where the splits at each feature is selected based on its impurity values which is calculated at each node where the feature with less impurity is selected. The algorithm is based on supervised machine learning

and made of nodes with the root , internal and leaf node. It is used for classification and regression tasks and work based on answering true or false to a question. The resulted structure of a tree based algorithm is in the form of a tree made of different nodes where the root node is the starting nodes while the leaf node determines the final classification category. Since the feature splits are selected based on the calculated impurity values, the algorithm is less computationally expensive which makes it suitable for low computational resource based seizure prediction. The algorithm has been implemented for seizure prediction in [3] and by different contestants in the seizure prediction competition in [21] where the results are better than other classifiers. In particular, extra trees was the classifier used in the second place algorithm that forms the benchmark model in this thesis.

2.4.5 Linear Discriminant Analysis

LDA is a classification algorithm that is based on a supervised learning method. It works by finding the linear combination of different features to distinguish two or more classes. The technique has been used for patient specific seizure prediction and demonstrated good performance of 98% for sensitivity and 0.39% of false positive rate [35]. However, to our knowledge,the algorithm performance is not yet studied for long term continuous recordings. In this thesis, its performance is compared with other classifiers.

2.4.6 Adaptive Boosting

The Adaboost classifier is a type of supervised learning learning algorithm that links together the performance of various weak classifiers to improve their performance. In its functionality, the weak learner is trained using a subset of a dataset. Thereafter, the misclassified data sample of a weak learner is given a high weighting. The high weighted samples are then used for training in the next weak classifier training. The objective of the Adaboost classifier is to create a learning classifier that takes care of the misclassified samples in the previous classification phase. The functionality and equations of Adaptive boosting are found in the following article [89]

2.4.7 Extreme Gradient Boosting

Extreme gradient boosting works using decision trees in a boosting process that starts with the first tree developed in the model. It works by improving the quality of fit for the base learner. Here, the boosting in gradient means the method of modifying the weak learners relative to the active ones. Doing this, each new tree to be trained is fitted to the modified version of the original data set. With the gradient boosting techniques, many learning models are trained gradually where the first trained model performance is improved in the second trained model which is improved by its next model and so forth. Each weak model learns from the previous models misclassification error[90]. Each model works to reduce the misclassification error and loss function to achieve more accurate predictions.

2.4.8 K Nearest Neighbor

The K Nearest Neighbor is a supervised learning algorithm that is applicable for both classification and regression problems. It is simple and easy to apply. The decision for a k nearest neighbor depends on the number of neighbors data points. The distance from the k points to other points is computed. Thereafter, the labels of each point are determined by the average class of the k nearest points.

2.4.9 Gaussian Naive Bayes

The Gaussian naive Bayes is a type of supervised machine learning based classifier that follows the Gaussian distribution. It is also based on the Naive Bayes theorem. It is considered as a simple supervised learning classifier with high functionality. It is mainly known for being able to work with high dimensionality datasets and can also be useful for complex classification problems. Gaussian Naive Bayes is specifically applicable for continuous data such as time series data and that the data has a Gaussian distribution shape [91]. For all these classification algorithms, various features have been used to train and evaluate the developed learning model. In the next section, we will discuss on the features Implemented for epileptic Seizures prediction the features. These features have been studied in [3] and in [21] and proven to be useful for epileptic seizure prediction.

2.5 The features Implemented for Epileptic Seizures Prediction

In this subsection, we discuss on the features used in the 2016 Kaggle competition especially in the second place algorithm in the contest[21] and in [3]. This is because, the algorithm developed using these features was the most computationally efficient method in the contest and still achieved high AUC performance. In particular, the following set of features formed a basis of 2000 features in the second place algorithm and these features map to a large proportion of the features that have been applied to seizure prediction in the past.

2.5.1 Mean and Variance

The mean and variance of the EEG signal amplitude indicate the information about the location and span of the amplitude distribution of a time series x_i. The mean is defined as in the equation 2.3 while the variance is defined as in the equation 2.4.

$$\mu = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{2.3}$$

$$\sigma^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i)^2 \tag{2.4}$$

These two features have been used as uni-variate linear measures in different seizure prediction model developments in combination with other uni-variate and bi-variate statistical moment features to extract the statistical distribution of the EEG signal[31]. The performance of the algorithms have shown to perform better than chance prediction.

2.5.2 Skewness and Kurtosis

Skewness and kurtosis provide the information on the shape and distribution of a time series[29]. The skewness is equal to zero when the amplitude distribution is symmetric and non-zero otherwise. On the other hand, it is negative or positive respective to the right and left-skewed amplitude distribution whereas kurtosis measures the level of peakedness of the amplitude distribution of an EEG time series[92]. The skewness is

defined as in equation 2.5 while kurtosis is defined as in equation 2.6 where σ is the standard deviation

$$\chi = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{x_i}{\sigma}\right)^3, \tag{2.5}$$

$$\kappa = \left[\frac{1}{N}\sum_{i=1}^{N}\left(\frac{x_i}{\sigma}\right)^4\right] - 3. \tag{2.6}$$

2.5.3 Standard Deviation

The standard deviation is defined as in equation 2.7. It indicates the degree of deviation from the mean value in a dataset. This measure was taken as a feature given that the literature [93] indicates that standard deviation increases in brain activity during seizure and this is the source of higher deviation from the mean. As such, it may also be a good aspect to consider in seizure prediction system modelling. In this project, the standard deviation is regarded as the variability of EEG data and is computed for each one of the channels among 16 channels in each EEG 10 minutes window segment. In some cases, the standard deviation can be replaced by the signal variance. It looks at how data are spread out from the mean by computing the square root of the variance (see equation 2.7), while the variance indicates the average degree at which every point differs from the mean. It is computed by making the average of the squared differences from the mean. In state-of-the-art seizure prediction, both measures have been indicated as useful features for seizure prediction [22].

$$Std_k = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \overline{x}_k)^2} \tag{2.7}$$

2.5.4 Zero-crossing

The Zero-crossing feature is a well known feature in signal processing. It is a point where the sign of a mathematical function changes from positive to negative and vice versa . It is indicated by an intercept of the x-axis (zero value) in the graph of the function. This is shown in the figure 2.2

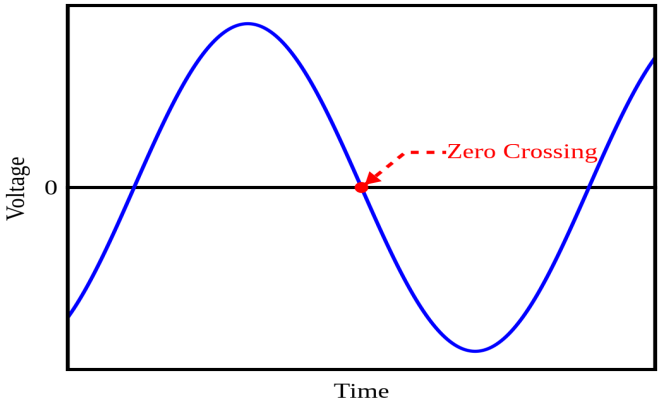


FIGURE 2.2: zero crossing values is the points of intersection on the x-axis when the signal changes from positive to negative sign or vice versa. [1]

The equation for the zero crossing value is described as in the equation:
$ZC(\mathrm{i}) = \frac{1}{2W_L} \sum |sgn[x_i(n)] - sgn[x_i(n-1)]|$, where $sgn(.)$ is the sign of a function. This means,

$$sgn[x_i(n)] = \begin{cases} 1, & x_i(n) \geq 0 \\ -1, & x_i(n) < 0 \end{cases}$$

The zero crossing value has been shown in some studies as a useful element in seizure prediction to discriminate preictal from interictal seizure states [26]. It is defined as the number of times that the signal changes its sign. These numbers may play a big role in seizure classification. In the study in [94], the zero-crossing value was applied as features and the accuracy of 94% on average has been achieved when the evaluation was done with CHB-MIT dataset for various classifiers. However, the choice of these features is not mentioned in the paper and the assessment of the influence of each of the features in the model performance is important. This importance can be accessed through different feature selection techniques.

2.5.5 Hjorth Parameters

Hjorth parameters are time domain statistical parameters. These parameters are activity, mobility and complexity. The parameter was first introduced in [95] and has been proven in[96] to produced better results in epileptic seizure prediction modeling with improved computational efficiency (decreased computation time). The important of these features was demonstrated in the competition for epileptic seizure prediction in [21] where four out of 8 teams have implemented at least one parameter of Hjorth. In

this study we will extract mobility and complexity because activity is about variance and this have been already extracted. The mobility is defined by the equation 2.8 where M stands for mobility while the complexity is defined in the equation 2.9 and C stands for complexity.

$$M = \sqrt{\frac{\sigma^2(y'(t))}{\sigma^2(y(t))}} \tag{2.8}$$

$$C = \frac{M(y'(t))}{M(y(t))} \tag{2.9}$$

2.5.6 Root Mean Square

The Root Mean Square (RMS) is defined as the square root of the mean of the squared values of a time-series. It is used to measure the magnitude of a signal in different frequency bands and regarded as a good signal power estimator in EEG frequency bands [97],[98]. In the study [99], an automatic seizure detection model has been developed considering RMS in combination with other features and achieved a better predictive performance of 84% in terms of sensitivity. In another study [100], the use of RMS alone achieved the best performance of 77.7% in terms of sensitivity in the training dataset compared to 21 other features studied. It beats minimum and maximum amplitude features by 3.6% which came after RMS while the combination of all 21 features including RMS achieved the best performance of 81.75%. RMS usefulness was also demonstrated in[101] where it has been identified as the best feature compared to 21 other features. However, it was also observed that there is a drop of 18% of performance in sensitivity when the best model is evaluated on other patients. Despite that, RMS is considered as a useful feature for seizure prediction and it is included in this research project as a feature extracted from the iEEG brain signal.

2.5.7 Entropy

Entropy is a property that indicates the degree of complexity in data. It was initially presented by Shannon in [102] and takes the name of Shannon entropy. Shannon entropy is used to quantify the uncertainty of the EEG data signal to hypothetically discriminate preictal from interictal brain states. In the paper [103] , the spectral entropy was indicated as increased before the occurrence of seizure though in another research paper in [104], the authors indicated that a combination of various entropy measures achieved

an amazing results in terms of accuracy of 90%. In this study, we will extract spectral entropy utilising the total power spectral density from the every channel among 16 channels.

2.5.8 Power Spectral Density

The power spectral density (PSD) indicates the spectral energy distribution that is obtained per unit time, since the over-all energy of such a signal over all time would generally be unbounded. The PSD captures the degree of noise power over different frequencies in Hertz within in a signal. It is among the well known methods applied in seizure prediction to get frequency parameters and can lead to various success depending on the type of PSD utilized[80][65]. The following features can be computed from the PSD ,these are the maximum frequency which indicates the top most peak of the PSD and sum of frequency which indicates the total PSD amplitude in the EEG data signal. These features are important to make separation between seizure and non-seizure phases[99]. In this study, we applied the maximum together with total frequency of the PSD by sampling frequencies up to 80 Hertz for every channel from 16 channels.

2.5.9 Spectral Bands

In this stage, the summed energy is extracted from a set of frequency bands such as 0.1-4,4-6,6-12,12-30 and 30-40 Hz. The approach is similar as the one stated in the section of power spectral density (PSD). These are considered as longer-term spectral features calculated on 10 minute windows. In addition, shorter-term spectral features were computed from various frequency bands of 0-4, 4-8, 8-12, 12-30, 30-70 and 70-180 Hz obtained from 30 second windows of overlapping EEG data (see [3] and the second place method in the 2016 crowd-sourcing reproducible seizure prediction competition [21]). The five frequency bands fro EEG signal is presented in the table 2.4

2.5.10 Summary

Here we have reviewed the field of seizure prediction and closed with a overview of the classifiers and features that have been and can be used for computationally efficient

seizure prediction. In this thesis, we aim to analyse the performance of the most efficient and well known machine learning classifiers for seizure prediction to find the best algorithm for seizure prediction using long term data recordings and applying short-time scale analysis. There after, the goal would be that the best classification algorithm is used to design a transparent seizure prediction model with high predictive power but low computational cost during prediction time via the use of feature selection approaches. Feature selection is used to find the best and relevant features characterising EEG signals to be using for model training and evaluation. We will study different EEG features including time and frequency domain such as statistical moments features computed from iEEG brain signal recorded from 15 epileptic patients in the first-in-man long-term trial dataset known as the NeuroVista dataset. This dataset contains brain EEG recordings that have been obtained from each patient for a period between 6 months to 3 years depending on the patient. In the next chapter, we will discuss our methodology for seizure prediction model development in this thesis.

TABLE 2.4: Five Main Frequency Bands for EEG Signal

Frequency Name	Symbol	Frequency range
Delta	δ	0-4 Hz
Theta	θ	4-8Hz
Alpha	α	8-15 Hz
Beta	β	15-32 Hz
Gamma	γ	32-100 Hz

Chapter 3

Methodology and Design

The proposed method in this thesis consists of the following 2 parts and their associated sub-parts:

Part1: Comparison of 9 classifiers on a large set of seizure prediction features on long-term data using short-time scale analysis.

a. Definition and explanation of 9 supervision based machine learning classifiers and a large set of EEG features that have been mostly used in the past for seizure prediction model development.

b. Definition of the contest data, a tractable dataset used to explore different cross validation approaches.

c. Definition of the full trial dataset, a long-term dataset used for examining the reliability of seizure prediction methods.

d. The outline of the statistical analysis of the individual and group level evaluation of seizure prediction performance.

Part 2: The development of a computationally efficient seizure prediction algorithm

a. The definition of the proposed algorithm

b. The outline of the statistical analysis of the individual and group level evaluation of seizure prediction performance on both contest and full trial dataset.

c. The plan of the evaluation for computational efficiency and performance on the full trial dataset

d. Extending the statistical analysis using clinical relevant evaluation methods.

Our methodology design follows the general structure in machine learning based system development. This consists of data collection, data preprocessing, feature extraction, model training and evaluation. The proposed workflow of the algorithm implementation in this thesis is depicted in the figure 3.1. The main components and activities carried out in each block diagram to achieve our objective for this thesis are discussed in this chapter starting from the features extracted from iEEG data to the model evaluation process. The EEG data is split into non-overlapping 10 minute window segments. In this thesis, the original NeuroVista trial continuous data for each patient were split into three sets referred to as Mode 1, 2 and 3 and containing in chronological order the first 50%, next 25% and last 25% of seizures, respectively. The mode 1 data segments for each patient were divided into training and test sets where the training data segment is 90% while the test set is the last 10% of the full set of data segments for each patient. As such 90% of mode 1 data was used for model training and the remaining 10% of mode 1 data was used for model validation. For each set of training and testing data segments, the features are extracted to create a training and testing set. The model is trained via 10 fold stratified cross validation for model hyper-parameter optimization and the model optimal model threshold is then found. Thereafter, the out-of-sample model evaluation is performed for mode 2 which consists of the data designed for model evaluation after deployment. The same for the mode 3 data. The evaluation in this thesis is based on performance evaluation of AUC and clinical evaluation of model sensitivity and proportion/percent of time in warning (TiW)[77]. In the next sections 3.2, we make description of the human long-term intracranial EEG dataset and the machine learning algorithms that have been developed using the same dataset for state-of-the-art seizure prediction.

3.1 The Definition of Contest Data in Kaggle Competition

The contest data used in seizure prediction competition is a dataset made of the data recording for only three patients. This dataset was chosen based on the performance

where in the initial study in[13] the algorithm have shown poor performance for those patients. The data set was divided into training and testing data segments where the testing or evaluation dataset was made of private and public dataset for evaluation purpose. This dataset is presented in the table 3.2 with the proportion of the preictal and interictal window segments.

3.2 The Definition of NeuroVista Intracranial EEG Data Full Trial Dataset

The iEEG data used is in this project are described in [13]. The data have been recorded continuously from 15 epileptic patients with drug resistant epilepsy using the NeuroVista Seizure Advisory System. The recording took 6 months to 3 years per patient. Here, 16 electrodes (4×4 contact strips) were implanted in every patient, directed to the presumed seizure focus and connected to a telemetry unit embedded in the sub-clavicular area and a rechargeable battery powered the embedded device. Data were sampled at 400 Hz with signed 16-bit resolution and wirelessly transmitted to an external, hand-held personal advisory device. Recorded iEEG from the 16 electrode contacts were referenced to the group average across all electrode channels and continuously saved to removable flash media. The electrode contacts were made of platinum-iridium and embedded in silicon. In the subsection 3.3, we discuss on the proposed algorithm for epileptic seizure prediction.

3.3 The Definition of the Proposed Algorithms for Seizure Prediction

In this thesis, for each patient, the data is divided into 10 minute long window segments and there are 16 channels and 240000 time samples per segment. We perform a two stage evaluation, first on the 2016 Kaggle contest data [21] and then on the full trial data from 15 patients. This enables review of the results both within the context of the previous contest results, as well as results obtained on the much larger dataset. The Kaggle contest data is a small dataset of data from the three patients (out of the 15) that had the worst seizure prediction performance in the original trial [13]. These 3

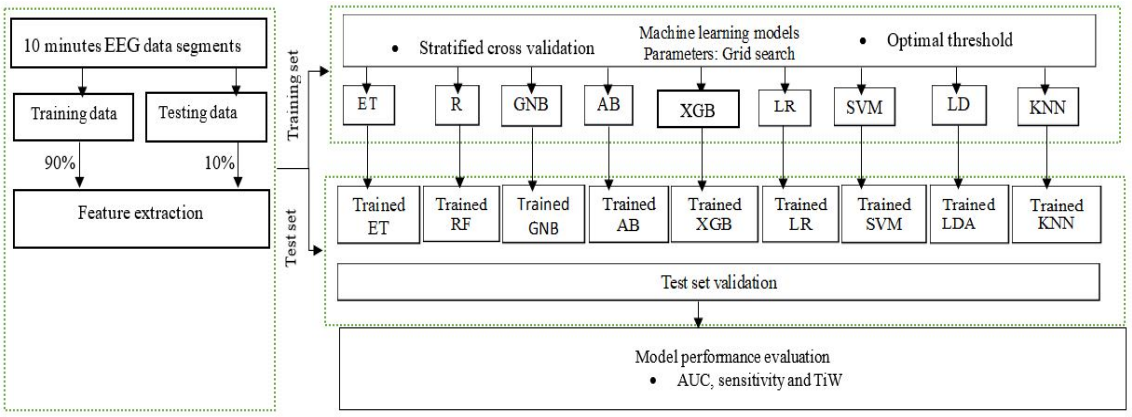


FIGURE 3.1: The flowchart of the new proposed approach for seizure prediction model development. The 10 minutes window segments are divided into train and test segments, there after different models are trained on a training set through stratified cross validation and the trained model is evaluated on test set using the AUC metric. The optimal threshold helps us to evaluate the model for its sensitivity and time in warning. The prediction probability of each segment is computed and if it is greater than the threshold the segment is classified as true prediction (pre-ictal segment) otherwise it is a false prediction (inter-ictal segment)

patients, labelled patient 3, 9 and 10 provided 826, 2058 and 2163 window segments, respectively. The contest data were split into train and test sets as described in [21]. The trial dataset provides between 6 months and 3 years of data per patient for 15 patients (equivalent to the average of 39836 segments per patient). For each patient, the data was split into three roughly equally sized chronologically ordered groups of segments, the first for training and the last two groups for testing. Two test sets were applied to account for possible non-stationarity of the data and class distributions. The patients in this trial dataset are labelled patients 1 to 15, where patients 3, 9 and 10 are the same as in the contest data. For both the contest and trial data, 90% of the training set was used for patient-specific training and the remaining 10% was used for validation. Further details on the patient diagnosis/meta-data are provided in the original papers for the trial [13] and the 2016 Kaggle competition [21]. In this thesis, the most used algorithms for seizure prediction in the past for different datasets are evaluated. The classification algorithms have been detailed in the literature review section2 especially in the subsection of 2.4.1, 2.4.2,, 2.4.4, 2.4.5, 2.4.6, 2.4.8 and 2.4.9.. In the section 3.4, we discuss on the brain characteristic features for seizure prediction system development that have been applied here.

3.4 The Brain Characteristic EEG Features for Seizure Prediction Model Development

The brain is a complex system and composed with various neurons which exchange information in a form of wave signals. EEG signals, which indicate the brain activity are discrete time series and techniques applied in signal processing can be applied to extract relevant information to characterize brain activities. It is hard to know which features characterize brain behaviours or activity during seizure so that these features can be used to train a machine learning model. These EEG features that can be extracted from EEG signals are grouped into two broad categories such as time-domain including statistical moments features which characterize the amplitude distribution in time series x_i and frequency-domain based features. In time domain based features, various statistical features can be extracted from a time series. There are four well known statistical moment features: variance, mean, skewness and kurtosis. There are also Zero-crossing, Hjorth parameters and Root Mean Square[105]. In frequency-domain, EEG time series are transformed into various frequency bands(see table 2.4) using the Fourier transform technique so that the frequencies of the original signal can be observed. Thereafter, different features are computed from each bands and fed into a machine learning algorithms to evaluate the model performance.

In the 2016 seizure prediction kaggle competition[21], various numbers of features have been used to train different learning models and the features in the algorithm in the second place algorithm in the contest seems to be more computationally efficient than others features used in the top 8 algorithms in the competition. The full description of the features in the seizure prediction competition especially those in the second place algorithm have been discussed in literature review of this thesis and are available in [21] as well as in [3]. In the subsection 3.4.1, we talk about the process of EEG features extraction.

3.4.1 Feature Extraction

Feature extraction is a process of computing the possible EEG characteristic features that describe the brain activities. In literature many features have been extracted to be used in seizure prediction model development including those in the time or frequency

domain. In this thesis, the features have been extracted in a 10 minutes window segments and the extracted features are adopted from the second place algorithm in the 2016 kaggle competition for seizure prediction in [21] and are discussed in the Literature review. The subsection 3.4.2 is about the process of the selection of the relevant features for seizure prediction model development.

3.4.2 Feature Selection

In machine learning, not all features are relevant for classification. Some features are redundant and have no influence on the performance of the trained model in the future unseen data. Instead, many irrelevant features increase model computational time, complexity and may be the cause of poor model performance. This is not a good aspect in the development of a seizure prediction system where an efficient, low computational cost and high performance model is important to develop an affordable/implantable device with energy efficiency to monitor brain activity. Therefore, only relevant features are required. In the next section, we discuss feature selection approaches and make a comparison table of each feature selection method.

3.4.2.1 Correlation Based Feature Selection (CFS)

Feature selection is a technique of selecting relevant features from a set of features for classification to achieve a high performing and interpretable model. There are three popular types of feature selection such as wrapper, filter and embedded method. The wrapper method itself applies three search approaches which are exhaustive, heuristic and random search methods [106]. The filter method is based on the relationship between features where difference evaluation criteria are used. This includes the measure of distance metrics to analyse the inter-class distance, such as the correlation coefficient to analyse the inter-dependency among features. The mutual information which works like correlation based feature selection but basically quantifies the information obtained from one feature via the other features. A consistency metric is normally computed as the chi-square statistic between features in a data set and the target class. The main group of type of feature selection methods are described in [107]. Table 3.1 shows the role and use of these methods. In this thesis we adopt the features in[3] where these features have been identified using the same technique of correlation feature selection

(CFS) techniques with backward elimination [CFS, 110] to select the relevant features for classification. It is simple, not based on a specific machine learning algorithm and less prone to overfitting. For example, given two random variables X and Y, the correlation is computed using the Pearson correlation coefficient. See equation 3.1 where left is the population formula and right is the sample formula.

$$\rho_{xy} = \frac{\text{cov}(X,Y)}{\sigma_x \sigma_y} \quad \text{or} \quad r_{xy} = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2 (y_i - \overline{y})^2}} \tag{3.1}$$

Suppose we have a set of N features extracted from 16 channels of EEG signal data. The variable x_i is the i^{th} feature where i takes values from 1,2,...,n and y_j is a target class. Let $r(x_i, x_k)$ be the correlation coefficient between the i^{th} and k^{th} feature and $r(x_i, y_j)$ be the correlation coefficient between the i^{th} feature and j^{th} target class. The relevant features are selected in such a way that $r(x_i, x_k)$ is less than a fixed threshold r_{t1} and $r(x_i, y_j)$ is greater than a threshold r_{t2}.

TABLE 3.1: Summary of type of feature selection.

Filter method	**Wrapper methods**	**Embedded methods**
It is a generic set of methods which do not incorporate as specific machine learning algorithm	It works for a specific machine learning method to find the optimal features to fit on a given data set and uses a grid search method studying the performance of a model on a possible set of features based on an evaluation performance criteria.	Embeds features during model building process. Feature selection is done by observing each iteration of a model training phase. It uses a grid search approach to test all plausible combinations of features for a given set of performance metrics.
It is faster compared to wrapper method in terms of time complexity	It is not recommended for data sets with high numbers of features (High computation time)	In terms of computation time. It is higher than the filter and less than that of the wrapper method
It is less prone to over fitting	It has high chance of over fitting due to its property of training a model with various combination of features	It is used mostly in reduction of over fitting by penalizing the coefficients of a model being too large.
Example: Correlation, Chi-Square test,Information gain, ANOVA etc	Examples are: Forward selection, backward elimination, stepwise or Bi-directional elimination selection etc	Examples: Lasso, Elastic Net, Ridge regression[108],[109] etc

From a machine learning perspective, the performance of an algorithm is influenced by many factors including the relevant features as well as the the proportion of class labels. Therefore, imbalance in a data set during model training is a problem. One of the approaches to alleviate the issue is the creation of synthetic observations of the minority class known as synthetic minority oversampling or different types of cross validation to avoid the model overfitting issue. Feature selection for seizure prediction model development has previously been applied on the features in the second place algorithm and only 208 features have been selected from 2022 features in the second place algorithm [3] [21]. The results have shown improvement for both in terms of AUC and computation complexity in memory usage, training and prediction time.

In this thesis, we will first use all 208 features to analyse the algorithm performance for seizure prediction to find which algorithm performs better than others across all patients. Second, we will apply SHapley Additive exPlanation (SHAP) to find the best features based on features importance to achieve simpler and more interpretable models.

3.4.2.2 SHapley Additive exPlanation (SHAP)

SHAP is a feature selection technique that is based on the gradual decline of model performance results when a specific feature is used. It is a replacement of permutation feature importance [111]. In SHAP feature selection techniques, each feature is given a SHAP value which describes its ranking in influencing the model performance. It is useful for ranking based algorithms and works better for tree and boosting based classifiers. The SHAP algorithm gives the idea that is related to the input feature contributions for a particular prediction. For simplicity, the SHAP algorithm is expressed as a linear regression function where x_i is the i_{th} feature, w_i refers to the weight associated with the i_{th} feature and $\hat{y}$ is the prediction made by the model. $\hat{y} = w_0x_0 + w_1x_1 + w_2x_2 + \cdots + w_nx_n$. In machine learning model development, the algorithm or a learning model can be biased due to the class imbalance in the dataset. Hence, it is good idea to avoid class imbalance during model training.

3.5 Class Imbalance

In machine learning, class imbalance is a common issue. This happens when one class is highly over-represented than the others and then a trained classifer becomes biased towards the over-represented class. In epileptic seizure prediction, within a dataset preictal or interictal classes may have different data representation where we may have more interictal or preictal segments. In the NeuroVista dataset, in general there are more interictal than preictal window segments. For example in the 2016 contest data. See table 3.2, more than 50% of the data are interictal window segments while less than 30% are preictal with patient 1 which has many preictal segments. The table 3.2 shows the proportion of preictal and inteictal segments in the contest data where it is clear that there is imbalance in the class labels(high number of interictals that preictals segments)

3.5.1 Cross Validation to Avoid Class Imbalance

In this thesis, we applied three different types of cross validation such as k-fold, stratified and group k-fold cross validation to overcome the issue of class imbalance. Group k-fold cross validation takes care of the successive of 10 minutes segments to make one hour during cross validation at each fold while stratified cross validation works in a way that at each fold, the proportion of class labels is represented and the general model performance is obtained as an average of all performance at each fold. The average results demonstrate how the model is expected to perform on new unseen data. In this thesis, the cross validation is performed on 90% of the full dataset for each patient to avoid overfiting as well as optimizing the model hyper-parameters of each of the analysed classification algorithms while the last 10% of the data is used for model validation.

TABLE 3.2: The proportion of preictal and interictal data segments for the three patients from the NeuroVista human iEEG dataset used in the 2016 kaggle contest.

Patients	**Preictal**	**Interictal**	**Total**	**Preictal %**	**Interictal %**
P1	256	570	826	30	70
P2	222	1836	2058	11	89
P3	256	1908	2164	12	88

3.6 Method Implementation

The implementation of our proposed methodology involves the analysis of the performance of 9 classification algorithms. We used the NeuroVista intracranial electroencephalogram data recording from 15 patients. The algorithm is trained with 90 percent of data recording for each patient from mode 1 and the evaluation is done on last 10 percent of the same data recording. In this thesis, as the part 1 task defined at the start of the methods, we have evaluated the capability of each of the nine most applied machine learning algorithms to distinguish preictal from interictal window segments for epileptic seizure prediction. The evaluation focused on the analysing of the performance of 9 learning algorithms first using the 2016 Kaggle contest dataset with 3 patients as a small dataset. Following this, the evaluation then focused on the full long-term continuous EEG recording of 15 patients from the NeuroVista Trial. For the contest data, we have considered the evaluation of the learning algorithms for different types of cross validation with grid search for hyperparameter tuning and to avoid model over-fitting. This is done, to find the best type of cross validation to apply during the evaluation on the full continuous trial dataset and the search for the best classification algorithm.

To access the performance of each algorithm on unseen data, we have used three test sets noted as test 1, test 2 and test 3 which are the last 10% of mode 1, 2 and 3, respectively. The test 1 is for the model evaluation on the model training set while test 2 and 3 are used to simulate the model evaluation after the model is deployed.

After computation of the outputs of each classifier for each patient, the cross patient performance is accessed based on average performance and test statistics are calculated to assess for statistical significant differences between various classifiers' performance at the group level. To do this, each classifier's performance is compared to the extra tree classifier as a our benchmark in this study. As noted above, this method was chosen as the benchmark for its computational efficiency and good performance as proven in the 2016 Kaggle competition using the long-term human contest dataset.

As the part 2 task defined at the start of the methods section, we looked at taking the second place algorithm from the 2016 Kaggle contest and developing a more computationally efficient algorithm by using SHAP feature selection to further reduce the

number of features within a patient-specific algorithm whilst maintaining prediction performance. This involved analysis on the contest and long-term trial datasets in a similar fashion to the part 1 task. Note SHAP was not applied in the part 1 task.

In the next section 3.7, we describe the test statistics used in this thesis for algorithm comparison while in 3.8, we talk about algorithm performance evaluation process and discussion on each algorithm performance.

3.7 Statistical Analysis Test

In this thesis, we perform two sample t-tests to compare the AUC of classifiers and find if there is a significant difference between the performance among classification algorithms studied in this study. This includes using the extra tree classifier as a reference as it is the chosen benchmark in state of the art seizure prediction model development because it represents the second place algorithm of the 2016 Kaggle contest. One sample t tests were also used to compare the individual classifier against chance prediction. For both the one and two sample t-tests a single data point is the AUC for a single subject for a specific algorithm. As such the t-tests can be viewed as group level tests. For both tests correction for multiple comparisons was performed using adjusted significance levels based on Bonferroni correction. Statistical comparison of the two AUC values of classifier pairs obtained for the same individual was performed using the Hanley-McNeil test. As such this represents an individual level test. The method works by taking into account the correlation between the areas that is induced by the paired nature of the data. In the subsection 3.8, we discuss the evaluation approach in this study in more detail.

3.8 Windowed Segment Classification Evaluation'

After training the model, we evaluate its capability to classify preictal from interictal segments[20]. This is achieved with the help of AUC which gives the information on the capability of a given model classifier to separate interictal from preictal EEG data segments. The evaluation also considers the model sensitivity and specificity. The goal in machine learning is to have a trained model with high true positive rate for low false positive rate. The better performance is achieved when the plot is near the point

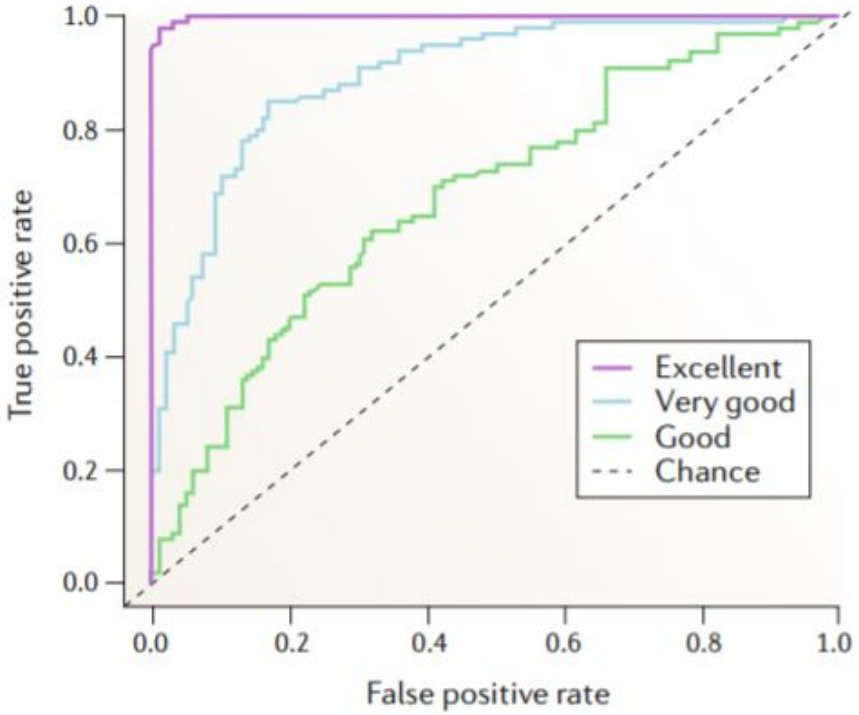


FIGURE 3.2: Performance evaluation for seizure prediction system. The higher AUC, the better model we get. The metric is taken from "Seizure prediction – ready for a new era" by Kuhlmann et al. [1]

($Truepositiverate = 1, Falsepositiverate = 0$) in the left side while when the AUC value is 0.5 this indicates no proper decision can be made. The dashed line is symmetric and is known as the baseline or chance prediction line. Again, model computational efficiency is measured by means of the total memory usage by the algorithm while being executed as well as run time. This can be monitored by using Performance monitor in Windows operating systems while the time taken by the model for classification can be recorded by using time package in SciKit Learn from the Python environment.

As noted in literature review, the ROC curve presents the performance of a designed model to make the real prediction for each possible threshold. Therefore, by considering the space under the ROC curve, we are able to get the AUC values which is normally between 0 and 1. The best model achieves the AUC of 1 while the poorest model achieves AUC of 0.

The ROC curve involves plotting sensitivity (true positive rate (preictal segment occurs and is predicted)) against false positive rate (1-specificity) (predicted a preictal segment that has not occurred) see figure 3.2.

AUC is mostly used in assessing the performance of seizure prediction classifier models. However, it is not the only metric that can be used. Clinically relevant metrics can also be used.

3.9 Clinically Relevant Performance Evaluation

In addition to the above 3.8 performance measures, clinically relevant pseudo-prospective evaluation of seizure prediction models is highly important. This involves the measures of model sensitivity and the percent of time in high risk or time in warning. The evaluation method has been initially published in [77] and used in [21]. In this thesis, only for the part 2 task, the same evaluation has been performed for evaluation of the clinically relevant performance of each classification algorithm on the full trial data set for better comparison with the benchmark algorithm for the same dataset.

In the NeuroVista seizure advisory system, the device utilised a three colour light based warning system involving the colours of red, white and blue to indicate the proportion of seizure risk occurrence. Here, red light, white and blue light show high, moderate and low seizure risk, respectively. A machine learning model generates the probability of seizure associated with each preictal window to be predicted and two thresholds are applied, one for upper threshold and another for lower threshold. In the case when the predicted probability is above the upper threshold, a warning is issued to indicate the high risk of seizure occurrence (red colour light) while when it is below the lower threshold, a low risk warning is issued (in this case a blue light colour is used). The warning indicator is issued for a specific time duration[77].

Clinically, the seizure prediction performance measures consider the model sensitivity which is the proportion of seizures that are correctly classified, which means the ratio of the count of seizures happening during high-seizure-risk and the total count of seizures. (This is in contrast to how sensitivty is typically defined in machine learning as the percent of the positive class/preictal segments correctly classified, as was defined above.) Moreover, the percentage of time in warning is defined as the ratio between the amount of time spent in warning supposing a fixed warning time duration generated after each prediction and the total recording time expressed as a percentage.

In this thesis, the performance evaluation considers the evaluation using three different datasets where the first dataset is used to find the best classification algorithm. The second test set is used for validation purpose to while the third test set is used to analyse the performance variability between the performance of each classifier from patient to patient.

3.10 Optimal Threshold for Clinical Evaluation

For the clinical evaluation for model sensitivity and time in warning, we have computed the optimal threshold based on the training data and use this threshold to find the model sensitivity and time in warning for a validation set, See figure 3.4. The optimal threshold is computed based on the Youden index where the optimal threshold corresponds to the point where the true positive rate is highest while the false positive rate is lowest. Such that the green vertical line, depicted in figure 3.3 and representing the magnitude of the difference in true and false positive rates, is longest.

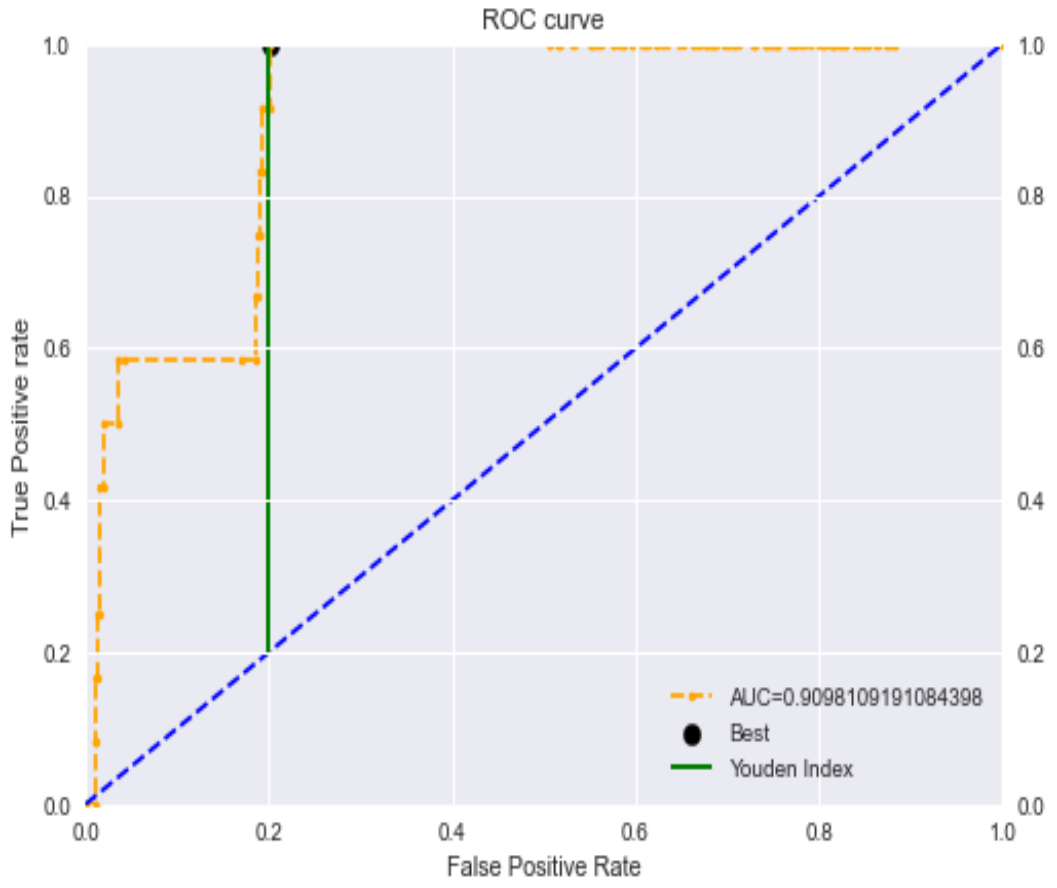


FIGURE 3.3: The Youden index plot. The point close to one on the upper right corner in this figure indicates the optimal threshold value

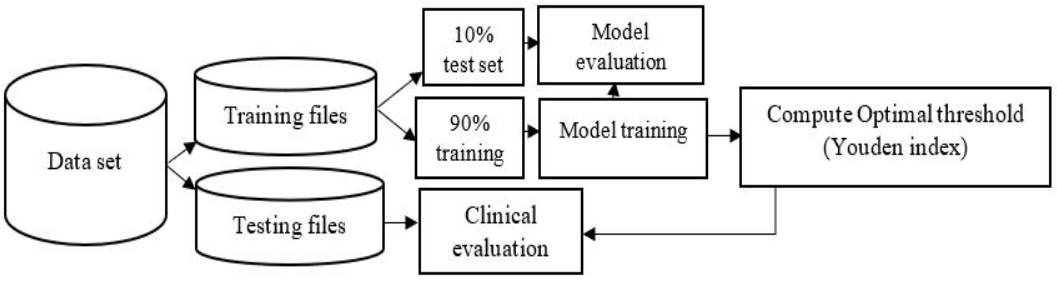


FIGURE 3.4: The process of optimal threshold computation for clinical evaluation. This threshold is computed based on the Youden index where the optimal threshold corresponds to the maximum values of Youden index

Chapter 4

Results and Discussions

4.1 Part 1 task: Comparison of the 9 Classification Algorithms

In this section, we first evaluate and discuss on the performance of these classifiers on contest data for 3 patients and then evaluate the performance on the full trial dataset of 15 patient. This will be followed by the statistical analysis.

4.1.1 The Analysis of Seizure Prediction Algorithm Performance using Short-time Scale Feature Analysis on Contest Data

The results for the contest data are tabulated in the table 4.1 for public test set and in the table 4.2 for private test set.

Clf	**K Fold**				**GroupKFold**				**Stratified Kfold**			
	P1	P2	P3	AV	P1	P2	P3	AV	P1	P2	P3	AV
ET	0.714	**0.757**	0.648	0.706	**0.710**	0.734	**0.804**	**0.749**	0.745	**0.761**	0.639	0.715
RF	0.565	0.639	**0.789**	0.664	0.693	0.625	0.758	**0.692**	0.588	0.641	**0.786**	0.671
GNB	0.580	0.594	0.407	0.527	0.580	0.586	0.408	**0.524**	0.580	0.594	0.407	0.527
AB	**0.750**	0.656	0.348	0.527	0.677	0.638	0.631	**0.648**	**0.750**	0.656	0.348	0.584
XGB	0.675	0.622	0.598	0.631	0.675	0.625	0.620	**0.640**	0.486	0.518	0.486	0.496
LR	0.471	0.599	0.640	0.570	0.471	**0.748**	0.667	**0.628**	0.471	0.599	0.640	0.570
SVM	0.621	0.467	0.527	0.533	0.621	0.568	0.524	**0.621**	0.621	0.568	0.524	0.571
LDA	0.321	0.708	0.406	0.478	0.603	0.740	0.420	**0.587**	0.321	0.740	0.402	0.584
KNN	0.468	0.543	0.740	0.583	0.460	0.550	0.779	**0.596**	0.451	0.546	0.741	0.579

TABLE 4.1: The performance of classifiers on the public test set in terms of AUC for each patient with respect to the different type of cross validation method. Clf stands for classifiers,ET(extra tree),RF(Random Forest),GNB(Gaussian Naive Bayes),AB(Adaptive Boosting),XGB(Extreme Gradient Boosting),LR(Linear Regression),SVM(Support Vector Machine),LDA(Linear Discriminant Analysis),KNN(K-Nearest Neighbor). The Bold values indicates the best classifier AUC Values for a specific patient and the highest average performance values across all three patients in the contest data

Clf	**K Fold**				**GroupKFold**				**Stratified Kfold**			
	P1	P2	P3	AV	P1	P2	P3	AV	P1	P2	P3	AV
ET	0.714	0.757	0.648	0.706	0.710	0.734	**0.804**	**0.749**	**0.745**	0.761	0.639	0.715
RF	0.631	0.745	**0.758**	0.711	0.719	0.746	0.732	**0.732**	0.654	0.751	**0.735**	0.713
GNB	0.579	0.569	0.586	0.578	0.579	0.700	0.586	**0.621**	0.579	0.569	0.586	0.578
AB	0.726	**0.809**	0.488	0.578	0.662	0.711	0.637	**0.670**	0.726	0.685	0.488	0.633
XGB	**0.729**	0.729	0.643	0.700	**0.729**	0.729	0.664	**0.707**	0.729	0.718	0.532	0.659
LR	0.471	0.599	0.640	0.570	0.471	0.748	0.667	**0.628**	0.471	0.599	0.640	0.570
SVM	0.621	0.467	0.527	0.533	0.621	0.568	0.524	**0.571**	0.621	0.568	0.524	0.571
LDA	0.387	**0.809**	0.562	0.586	0.642	**0.826**	0.579	**0.682**	0.387	**0.827**	0.541	0.585
KNN	0.568	0.578	0.667	0.604	0.573	0.589	0.672	**0.611**	0.545	0.595	0.667	0.602

TABLE 4.2: *AUC* The performance of classifiers on the private test set in terms of AUC for each patient with respect to the different type of cross validation method. Clf stands for classifiers,ET(extra tree),RF(Random Forest),GNB(Gaussian Naive Bayes),AB(Adaptive Boosting),XGB(Extreme Gradient Boosting),LR(Linear Regression),SVM(Support Vector Machine),LDA(Linear Discriminant Analysis),KNN(K-Nearest Neighbor). The Bold values indicates the best classifier AUC Values for a specific patient and the highest average performance values across all three patients in the contest data

On average, LDA performs better than other classifiers on test set 2 followed by the extra tree classifier. The average comparison of different classifier is shown in figure 4.2

In general, most of the classification algorithms evaluated in this thesis have achieved above chance prediction performance for most of the subjects with the exception of Patients 5, 7 and 12 where the performance was lower than chance prediction AUC of 0.5 for all classifiers with the exception of LDA for Patients 5 and 7 as well as GNB for Patients 7 and 12. The difference in AUC performance values is the explanation that each patient has its own characteristic and the seizure prediction model is patient specific and the brain behaviours for one patient may change from time to time and one classifier may give different performance for the same patient.

The results demonstrated that in terms of the type of cross validation, group k-fold has on average achieved good performance compared to other type of cross validation while extra tree classifier and random forest have good performance than other classifiers evaluated in this thesis. However, some of the classifiers like boosting based algorithms have achieved better performance than chance prediction.

Due to class imbalance in our dataset, the Group K fold cross validation was not implemented on full continuous dataset since it does not consider the class representation in each fold as stratified cross validation does. Hence, there might be overfitting on the algorithm outputs. Therefore, we have implemented the algorithm evaluation using the stratified cross validation since on the contest data it has shown on average to give the closest performance to that of group k-fold cross validation and the performance is better than chance prediction. Stratified cross validation takes care of the problem of over-fitting in supervised learning approaches for binary classification. In the subsection 4.1.2, we extend the analysis of this study on the full trial for 15 patients.

4.1.2 The Analysis of Seizure Prediction Algorithm Performance Using Short-time Scale Feature Analysis on Full Trial Dataset

The results that compare the average performance of different classifiers for each patient across all 15 patients in the dataset is presented in the figure 4.1 for test 1 as a test set and 4.2 for the test set 2 as a validation test set.

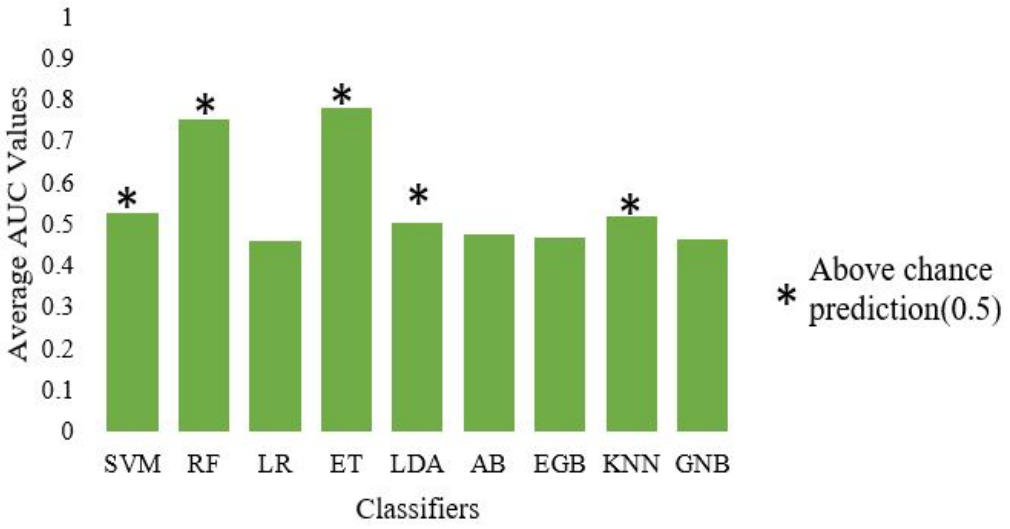


FIGURE 4.1: The AUC average comparison of the classification algorithms against chance prediction on long-term continuous intracranial EEG for test set 1. The star symbol " $*$ " indicates that the classifier with AUC performance is above chance prediction.

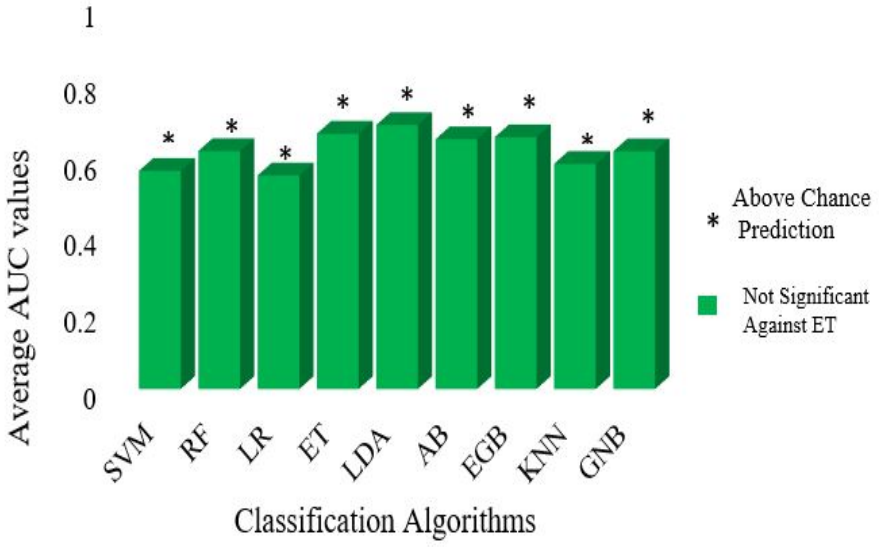


FIGURE 4.2: The AUC average comparison of the classification algorithms on long term continuous intracranial EEG against ET as the bench mark in this study. The " $*$ " show that the average performance from all classifiers across all 15 patients is above chance prediction at alpha =0.05 while the dark green color shows that there is no significant difference when the performance of each classifier is compared to the performance of extra tree classifier after two-sided two-sample t-tests corrected for multiple comparisons.

The results on the training set, see figure 4.1 show that on average the extra tree classifier is the best classification algorithm among others with AUC of 78%. This followed by

Random Forest, Support Vector Machine, Linear Discriminant Analysis and Near rest neighbor and the performance is above chance prediction while the poor performance is for Gaussian Naive Bayes, Logistic Regression, adaptive boosting and extreme gradient boosting with almost the the same AUC values. In the next section, we perform statistical analysis to find if there is the significant difference between the performance of these classifier against chance prediction and the best classifier in the test 2 which is also the second place algorithm in the Kaggle competition for seizure prediction.

4.1.3 Statistical Analysis of the Algorithms Performance at the Group Level(Across all 15 Patients)

The performance of each classifier for test set 2 for a given patient is presented in the table 4.3. The results show that a given classifier presents different performance for different patients.

In our results for test set 2, it is clear that Linear Discriminant Analysis, Extra Tree classifier, Adaptive Boosting, Extreme Gradient Boosting and SVM have beaten other classifiers. In this thesis, LDA and AB were the best classifier each for 5 times out of 15 possible options for different patients, ET and EGB were the best 3 times each out of 15 possible options for different patients, while SVM and RF each become the best only once in 15 possible options as shown on figure 4.2. In general, the LDA has achieved

TABLE 4.3: The Algorithm Performance for Each Patient on Test 2

Patients	SVM	RF	LR	ET	LDA	AB	EGB	KNN	GNB
P1	0.559	0.678	0.559	0.69	0.734	0.671	0.678	0.56	0.594
P2	0.611	0.678	0.432	0.715	0.658	0.598	0.61	0.539	0.588
P3	0.608	0.617	0.608	0.684	0.898	0.724	0.859	0.844	0.89
P4	0.559	0.656	0.559	0.714	0.583	0.805	0.5	0.605	0.658
P5	0.444	0.258	0.444	0.244	0.77	0.486	0.5	0.247	0.135
P6	0.367	0.865	0.367	0.854	0.71	0.543	0.727	0.507	0.616
P7	0.432	0.479	0.432	0.482	0.576	0.561	0.497	0.508	0.517
P8	0.57	0.585	0.57	0.706	0.567	0.824	0.705	0.671	0.733
P9	0.587	0.696	0.587	0.807	0.808	0.814	0.806	0.708	0.686
P10	0.471	0.613	0.471	0.659	0.551	0.72	0.742	0.62	0.625
P11	0.639	0.682	0.639	0.837	0.846	0.871	0.879	0.774	0.783
P12	0.758	0.356	0.758	0.436	0.465	0.177	0.5	0.485	0.731
P13	0.479	0.624	0.479	0.569	0.537	0.676	0.653	0.562	0.486
P14	0.803	0.878	0.803	0.942	0.961	0.857	0.5	0.64	0.572
P15	0.658	0.656	0.658	0.659	0.677	0.454	0.708	0.549	0.702

better performance than chance prediction and compared to the average performance of other classifiers studied in this thesis. This average results comparison is presented on the figure 4.2.

Statistically, all classifiers achieved better results than chance prediction as indicated by a statistical analysis based on one sample t-test. This analysis is presented in the table 4.4 for significant level of alpha=0.05. However, for deeper analysis with Bonferroni correction alpha/9, for 9 comparison, only the performance of Random Forest, Extra Tree, Linear Discriminant Analysis, Adaptive Boosting and Extreme Gradient Boosting classifiers are statistically significant compared to chance prediction (0.5).

Next we tested to see if there are any statistically significant differences in AUC performance if we compare each of the evaluated algorithms with the benchmark, which is the algorithm in the second place in the Kaggle competition. Here the second place algorithm in the Kaggle contest for epileptic seizure prediction is the Extra Tree classifier. To access this, we use a two sample t-test statistical analysis with Bonferroni correction since we are performing 8 comparisons. The results in table 4.5 shows that there is no significant difference between the average performance of the second place algorithm with other classifiers studied in this thesis.

However, the individual analysis in table 4.2 shows that for some classifiers there are statistically significant differences between the classifiers' performance. This is the case for SVM versus LDA and LR versus AB for a significance level of alpha=0.05.

TABLE 4.4: The statistical analysis comparison of algorithm performance against chance prediction for test set 2

Classifiers	Statistical test	T Values	P values	Significant?	Alpha/9
SVM	One sample t-test	2.2574	0.0202	NO	0.005
RF	One sample t-test	2.9023	0.0057	YES	0.005
LR	One sample t-test	1.8038	0.0464	NO	0.005
ET	One sample t-test	3.6291	0.0013	YES	0.005
LDA	One sample t-test	5.0293	0.0000	YES	0.005
AB	One sample t-test	3.1307	0.0036	YES	0.005
EGB	One sample t-test	4.5122	0.0002	YES	0.005
KNN	One sample t-test	2.4493	0.0140	NO	0.005
GNB	One sample t-test	2.7571	0.0077	NO	0.005

4.1.4 The Analysis of Various Classification Algorithms Performance for the Same Patient for Different Test Sets

The results in the figure 4.2 show that for two different test sets, on the same patients, one may performs better for one test set with poor performance on the other test set.

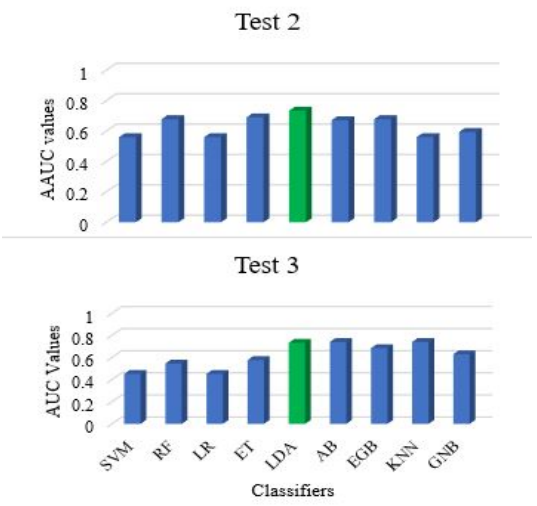
The variability in performance can be seen in figure 4.2 for Patients 2, 5, 12 and 15. This makes the seizure prediction model development a complex problem in real time application. The cross subject analysis using statistical analysis shows that there is no significant difference between the performance of Extra Tree classifiers versus Random Forest, Linear Discriminant Analysis, Adaptive Boosting, Extreme Gradient Boosting and Gaussian Naive Bayes where the P-values were greater than the significance level of 0.05. However, there is a significant difference in performance for Extra Tree classifier versus SVM, Logistic Regression and K Nearest Neighbor with p-values less than significant level of 0.05. See table A.1 in the Appendix.

4.2 Statistical Analysis of Algorithm Performance on Individual Basis.

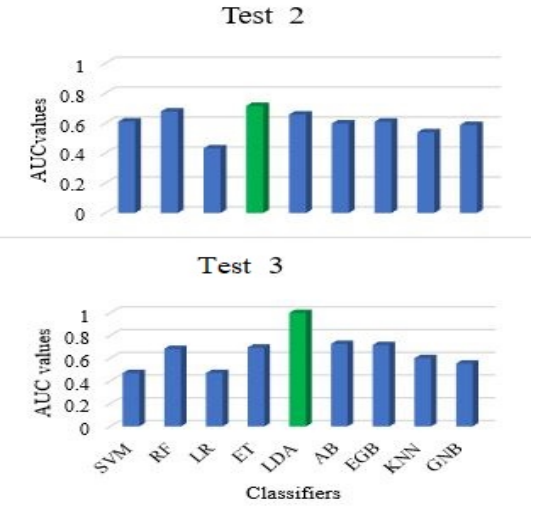
We have seen that different algorithms achieve different performance for different patients. In this section, we analyse whether there is a significant difference in performance for different classifier pairs at the individual level. In particular, our comparison refers to the performance of the second place algorithm from the 2016 kaggle contest with any one of the other 8 classification algorithms evaluated in this thesis. The results

TABLE 4.5: The statistical analysis comparison of algorithm performance against The second place algorithm in the 2016 Kaggle contest for test set 2

Classifiers	Statistical test	T Values	P values	Significant?	Alpha/8
ET vs SVM	Two Sample T-test	1.7516	0.0920	NO	0.00625
ET vs RF	Two Sample T-test	0.72689	0.4733	NO	0.00625
ET vs LR	Two Sample T-test	1.9447	0.0631	NO	0.00625
ET vs LDA	Two Sample T-test	-0.3852	0.7031	NO	0.00625
ET vs AB	Two Sample T-test	0.2165	0.8301	NO	0.00625
ET vs EGB	Two Sample T-test	0.1549	0.8780	NO	0.00625
ET vs KNN	Two Sample T-test	1.3490	0.1889	NO	0.00625
ET vs GNB	Two Sample T-test	0.7158	0.4799	NO	0.00625

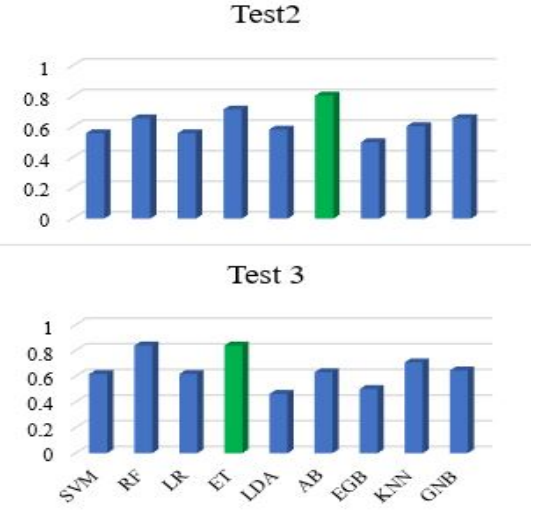


(A) P1

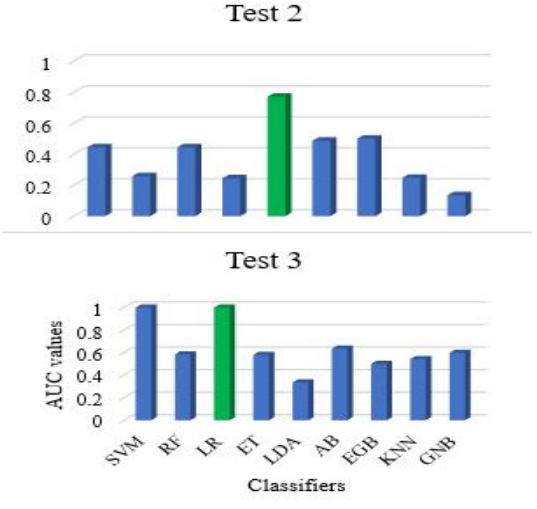


(B) P2

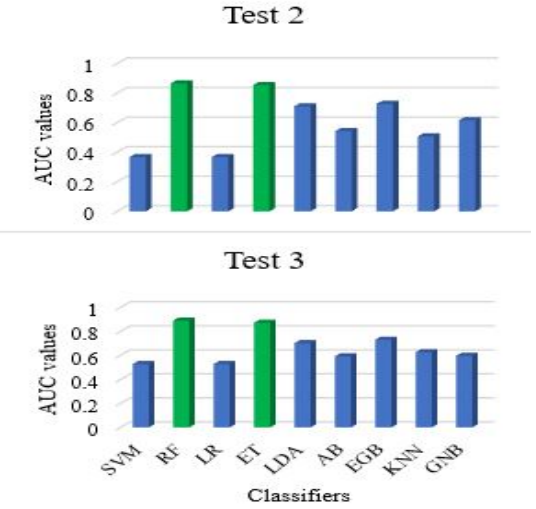
(C) P3



(D) P4

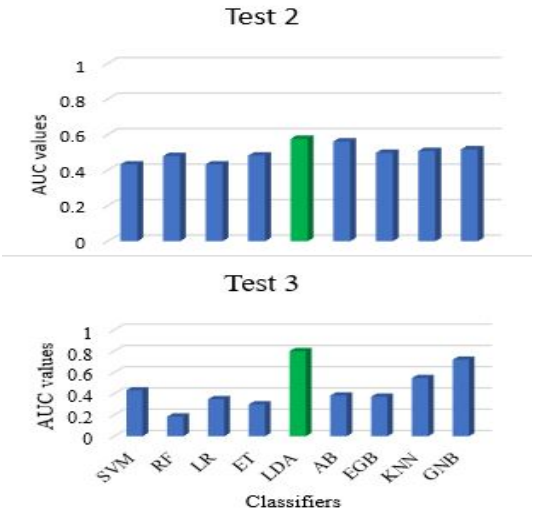


(E) P5

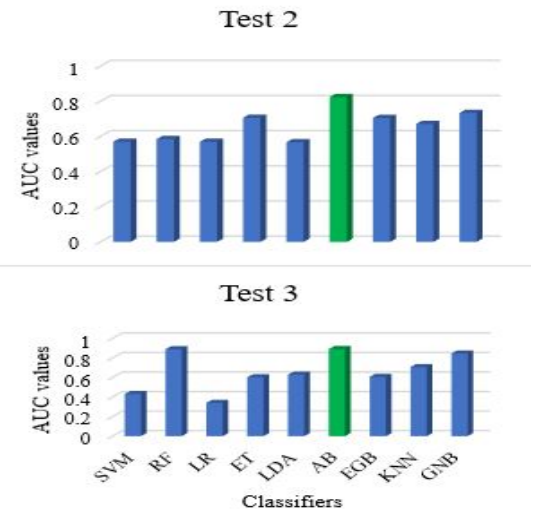


(F) P6

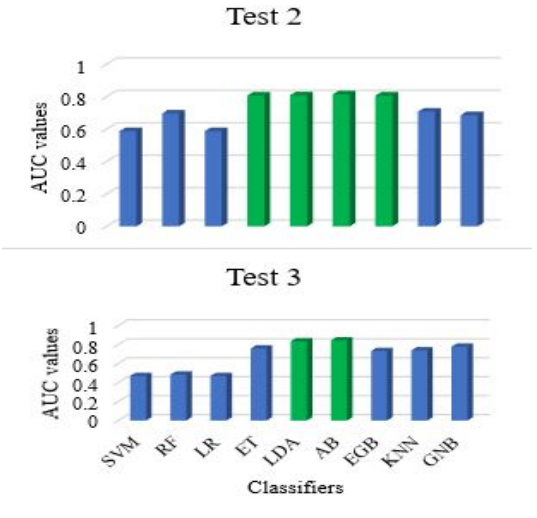
FIGURE 4.2: The performance of different classifiers for each patient in the long term continuous intracranial EEG with different test set: Patients 1 to 6. The green bar indicates the bets classification algorithm for a given test set on the same patient.



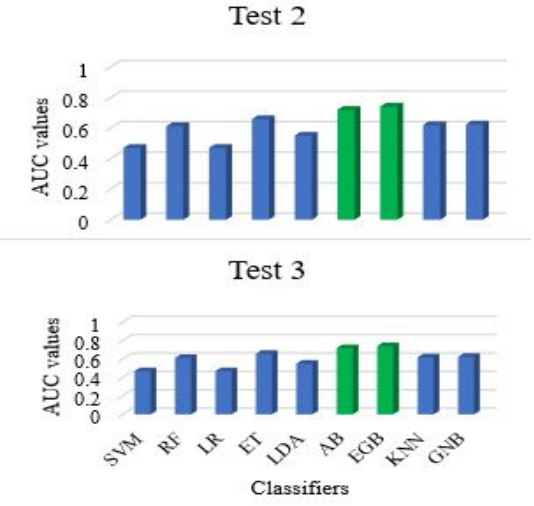
(G) P7



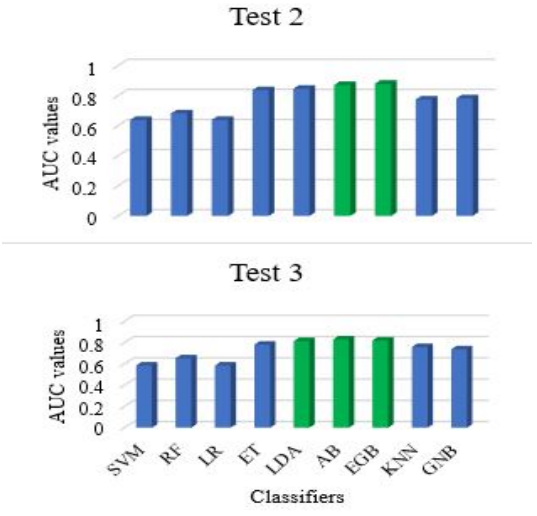
(H) P8



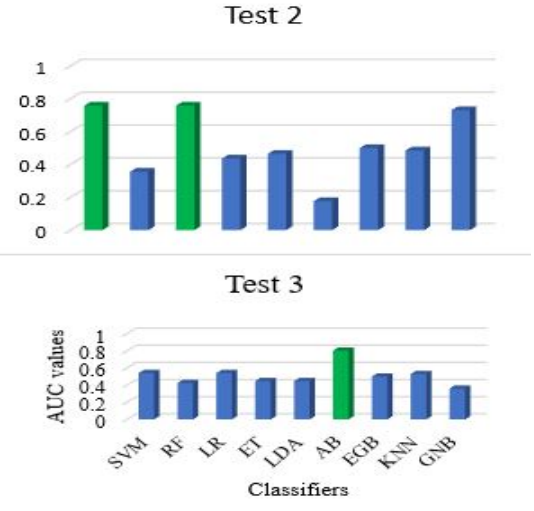
(I) P9



(J) P10

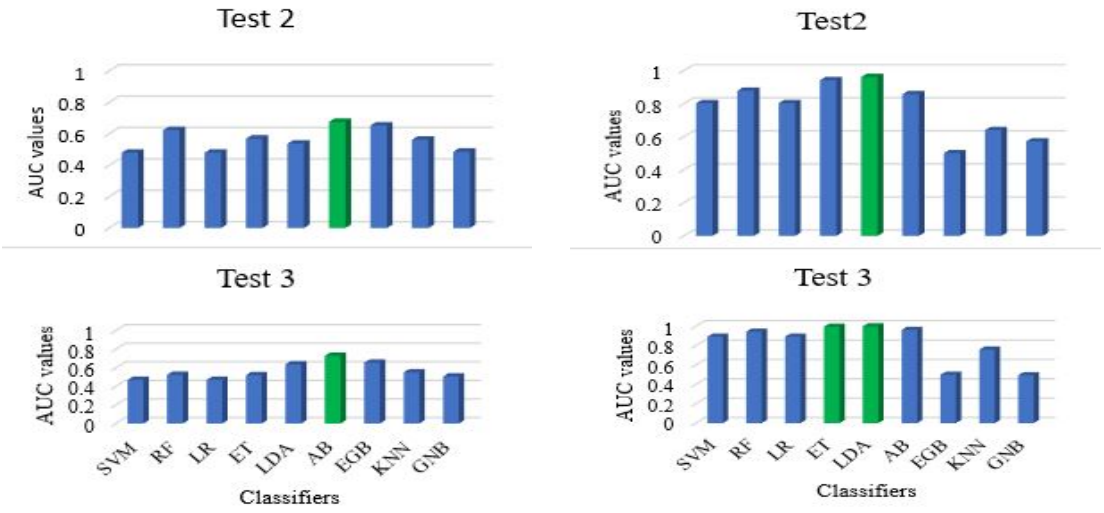


(K) P11



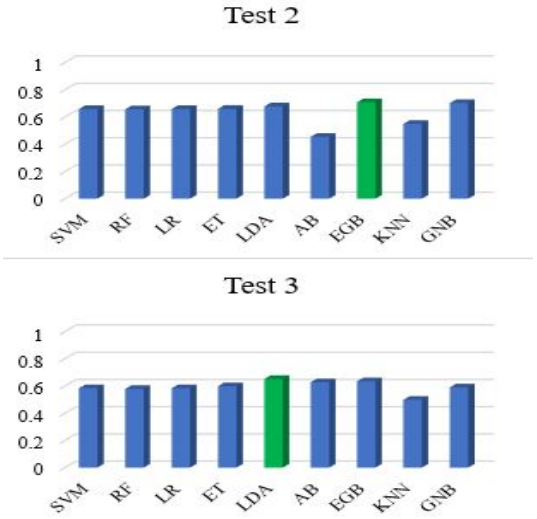
(L) P12

FIGURE 4.2: Cont. The performance of different classifiers for each patient in the long term continuous intracranial EEG with different test set: Patients 7 to 12. The green bar indicates the bets classification algorithm for a given test set on the same patient.



(M) P13

(N) P14



(O) P15

FIGURE 4.2: Cont. The performance of different classifiers for each patient in the long term continuous intracranial EEG with different test set: Patients 13 to 15. The green bar indicates the bets classification algorithm for a given test set on the same patient.

are presented in the figure 4.3. The presented results show that mostly, there is not a significant difference between the performance of the extra tree classifier and the other classification algorithms and the difference can be significant for one patient and not significant when the same classifier is studied for another patient. This can be observed for example when considering the performance of Random Forest (RF) for patients 1 and 13. The performance of all classifiers for patient 12 shows no significant differences. In general, the performance of each classifier changes from patient to patient which confirms the need for patient-specific seizure prediction systems.

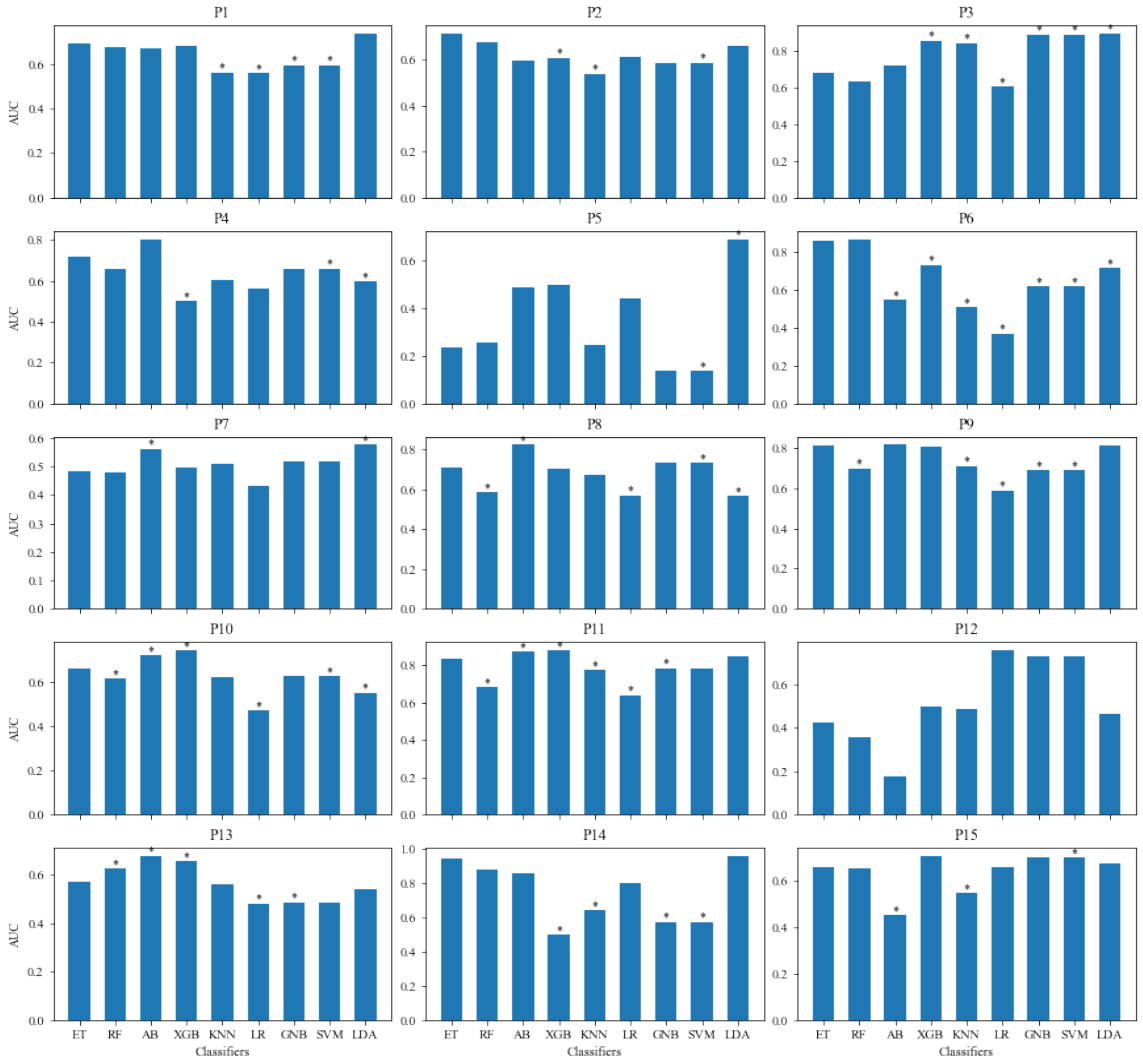


FIGURE 4.3: The comparison of the algorithm performance for each patient against the second place algorithm and statistical analysis of the performance of the algorithms for each patient. The * shows that there is a significant difference between the performance of extra tree classifiers and the specific classifier for the same patient based on the Hanley McNeil test with Bonferroni correction of alpha=0.05/8

In the next section 4.3, we will focus on the improvement of the model performance based on the feature selection.

4.3 Part 2 task: Making More Efficiency Algorithm with Feature Selection

As we have mentioned in the section 3.4.2 of the literature review, the relevant features can improve the performance of a given classifier and produce an efficient algorithm in terms of having low computational complexity. In this section 4.3 we apply feature

selection to produce an efficient seizure prediction algorithm to refine the benchmark method.

4.3.1 SHAP Feature Selection Based on Extra Tree Classifiers for Epileptic Seizure Prediction

We propose SHAP (SHapley Additive exPlanation) algorithm. This is a feature selection technique that selects features based on the highest Shapley values instead of the feature importance generated by a classifier. This algorithm has been applied given that it is designed for tree based classifiers, like extra trees [112]. The features are extracted from each 10 minute window segment and fed to the extra tree classifier. The SHAP values for each feature are ranked in descending order where only the first 20 features are used for model training and evaluation. The figure 4.7 illustrates the example of the features selected by the SHAP algorithm where the colour represents the values of each feature from high to low and each point is the SHAP value for a feature. The overlap of the points indicates the distribution of SHAP values for each feature.

The SHAP outcomes show that only two types of features are the best EEG characteristic measures for seizure prediction using extra tree classifiers. These are maximum frequency and zero-crossing values (maximum linear cross-correlation) from a few channels. We have seen that for different patients, different electrodes are important for feature selection to contribute to the performance of a prediction model.

For different patients, the same type of features have been selected but from different electrodes. This is because channel placement positions on the patient's brain is different from patient to patient. The overall evaluation focuses purely on the Kaggle contest data to be able to compare the novel algorithm with the benchmark algorithms on the same dataset. This evaluation consists of the average AUC performance across all three patients and the combined training and testing (prediction) time for all segments in each of the three patients. We have seen that on average, maximum linear cross-correlation appears in 10 out off 15 patients while the maximum frequency feature appears in 8 out off 15 patients.

The results are presented in the table 4.6. These results show that even though only 20 features have been used for model training in our proposed algorithm as compared

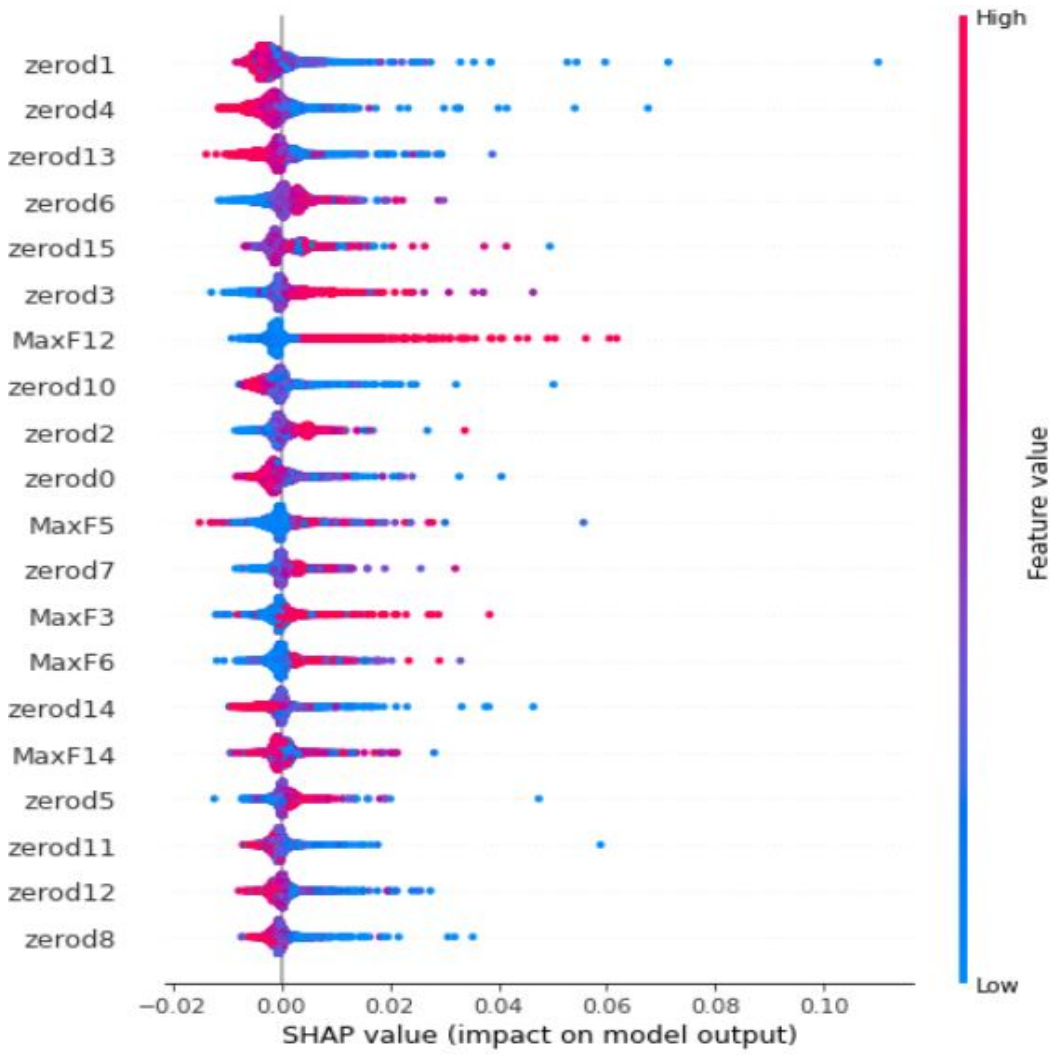


FIGURE 4.4: An example of feature selection with the SHAP algorithm for patient 1. We can see that only two types of features, zero crossing and maximum frequency are selected.

to 2264 features in [21] and 208 features in [3], the novel method improves upon the average performance of these benchmark algorithms by 6% and 8%, respectively. For patient specific performance, the improvement has been achieved for patient 1 and 2 with a decline in performance for patient 3 but for all patients the performance is better than chance prediction of AUC=0.5. The algorithm comparison in terms of training and prediction time is captured in figure 4.5 and shows that the new approach utilises less computation time for both combined training and prediction time compared to its benchmarks.

The extension of this study to the full data set of 15 patients shows better performance than chance prediction across all 15 patients with average AUC performance of 0.73,

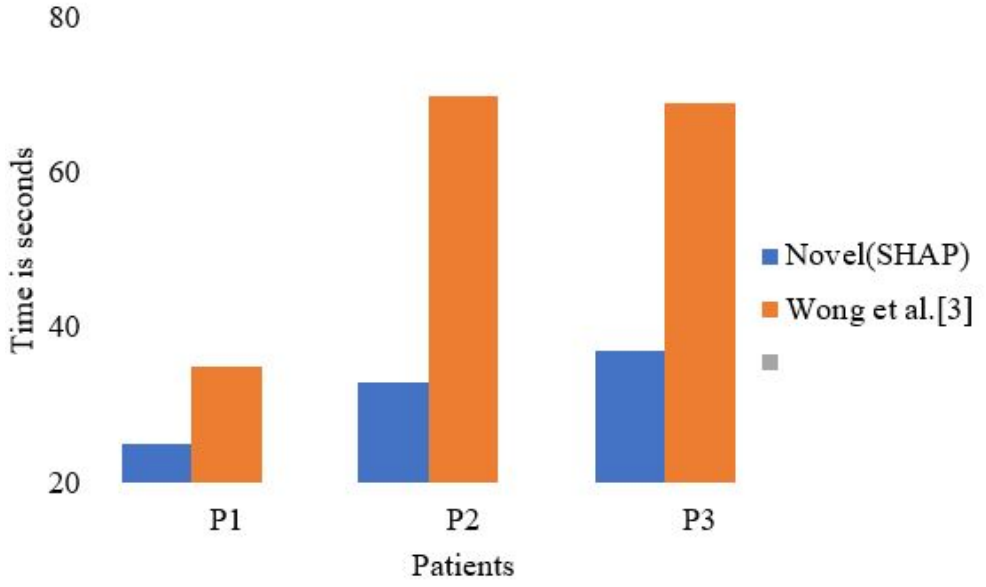


FIGURE 4.5: The algorithm comparison based on computational time for test set 2. The new designed algorithm is less computationally complex than the benchmark[3]

Patients	The second place[21]	Wong et al.[3]	Novel algorithm
1	0.620	0.68	0.830
2	0.740	0.71	0.871
3	0.810	0.84	0.699
Av	0.720	0.740	0.800

TABLE 4.6: The performance comparison of the new algorithm and the benchmarks in terms of AUC value.

0.65, 0.69 for test 1, test 2 and test 3, respectively. See Figure4.2(a). However, the performance for each patient for both algorithms showed poor performance that was less than chance prediction for patient 5 and 7 while on 98% of the patients, the both classifiers have better performance than chance prediction. See figure 4.2(b). However, statistically there were no significant differences between the performance of the two algorithms. Our study has shown that maximum linear cross correlation(Zero crossing values) can be a better predictor for seizure prediction algorithm for a generalized seizure prediction model. This is presented in the figure 4.7 where most of the features selected by SHAP are zero crossing for various electrodes

The approach is extended to the full long-term continuous dataset of 15 patients and the performance is compared to the algorithm in [3]. The results show that, using 20 features achieved almost the same results with the algorithm with 208 features. However, statistical analysis shows that there is a weak significant difference between the performance of the two algorithms for test 2 with (P-value = 0.0282) at the significant

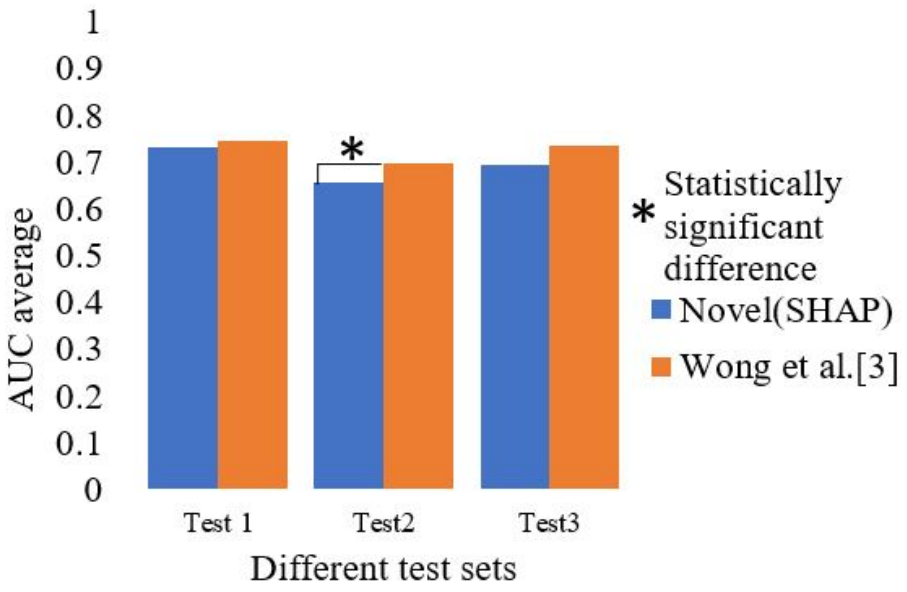


FIGURE 4.6: The algorithm comparison based AUC values for the different test sets. Our new proposed algorithm has almost similar performance for two test sets and there is significant difference only for the test set 2.

level of 0.05 and no significant difference for test 1 (P-values=0.1148) and 3(P-values =0.9092) for the same significant level. Hence, the new approach has improved the model computational cost as well as maintaining the performance with slightly different AUC values where the highest difference is 1%. This is shown on the figure 4.6 .

4.3.2 Clinical Evaluation

In our study, the extra tree classifier is among the studied algorithms that have shown better performance in terms of area under the curve as well as being the most computationally efficient in the 2016 Kaggle competition. We would like to evaluate the model using the clinically relevant performance metrics defined in section 2. To perform such an evaluation, we have computed the optimal threshold on the training data and use these optimal threshold values for each patient to compute the model sensitivity and percent of time in warning on test set 2 (validation set). The results are presented in the table 4.7 including the comparison with the recent benchmark algorithms.

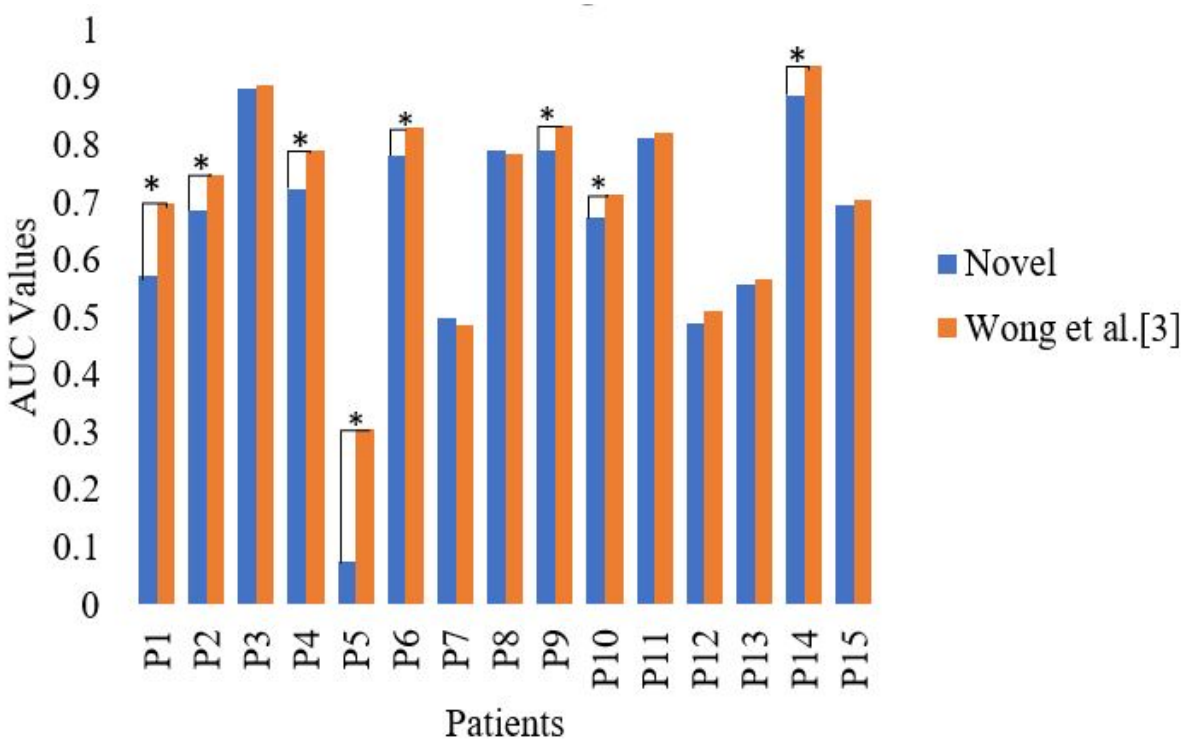


FIGURE 4.7: The algorithm comparison based AUC values for test set 2 (validation set) with the benchmark and reduced features using SHAP.

P	CW-Bayes [49]		Kaggle[21]		CW-Kaggle		Critical [9]		ET	
	Sens %	TiW %	Sens%	TiW %	Sens %	TiW %	Sens %	TiW %	Sens %	TiW %
1	61	27	n/a	n/a	n/a	n/a	68	7	44	5
2	n/a	n/a	n/a	n/a	n/a	n/a	75	0.7	37	6
3	55	29	66	29	60	26	n/a	n/a	40	9
4	n/a	n/a	n/a	n/a	n/a	n/a	64	0.02	44	8
5	n/a	n/a	n/a	n/a	n/a	n/a	72	2	33	5
6	n/a	n/a	n/a	n/a	n/a	n/a	72	2	48	3
7	n/a	n/a	n/a	n/a	n/a	n/a	76	21	42	9
8	76	28	n/a	n/a	n/a	n/a	67	0.14	45	6
9	45	11	39	11	52	11	75	10	48	6
10	52	17	48	17	53	17	69	13	44	7
11	58	15	n/a	n/a	n/a	n/a	83	17	49	7
12	n/a	n/a	n/a	n/a	n/a	n/a	54	0.000002	n/a	n/a
13	76	28	n/a	n/a	n/a	n/a	64	14	19	5
14	n/a	n/a	n/a	n/a	n/a	n/a	75	0.00006	n/a	n/a
15	60	21	n/a	n/a	n/a	n/a	88	2	37	5
Average	(61)	(21)	n/a	n/a	n/a	n/a	72(71)	8(9)	41(35)	6(5)

TABLE 4.7: The results comparison of the novel model with some of the algorithms in 2.3 developed with long-term human NeuroVista EEG dataset. The two values in brackets shows the average without n/a (not available)and with all n/a (15 patients). The new developed algorithm present the results for all patients except, patient 12 and 14 and the 10 minutes segment can be classified with less time in warning that the previous algorithms developed on the same dataset.

The results show that in almost all patients, the preictal segments can be classified with the exception of patient 12 and 14. This is due to the fact that in the dataset for those two patients, there are not many preictal segments and the prediction probabilities calculated for each segment are very small relative to the optimal threshold. We compare the prediction probability for each segment with the optimal threshold where all segments that have a prediction probability less than optimal threshold are classified as interictal and all segments where the prediction probability is greater than the optimal threshold are classified as preictal segments. Comparing the model sensitivity for the critical slowing down paper in [9] with the results in this thesis, our results are poor in terms of sensitivity with improved percent of time in warning. The critical slowing down algorithm is a long-time scale analysis algorithm and also represents the state of the art in seizure prediction across both short- and long-time scale feature analysis because it has achieved the highest performance yet of any algorithm applied to the full NeuroVista trial dataset. As such, based on the comparison here with the short-time scale analysis algorithm, it seems long-time scale analysis is the best path moving forward when trying to find algorithms that can achieve clinically useful levels of seizure prediction performance.

Chapter 5

Conclusion

In this thesis, the aim was to reliably estimate the performance of short-time scale analysis seizure prediction algorithms on long-term intracranial EEG recordings, and to see if there are any stand-out EEG features or classifiers for seizure prediction. Through the proposed work we have analysed the performance of 9 classification algorithms for seizure prediction (Part 1 task) and also developed a more efficient high performing seizure prediction algorithm based on simplifying the benchmark algorithm used in this study (part 2 task). This has been done for contest data for 3 patients and for long-term continuous intracranial EEG data of 15 patients. The same features in the second place algorithm in Kaggle for seizure prediction have been used given that this algorithm was the most computationally efficient algorithm in the competition, and therefore provides a sensible benchmark for developing efficient seizure prediction algorithms for implantable/wearable devices. For the contest data, on average extra tree classifiers perform better than other classifiers with group k-fold cross validation. On the long-term continuous EEG dataset of 15 patients, on average LDA outperforms others with stratified cross validation in terms of AUC. We have statistically compared the performance of the algorithms and found that on average LDA outperforms other classifiers at the group level and there is a statically significant difference in performance based on Bonferroni correction of alpha (0.05/9) when the algorithm performance is compared against algorithm chance prediction (AUC=0.5). The comparison of the performance of the benchmark algorithm (extra tree classifier) against other classifiers in this study shows no significant difference between its performance and the performance of each of the other classifiers in this study based on two sample t-testing. The performance analysis of a classifier for each patient

for different test sets shows that there is a variability of algorithm performance and a classifier may give different prediction results for different patients. This shows that finding a good seizure prediction algorithm is a complicated problem and there is no one size that fits all algorithm. Future work, could consider the complexity of algorithms to reduce the computational complexity by focusing of the importance of features to distinguish preictal from interictal window segments for seizure prediction model based on LDA as the best classifier in this thesis.

Overall the results suggest that short-time scale analysis can produce above chance prediction performance, however, it remains a challenge to produce clinically useful levels of seizure prediction performance. Perhaps a combination of short- and long-time scale analysis could yield super-additive results that push performance levels above those yielded by state of the art long-time scale analysis algorithms.

Appendix A

An Appendix

TABLE A.1: The statistical analysis comparison of algorithm performance against each other this equal to 36 comparisons

Classifiers	Statistical test	P values	Alpha	Significant?	Alpha/36	significant?
ET vs SVM	Two Sample T-test	0.0923	0.05	NO	0.0013	NO
ET vs RF	Two Sample T-test	0.4733	0.05	NO	0.0013	NO
ET vs LR	Two Sample T-test	0.0631	0.05	NO	0.0013	NO
ET vs LDA	Two Sample T-test	0.7031	0.05	NO	0.0013	NO
ET vs AB	Two Sample T-test	0.8301	0.05	NO	0.0013	NO
ET vs EGB	Two Sample T-test	0.8780	0.05	NO	0.0013	NO
ET vs KNN	Two Sample T-test	0.1887	0.05	NO	0.0013	NO
ET vs GNB	Two Sample T-test	0.4799	0.05	NO	0.0013	NO
SVM vs RF	Two Sample T-test	0.3281	0.05	NO	0.0013	NO
SVM vs LR	Two Sample T-test	0.7903	0.05	NO	0.0013	NO
SVM vs LDA	Two Sample T-test	0.0203	0.05	**YES**	0.0013	NO
SVM vs AB	Two Sample T-test	0.1632	0.05	NO	0.0013	NO
SVM vs EGB	Two Sample T-test	0.0696	0.05	NO	0.0013	NO
SVM vs KNN	Two Sample T-test	0.7025	0.05	NO	0.0013	NO
SVM vs GNB	Two Sample T-test	0.3464	0.05	NO	0.0013	NO
RF vs RL	Two Sample T-test	0.2375	0.05	NO	0.0013	NO
RF vs LDA	Two Sample T-test	0.2371	0.05	NO	0.0013	NO
RF vs AB	Two Sample T-test	0.6361	0.05	NO	0.0013	NO
RF vs EGB	Two Sample T-test	0.5121	0.05	NO	0.0013	NO
RF vs KNN	Two Sample T-test	0.5487	0.05	NO	0.0013	NO
RF vs GNB	Two Sample T-test	0.9956	0.05	NO	0.0013	NO
LR vs AB	Two Sample T-test	0.0126	0.05	**YES**	0.0013	NO
LR vs EGB	Two Sample T-test	0.0441	0.05	NO	0.0013	NO
LR vs KNN	Two Sample T-test	0.5351	0.05	NO	0.0013	NO
LR vs LDA	Two Sample T-test	0.2536	0.05	NO	0.0013	NO
LR vs GNB	Two Sample T-test	0.3276	0.05	NO	0.0013	NO
LDA vs AB	Two Sample T-test	0.5484	0.05	NO	0.0013	NO
LDA vs EGB	Two Sample T-test	0.5408	0.05	NO	0.0013	NO
LDA vs KNN	Two Sample T-test	0.0612	0.05	NO	0.0013	NO
LDA vs GNB	Two Sample T-test	0.2474	0.05	NO	0.0013	NO
AB vs GNB	Two Sample T-test	0.9269	0.05	NO	0.0013	NO
AB vs KNN	Two Sample T-test	0.2973	0.05	NO	0.0013	NO
AB vs GNB	Two Sample T-test	0.6395	0.05	NO	0.0013	NO
EGB vs KNN	Two Sample T-test	0.1752	0.05	NO	0.0013	NO
EGB vs GNB	Two Sample T-test	0.5202	0.05	NO	0.0013	NO
KNN vs GNB	Two Sample T-test	0.5637	0.05	NO	0.0013	NO

Bibliography

[1] Levin Kuhlmann, Klaus Lehnertz, Mark P Richardson, Björn Schelter, and Hitten P Zaveri. Seizure prediction—ready for a new era. *Nature Reviews Neurology*, 14(10):618–630, 2018.

[2] Alexandre Gramfort, Martin Luessi, Eric Larson, Denis A. Engemann, Daniel Strohmeier, Christian Brodbeck, Roman Goj, Mainak Jas, Teon Brooks, Lauri Parkkonen, and Matti S. Hämäläinen. MEG and EEG data analysis with MNE-Python. *Frontiers in Neuroscience*, 7(267):1–13, 2013. doi: 10.3389/fnins.2013.00267.

[3] Sheng Wong and Levin Kuhlmann. Computationally efficient epileptic seizure prediction based on extremely randomised trees. In *Proceedings of the Australasian Computer Science Week Multiconference*, pages 1–3, 2020.

[4] Tarun Dua, Hanneke M De Boer, Leonid L Prilipko, and Shekhar Saxena. Epilepsy care in the world: results of an ilae/ibe/who global campaign against epilepsy survey. *Epilepsia*, 47(7):1225–1231, 2006.

[5] Michael I Jordan and Tom M Mitchell. Machine learning: Trends, perspectives, and prospects. *Science*, 349(6245):255–260, 2015.

[6] K Lehnertz and CE Elger. Spatio-temporal dynamics of the primary epileptogenic area in temporal lobe epilepsy characterized by neuronal complexity loss. *Electroencephalography and clinical Neurophysiology*, 95(2):108–117, 1995.

[7] Robert S Fisher, Helen E Scharfman, and Marco DeCurtis. How can we identify ictal and interictal abnormal activity? *Issues in Clinical Epileptology: A View from the Bench*, pages 3–23, 2014.

[8] Maxime O Baud, Jonathan K Kleen, Emily A Mirro, Jason C Andrechak, David King-Stephens, Edward F Chang, and Vikram R Rao. Multi-day rhythms modulate seizure risk in epilepsy. *Nature communications*, 9(1):1–10, 2018.

[9] Matias I Maturana, Christian Meisel, Katrina Dell, Philippa J Karoly, Wendyl D'Souza, David B Grayden, Anthony N Burkitt, Premysl Jiruska, Jan Kudlacek, Jaroslav Hlinka, et al. Critical slowing down as a biomarker for seizure susceptibility. *Nature Communications*, 11(1):1–12, 2020.

[10] Vibha Patel, Sanjay Buch, and Amit Ganatra. A review on eeg based epileptic seizure prediction using machine learning techniques. In *International Conference on Intelligent Computing, Information and Control Systems*, pages 384–391. Springer, 2019.

[11] A Gorji and M Khaleghi Ghadiri. History of epilepsy in medieval iranian medicine. *Neuroscience & Biobehavioral Reviews*, 25(5):455–461, 2001.

[12] Klaus Lehnertz, Florian Mormann, Hannes Osterhage, Andy Müller, Jens Prusseit, Anton Chernihovskyi, Matthäus Staniek, Dieter Krug, Stephan Bialonski, and Christian E Elger. State-of-the-art of seizure prediction. *Journal of Clinical Neurophysiology*, 24(2):147–153, 2007.

[13] Mark J Cook, Terence J O'Brien, Samuel F Berkovic, Michael Murphy, Andrew Morokoff, Gavin Fabinyi, Wendyl D'Souza, Raju Yerra, John Archer, Lucas Litewka, et al. 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, 2013.

[14] Dieter Schmidt and Steven C Schachter. Drug treatment of epilepsy in adults. *Bmj*, 348, 2014.

[15] Patrick Kwan and Martin J Brodie. Early identification of refractory epilepsy. *New England Journal of Medicine*, 342(5):314–319, 2000.

[16] Nathan Zasler, Douglas Katz, and Ross Zafonte. *Brain injury medicine: Principles and practice*. Demos Medical Publishing, 2007.

[17] Tobias Loddenkemper, Andrew Pan, Silvia Neme, Kenneth B Baker, Ali R Rezai, Dudley S Dinner, Erwin B Montgomery Jr, and Hans O Lüders. Deep brain stimulation in epilepsy. *Journal of Clinical Neurophysiology*, 18(6):514–532, 2001.

[18] Arman Ahnood, Duy Truong, Bobbi Fleiss, Armin Nikpour, and Omid Kavehei. Novel electrode architecture for subgaleal electroencephalography: A feasibility study. *Sensors & Diagnostics*, 2022.

[19] Bill Byrom, Marie McCarthy, Peter Schueler, and Willie Muehlhausen. Brain monitoring devices in neuroscience clinical research: the potential of remote monitoring using sensors, wearables, and mobile devices. *Clinical Pharmacology & Therapeutics*, 104(1):59–71, 2018.

[20] Benjamin H Brinkmann, Joost Wagenaar, Drew Abbot, Phillip Adkins, Simone C Bosshard, Min Chen, Quang M Tieng, Jialune He, FJ Muñoz-Almaraz, Paloma Botella-Rocamora, et al. Crowdsourcing reproducible seizure forecasting in human and canine epilepsy. *Brain*, 139(6):1713–1722, 2016.

[21] Levin Kuhlmann, Philippa Karoly, Dean R Freestone, Benjamin H Brinkmann, Andriy Temko, Alexandre Barachant, Feng Li, Gilberto Titericz Jr, Brian W Lang, Daniel Lavery, et al. Epilepsyecosystem. org: crowd-sourcing reproducible seizure prediction with long-term human intracranial eeg. *Brain*, 141(9):2619–2630, 2018.

[22] Paul R Carney, Stephen Myers, and James D Geyer. Seizure prediction: methods. *Epilepsy & behavior*, 22:S94–S101, 2011.

[23] SS Viglione and GO Walsh. Proceedings: Epileptic seizure prediction. *Electroencephalography and clinical neurophysiology*, 39(4):435–436, 1975.

[24] Z Rogowski, I Gath, and E Bental. On the prediction of epileptic seizures. *Biological cybernetics*, 42(1):9–15, 1981.

[25] Brian Litt, Rosana Esteller, Javier Echauz, Maryann D'Alessandro, Rachel Shor, Thomas Henry, Page Pennell, Charles Epstein, Roy Bakay, Marc Dichter, et al. Epileptic seizures may begin hours in advance of clinical onset: a report of five patients. *Neuron*, 30(1):51–64, 2001.

[26] Xiaoli Li and G Ouyang. Nonlinear similarity analysis for epileptic seizures prediction. *Nonlinear Analysis: Theory, Methods & Applications*, 64(8):1666–1678, 2006.

[27] Ali Shoeb, Herman Edwards, Jack Connolly, Blaise Bourgeois, S Ted Treves, and John Guttag. Patient-specific seizure onset detection. *Epilepsy & Behavior*, 5(4): 483–498, 2004.

[28] M Valderrama, S Nikolopoulos, C Adam, Vincent Navarro, and M Quyen. Patient-specific seizure prediction using a multi-feature and multi-modal eeg-ecg classification. In *XII Mediterranean Conference on Medical and Biological Engineering and Computing 2010*, pages 77–80. Springer, 2010.

[29] Florian Mormann, Ralph G Andrzejak, Christian E Elger, and Klaus Lehnertz. Seizure prediction: the long and winding road. *Brain*, 130(2):314–333, 2007.

[30] Sriram Ramgopal, Sigride Thome-Souza, Michele Jackson, Navah Ester Kadish, Iván Sánchez Fernández, Jacquelyn Klehm, William Bosl, Claus Reinsberger, Steven Schachter, and Tobias Loddenkemper. Seizure detection, seizure prediction, and closed-loop warning systems in epilepsy. *Epilepsy & behavior*, 37:291–307, 2014.

[31] Florian Mormann, Thomas Kreuz, Christoph Rieke, Ralph G Andrzejak, Alexander Kraskov, Peter David, Christian E Elger, and Klaus Lehnertz. On the predictability of epileptic seizures. *Clinical neurophysiology*, 116(3):569–587, 2005.

[32] Nhan Duy Truong, Anh Duy Nguyen, Levin Kuhlmann, Mohammad Reza Bonyadi, Jiawei Yang, Samuel Ippolito, and Omid Kavehei. Convolutional neural networks for seizure prediction using intracranial and scalp electroencephalogram. *Neural Networks*, 105:104–111, 2018.

[33] Rui P Costa, Pedro Oliveira, Guilherme Rodrigues, Bruno Leitao, and António Dourado. Epileptic seizure classification using neural networks with 14 features. In *International Conference on Knowledge-Based and Intelligent Information and Engineering Systems*, pages 281–288. Springer, 2008.

[34] Meng Xi, Liang Chen, Desanka Polajnar, and Weiyang Tong. Local binary pattern network: a deep learning approach for face recognition. In *2016 IEEE international conference on Image processing (ICIP)*, pages 3224–3228. IEEE, 2016.

[35] Turky N Alotaiby, Saleh A Alshebeili, Faisal M Alotaibi, and Saud R Alrshoud. Epileptic seizure prediction using csp and lda for scalp eeg signals. *Computational intelligence and neuroscience*, 2017, 2017.

[36] Mojtaba Bandarabadi, César A Teixeira, Jalil Rasekhi, and António Dourado. Epileptic seizure prediction using relative spectral power features. *Clinical Neurophysiology*, 126(2):237–248, 2015.

[37] Jacek Biesiada and Wlodzisław Duch. Feature selection for high-dimensional data—a pearson redundancy based filter. In *Computer recognition systems 2*, pages 242–249. Springer, 2007.

[38] Negin Moghim and David W Corne. Predicting epileptic seizures in advance. *PloS one*, 9(6):e99334, 2014.

[39] Klaus Lehnertz and Christian E Elger. Can epileptic seizures be predicted? evidence from nonlinear time series analysis of brain electrical activity. *Physical review letters*, 80(22):5019, 1998.

[40] Leonidas D Iasemidis and James Chris Sackellares. Seizure warning and prediction, October 16 2001. US Patent 6,304,775.

[41] Florian Mormann, Ralph G Andrzejak, Christian E Elger, and Klaus Lehnertz. Seizure prediction: the long and winding road. *Brain*, 130(2):314–333, 2006.

[42] Richard Aschenbrenner-Scheibe, Thomas Maiwald, Matthias Winterhalder, Henning U Voss, Jens Timmer, and Andreas Schulze-Bonhage. How well can epileptic seizures be predicted? an evaluation of a nonlinear method. *Brain*, 126(12):2616–2626, 2003.

[43] M Winterhalder, T Maiwald, HU Voss, R Aschenbrenner-Scheibe, J Timmer, and A Schulze-Bonhage. The seizure prediction characteristic: a general framework to assess and compare seizure prediction methods. *Epilepsy & Behavior*, 4(3): 318–325, 2003.

[44] Thomas Maiwald, Matthias Winterhalder, Richard Aschenbrenner-Scheibe, Henning U Voss, Andreas Schulze-Bonhage, and Jens Timmer. Comparison of three nonlinear seizure prediction methods by means of the seizure prediction characteristic. *Physica D: nonlinear phenomena*, 194(3-4):357–368, 2004.

[45] Wim De Clercq, Philippe Lemmerling, Sabine Van Huffel, and Wim Van Paesschen. Anticipation of epileptic seizures from standard eeg recordings. *The Lancet*, 361(9361):971, 2003.

[46] Klaus Lehnertz and Brian Litt. The first international collaborative workshop on seizure prediction: summary and data description. *Clinical neurophysiology*, 116 (3):493–505, 2005.

[47] Maryann D'Alessandro, George Vachtsevanos, Rosana Esteller, Javier Echauz, Stephen Cranstoun, Greg Worrell, Landi Parish, and Brian Litt. A multi-feature and multi-channel univariate selection process for seizure prediction. *Clinical neurophysiology*, 116(3):506–516, 2005.

[48] Benjamin H Brinkmann, Edward E Patterson, Charles Vite, Vincent M Vasoli, Daniel Crepeau, Matt Stead, J Jeffry Howbert, Vladimir Cherkassky, Joost B Wagenaar, Brian Litt, et al. Forecasting seizures using intracranial eeg measures and svm in naturally occurring canine epilepsy. *PloS one*, 10(8):e0133900, 2015.

[49] Philippa J Karoly, Hoameng Ung, David B Grayden, Levin Kuhlmann, Kent Leyde, Mark J Cook, and Dean R Freestone. The circadian profile of epilepsy improves seizure forecasting. *Brain*, 140(8):2169–2182, 2017.

[50] Isabell Kiral-Kornek, Subhrajit Roy, Ewan Nurse, Benjamin Mashford, Philippa Karoly, Thomas Carroll, Daniel Payne, Susmita Saha, Steven Baldassano, Terence O'Brien, et al. Epileptic seizure prediction using big data and deep learning: toward a mobile system. *EBioMedicine*, 27:103–111, 2018.

[51] Hinnerk Feldwisch-Drentrup, Björn Schelter, Michael Jachan, Jakob Nawrath, Jens Timmer, and Andreas Schulze-Bonhage. Joining the benefits: combining epileptic seizure prediction methods. *Epilepsia*, 51(8):1598–1606, 2010.

[52] Shouyi Wang, Wanpracha Art Chaovalitwongse, and Stephen Wong. Online seizure prediction using an adaptive learning approach. *IEEE transactions on knowledge and data engineering*, 25(12):2854–2866, 2013.

[53] Cao Xiao, Shouyi Wang, Leon Iasemidis, Stephen Wong, and Wanpracha Art Chaovalitwongse. An adaptive pattern learning framework to personalize online seizure prediction. *IEEE Transactions on Big Data*, 2017.

[54] J Chris Sackellares, Deng-Shan Shiau, Jose C Principe, Mark CK Yang, Linda K Dance, Wichai Suharitdamrong, Wanpracha Chaovalitwongse, Panos M Pardalos, and Leonidas D Iasemidis. Predictability analysis for an automated seizure prediction algorithm. *Journal of Clinical Neurophysiology*, 23(6):509–520, 2006.

[55] Levin Kuhlmann, Dean Freestone, Alan Lai, Anthony N Burkitt, Karen Fuller, David B Grayden, Linda Seiderer, Simon Vogrin, Iven MY Mareels, and Mark J Cook. Patient-specific bivariate-synchrony-based seizure prediction for short prediction horizons. *Epilepsy research*, 91(2-3):214–231, 2010.

[56] Mario Valderrama, C Alvarado, S Nikolopoulos, Jacques Martinerie, Claude Adam, Vincent Navarro, and Michel Le Van Quyen. Identifying an increased risk of epileptic seizures using a multi-feature eeg–ecg classification. *Biomedical Signal Processing and Control*, 7(3):237–244, 2012.

[57] Jalil Rasekhi, Mohammad Reza Karami Mollaei, Mojtaba Bandarabadi, Cesar A Teixeira, and Antonio Dourado. Preprocessing effects of 22 linear univariate features on the performance of seizure prediction methods. *Journal of neuroscience methods*, 217(1-2):9–16, 2013.

[58] Catalina Alvarado-Rojas, Mario Valderrama, A Fouad-Ahmed, H Feldwisch-Drentrup, M Ihle, CA Teixeira, F Sales, A Schulze-Bonhage, C Adam, Antonio Dourado, et al. Slow modulations of high-frequency activity (40–140 hz) discriminate preictal changes in human focal epilepsy. *Scientific reports*, 4:4545, 2014.

[59] César Alexandre Teixeira, Bruno Direito, Mojtaba Bandarabadi, Michel Le Van Quyen, Mario Valderrama, Bjoern Schelter, Andreas Schulze-Bonhage, Vincent Navarro, Francisco Sales, and António Dourado. Epileptic seizure predictors based on computational intelligence techniques: A comparative study with 278 patients. *Computer methods and programs in biomedicine*, 114(3):324–336, 2014.

[60] Jalil Rasekhi, Mohammad Reza Karami Mollaei, Mojtaba Bandarabadi, César A Teixeira, and António Dourado. Epileptic seizure prediction based on ratio and differential linear univariate features. *Journal of medical signals and sensors*, 5(1): 1, 2015.

[61] Bruno Direito, César A Teixeira, Francisco Sales, Miguel Castelo-Branco, and António Dourado. A realistic seizure prediction study based on multiclass svm. *International journal of neural systems*, 27(03):1750006, 2017.

[62] Kais Gadhoumi, Jean Gotman, and Jean Marc Lina. Scale invariance properties of intracerebral eeg improve seizure prediction in mesial temporal lobe epilepsy. *PloS one*, 10(4):e0121182, 2015.

[63] Vanessa Senger and Ronald Tetzlaff. New signal processing methods for the development of seizure warning devices in epilepsy. *IEEE Transactions on Circuits and Systems I: Regular Papers*, 63(5):609–616, 2016.

[64] Piotr Mirowski, Deepak Madhavan, Yann LeCun, and Ruben Kuzniecky. Classification of patterns of eeg synchronization for seizure prediction. *Clinical neurophysiology*, 120(11):1927–1940, 2009.

[65] Yun Park, Lan Luo, Keshab K Parhi, and Theoden Netoff. Seizure prediction with spectral power of eeg using cost-sensitive support vector machines. *Epilepsia*, 52 (10):1761–1770, 2011.

[66] Ardalan Aarabi and Bin He. A rule-based seizure prediction method for focal neocortical epilepsy. *Clinical Neurophysiology*, 123(6):1111–1122, 2012.

[67] James R Williamson, Daniel W Bliss, David W Browne, and Jaishree T Narayanan. Seizure prediction using eeg spatiotemporal correlation structure. *Epilepsy & behavior*, 25(2):230–238, 2012.

[68] Shufang Li, Weidong Zhou, Qi Yuan, and Yinxia Liu. Seizure prediction using spike rate of intracranial eeg. *IEEE transactions on neural systems and rehabilitation engineering*, 21(6):880–886, 2013.

[69] Yang Zheng, Gang Wang, Kuo Li, Gang Bao, and Jue Wang. Epileptic seizure prediction using phase synchronization based on bivariate empirical mode decomposition. *Clinical Neurophysiology*, 125(6):1104–1111, 2014.

[70] Amir Eftekhar, Walid Juffali, Jamil El-Imad, Timothy G Constandinou, and Christofer Toumazou. Ngram-derived pattern recognition for the detection and prediction of epileptic seizures. *PloS one*, 9(6):e96235, 2014.

[71] Ardalan Aarabi and Bin He. Seizure prediction in hippocampal and neocortical epilepsy using a model-based approach. *Clinical Neurophysiology*, 125(5):930–940, 2014.

[72] Zisheng Zhang and Keshab K Parhi. Low-complexity seizure prediction from ieeg/seeg using spectral power and ratios of spectral power. *IEEE transactions on biomedical circuits and systems*, 10(3):693–706, 2015.

[73] K.K others Zhang, Z.and Parhi. 37th annual international conference of the ieee engineering in medicine and biology society (embc). *IEEE*, 2015:5748–5751–6, 2015.

[74] Ali Shahidi Zandi, Reza Tafreshi, Manouchehr Javidan, and Guy A Dumont. Predicting epileptic seizures in scalp eeg based on a variational bayesian gaussian mixture model of zero-crossing intervals. *IEEE Transactions on Biomedical Engineering*, 60(5):1401–1413, 2013.

[75] Jui-Hong Chien, Deng-Shan Shiau, JJ Halford, KM Kelly, RT Kern, MCK Yang, Jicong Zhang, J Ch Sackellares, and PM Pardalos. A signal regularity-based automated seizure prediction algorithm using long-term scalp eeg recordings. *Cybernetics and Systems Analysis*, 47(4):586, 2011.

[76] Kais Gadhoumi, Jean-Marc Lina, and Jean Gotman. Seizure prediction in patients with mesial temporal lobe epilepsy using eeg measures of state similarity. *Clinical Neurophysiology*, 124(9):1745–1754, 2013.

[77] David E Snyder, Javier Echauz, David B Grimes, and Brian Litt. The statistics of a practical seizure warning system. *Journal of neural engineering*, 5(4):392, 2008.

[78] Philippa J Karoly, Dean R Freestone, Ray Boston, David B Grayden, David Himes, Kent Leyde, Udaya Seneviratne, Samuel Berkovic, Terence O'Brien, and Mark J Cook. Interictal spikes and epileptic seizures: their relationship and underlying rhythmicity. *Brain*, 139(4):1066–1078, 2016.

[79] Chip Reuben, Philippa Karoly, Dean R Freestone, Andriy Temko, Alexandre Barachant, Feng Li, Gilberto Titericz Jr, Brian W Lang, Daniel Lavery, Kelly Roman, et al. Ensembling crowdsourced seizure prediction algorithms using long-term human intracranial eeg. *Epilepsia*, 2019.

[80] Theoden Netoff, Yun Park, and Keshab Parhi. Seizure prediction using cost-sensitive support vector machine. In *2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 3322–3325. IEEE, 2009.

[81] Han-Tai Shiao, Vladimir Cherkassky, Jieun Lee, Brandon Veber, Edward E Patterson, Benjamin H Brinkmann, and Gregory A Worrell. Svm-based system for prediction of epileptic seizures from ieeg signal. *IEEE Transactions on Biomedical Engineering*, 64(5):1011–1022, 2016.

[82] Edward E Patterson. Canine epilepsy: an underutilized model. *ILAR journal*, 55 (1):182–186, 2014.

[83] Zisheng Zhang and Keshab K Parhi. Seizure prediction using polynomial svm classification. In *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 5748–5751. IEEE, 2015.

[84] Cristian Donos, Matthias Dümpelmann, and Andreas Schulze-Bonhage. Early seizure detection algorithm based on intracranial eeg and random forest classification. *International journal of neural systems*, 25(05):1550023, 2015.

[85] Yuxing Wang, Jiuwen Cao, Xiaoping Lai, and Dinghan Hu. Epileptic state classification for seizure prediction with wavelet packet features and random forest. In *2019 Chinese Control And Decision Conference (CCDC)*, pages 3983–3987. IEEE, 2019.

[86] Md Mursalin, Yuan Zhang, Yuehui Chen, and Nitesh V Chawla. Automated epileptic seizure detection using improved correlation-based feature selection with random forest classifier. *Neurocomputing*, 241:204–214, 2017.

[87] Gregor Stiglic, Petra Povalej Brzan, Nino Fijacko, Fei Wang, Boris Delibasic, Alexandros Kalousis, and Zoran Obradovic. Comprehensible predictive modeling using regularized logistic regression and comorbidity based features. *PloS one*, 10 (12), 2015.

[88] Abdulhamit Subasi and Ergun Ercelebi. Classification of eeg signals using neural network and logistic regression. *Computer methods and programs in biomedicine*, 78(2):87–99, 2005.

[89] Robert E Schapire. Explaining adaboost. In *Empirical inference*, pages 37–52. Springer, 2013.

[90] Tianqi Chen, Tong He, Michael Benesty, Vadim Khotilovich, Yuan Tang, Hyunsu Cho, Kailong Chen, et al. Xgboost: extreme gradient boosting. *R package version 0.4-2*, 1(4):1–4, 2015.

[91] Nathaniel R Goodman. Statistical analysis based on a certain multivariate complex gaussian distribution (an introduction). *The Annals of mathematical statistics*, 34 (1):152–177, 1963.

[92] Yusuf U Khan, Omar Farooq, and Priyanka Sharma. Automatic detection of seizure onset in pediatric eeg. *International Joural of Embeded Systems and Applications*, 2(3):81–89, 2012.

[93] DK Ravish, S Shenbaga Devi, SG Krishnamoorthy, and MR Karthikeyan. Detection of epileptic seizure in eeg recordings by spectral method and statistical analysis. *Journal of Applied Sciences*, 13(2):207–219, 2013.

[94] Sahar Elgohary, Seif Eldawlatly, and Mahmoud I Khalil. Epileptic seizure prediction using zero-crossings analysis of eeg wavelet detail coefficients. In *2016 IEEE conference on computational intelligence in bioinformatics and computational biology (CIBCB)*, pages 1–6. IEEE, 2016.

[95] Bo Hjorth. Eeg analysis based on time domain properties. *Electroencephalography and clinical neurophysiology*, 29(3):306–310, 1970.

[96] Seung-Hyeon Oh, Yu-Ri Lee, and Hyoung-Nam Kim. A novel eeg feature extraction method using hjorth parameter. *International Journal of Electronics and Electrical Engineering*, 2(2):106–110, 2014.

[97] Ali A Abdul-Latif, Irena Cosic, Dinesh K Kumar, Barbara Polus, and Cliff Da Costa. Power changes of eeg signals associated with muscle fatigue: the root mean square analysis of eeg bands. In *Proceedings of the 2004 Intelligent Sensors, Sensor Networks and Information Processing Conference, 2004.*, pages 531–534. IEEE, 2004.

[98] Kunjan Patel, Chern-Pin Chua, Stephen Fau, and Chris J Bleakley. Low power real-time seizure detection for ambulatory eeg. In *2009 3rd International Conference on Pervasive Computing Technologies for Healthcare*, pages 1–7. IEEE, 2009.

[99] Paul Fergus, David Hignett, Abir Hussain, Dhiya Al-Jumeily, and Khaled Abdel-Aziz. Automatic epileptic seizure detection using scalp eeg and advanced artificial intelligence techniques. *BioMed research international*, 2015, 2015.

[100] BR Greene, S Faul, WP Marnane, G Lightbody, I Korotchikova, and GB Boylan. A comparison of quantitative eeg features for neonatal seizure detection. *Clinical Neurophysiology*, 119(6):1248–1261, 2008.

[101] Lalit M Patnaik and Ohil K Manyam. Epileptic eeg detection using neural networks and post-classification. *Computer methods and programs in biomedicine*, 91 (2):100–109, 2008.

[102] Claude E Shannon. A mathematical theory of communication. *Bell system technical journal*, 27(3):379–423, 1948.

[103] Susana Blanco, Arturo Garay, and Diego Coulombie. Comparison of frequency bands using spectral entropy for epileptic seizure prediction. *ISRN neurology*, 2013, 2013.

[104] N Kannathal, Min Lim Choo, U Rajendra Acharya, and PK Sadasivan. Entropies for detection of epilepsy in eeg. *Computer methods and programs in biomedicine*, 80(3):187–194, 2005.

[105] Kamel Nidal and Aamir Saeed Malik. *EEG/ERP analysis: methods and applications*. Crc Press, 2014.

[106] Pinar Yildirim. Filter based feature selection methods for prediction of risks in hepatitis disease. *International Journal of Machine Learning and Computing*, 5 (4):258, 2015.

[107] Avrim L Blum and Pat Langley. Selection of relevant features and examples in machine learning. *Artificial intelligence*, 97(1-2):245–271, 1997.

[108] Arthur E Hoerl and Robert W Kennard. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1):55–67, 1970.

[109] Robert Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1):267–288, 1996.

[110] Mark Andrew Hall. Correlation-based feature selection for machine learning. 1999.

[111] Scott M Lundberg and Su-In Lee. A unified approach to interpreting model predictions. *Advances in neural information processing systems*, 30, 2017.

[112] Eyal Winter. The shapley value. *Handbook of game theory with economic applications*, 3:2025–2054, 2002.