

Analysis of Intracranial Electroencephalography for Prediction of Epileptic Seizure

Behrooz Abbaszadeh

Thesis submitted to the University of Ottawa
in partial Fulfillment of the requirements for the degree of
Doctor of Philosophy
in
Electrical and Computer Engineering

School of Electrical Engineering and Computer Science
Faculty of Engineering
University of Ottawa

Abstract

Epilepsy is one of the most common and devastating neurologic diseases, which affects over 70 million people around the world. For some patients, it can be managed with antiepileptic drugs. However, 20 to 30% of them would likely get worse after the initial improvement, and some may even remain refractory to the current medicine. Predicting seizures enough in advance could allow patients or caretakers to take appropriate actions and therefore, reduce the risk of injury.

This work focused on time-domain features to achieve discriminative information about EEG signals and at a lower CPU cost. Thus, Kruskal Wallis, a non-parametric approach, was found to perform better than other approaches due to its less resource consumption strategy while maintaining the highest Matthews's Correlation Coefficient (MCC) score. The performance of Kruskal Wallis may suggest considering the importance of univariate features like complexity and interquartile ratio along with auto regressive model parameters and maximum cross-correlation.

Furthermore, it has been demonstrated that dividing EEG to sub-bands will provide more discriminative information and that a 2-second window length provides the highest MCC score while dealing with the non-stationary behaviour of EEG signal for this window length. Consequently, the obtained results from 2-s window were fed to a binary (Extreme Gradient Boosting) classifier and the posterior probability of the test data extracted. The probabilistic framework requires the mean and maximum probability of the non-seizure and the seizure occurrence period segments. Once all these parameters were set for each patient, the medical decision maker can send alarm based on well-defined thresholds.

While finding a unique model for all patients is really challenging, our modelling results demonstrated that the proposed algorithm can be an efficient tool for reliable and clinically relevant seizure forecasting. Using iEEG signals, the proposed algorithm is capable of forecasting seizures, informing a patient about 75 minutes before a seizure would occur, a period large enough for patients to take practical actions to minimize the potential impacts of the seizure.

The proposed tool aims to be implemented in a low power portable device by considering few simple time-domain features for a specific sub-band. It should be noted that there are not many publications investigating frontal lobe epilepsy, making the findings in this work promising results.

Acknowledgments

Poem by Persian poet Saadi Shirazi (1210-1292):

"Human beings are members of a whole,

since in their creation they are of one essence.

When the conditions of the time brings a member (limb) to pain,

the other members (limbs) will suffer from discomfort.

You, who are indifferent to the misery of others,

it is not fitting that they should call you a human being."

In the name of the first supervisor in my life with the power of Psyche.

I am in the process of leaving from a cave, the Allegory of the Cave suggested by Plato, and I will continue to learn and reveal the hidden secrets of the world by scientific investigation. Eventually, I will leave the cave and fly to infinite.

I would like to thank all people who helped me during my PhD studies for their advices, guidance, and supports. I owe sincere thankfulness to my supervisor, Professor Mustapha C. E. Yagoub, the comprehensive and proposal examiners, and express my special feelings of gratitude to Professor Liam Peyton. I also gratefully acknowledge the financial assistance of the University of Ottawa in supporting during my PhD journey and I would not achieve my goals and dreams without this support.

Also, I was fortunate in building a successful academic network and having a fruitful collaboration with Professor Cesar A.D. Teixeira in the field of seizure prediction. I will never forget the support of my best friend, Ali Etemad, as well the help I received from Dr. Tahar Haddad and Dr. Ali Yousefi during my PhD journey.

Last but not least, I owe to express my appreciative feeling to my loving wife, “Leila”, for her sacrifice to take this journey with me and withstand all those intolerable and tough days. I am also thankful for my sweet son, “Danial”, for all the happiness, joy, and power that I received from him.

In addition, I appreciate my dear parents for all their supports and motivations. They may be far away, but they are always in my thoughts.

Contents

Chapter 1	Introduction.....	1
1.1	Motivations	1
1.2	Methodology	7
1.3	Contributions	8
1.4	Outlines.....	9
1.5	Publications.....	9
Chapter 2	Related concepts and State of the Art	11
2.1	Background.....	11
2.1.1	The central nervous system.....	11
2.1.2	The brain and its parts.....	12
2.1.3	Neurotransmission	13
2.2	EEG recordings.....	15
2.2.1	Scalp EEG.....	15
2.2.2	Intracranial EEG	16
2.3	Epilepsy	19
2.4	The Brain, Heart interactions in epilepsy	21
2.4.1	The autonomic nervous system.....	21
2.4.2	Influences of ANS on Cardiac Electrophysiology.....	24
2.4.3	Heart/Brain interaction in seizure period.....	26
2.5	Prediction of the seizure by analyzing electroencephalogram.....	27
2.6	Prediction of the seizure based on ECG signal	31
2.7	Prediction of the seizure based on EEG and ECG signals.....	31
2.8	Discussions	32
2.9	Conclusion	33
Chapter 3	Data Pre-Processing and Feature Engineering	35
3.1	Freiburg seizure prediction EEG database	36
3.2	Artifacts and Noise Removal	37
3.3	Brain wave sub bands	40
3.4	Preprocessing of ECG signal	47
3.5	Feature engineering.....	47

3.5.1 Introduction.....	47
3.5.2 Feature Extraction.....	48
3.5.3 Time domain features	49
3.5 Conclusion.....	54
Chapter 4 Preliminary results for prediction and classification	55
4.1 Classification	55
4.1.1 Support vector machine (SVM).....	56
4.1.2 Constraints with overfitting	57
4.1.3 Double cross-validation	58
4.1.4 Performance analysis	59
4.1.5 Results for SVM	61
4.2.1 Other Classifiers	70
4.2.2 Neural Networks (Multi-Layer Perceptron: MLP)	70
4.2.3 Random Forest	71
4.2.4 XGBOOST	72
4.2.5 Conclusion	73
Chapter 5 Comparative study of feature selection techniques	74
5.1 Methodology	74
5.2 Feature Extraction	75
5.3 Feature selection	76
5.4 Evaluation and Performance analysis	74
5.5 Dividing signals into frequency sub-bands	80
5.6 Feature selection methods comparison	87
5.7 Discussion	89
5.8 Conclusion	93
Chapter 6 Probabilistic approach for efficient prediction of epileptic seizure	94
6.1 Introduction.....	94
6.2 SOP and SPH definitions in seizure forecasting.....	95
6.3 Accuracy, an inadequate metric to evaluate an imbalanced dataset	97
6.4 The optimum window	100
6.5 Discussion about the choice of the window size	102
6.6 Proposed method for probabilistic approach	105
6.6.1 Preprocessing and feature extraction	106
6.6.2 Probabilistic framework.....	108

6.7 Results and discussions.....	110
6.8 Conclusion	116
Chapter 7 Conclusions and Future Work.....	118
7.1 Conclusions.....	118
7.2 Future work.....	119
References.....	127

List of Figures

Figure 1-1: The evolution of seizure prediction.....	18
Figure 1-2: Signal characteristics between EEG and iEEG signals.....	20
Figure 2-1: Brain cell-interconnection.....	12
Figure 2-2: Illustration of the different lobes of the brain.....	13
Figure 2-3: Illustration of the general form of an action potential in a neuron.....	15
Figure 2-4: An illustration of EEG recording in 10-20 system from a left sagittal.....	16
Figure 2-5: Different size and arrangement of subdural grid and strip electrodes.....	18
Figure 2-6: An intraoperative photography implanted.....	18
Figure 2-7: Non-invasive EEG compared to invasive EEG.....	19
Figure 2-8: Action potential in normal and during seizure activity.....	20
Figure 2-9: An illustration of CNS (in red and purple color) and PNS.....	22
Figure 2-10: A picture of the autonomic nervous system and its connections.....	23
Figure 2-11: The location of the hypothalamus near the brainstem.....	24
Figure 2-12: The innervation of the heart by parasympathetic and sympathetic systems.....	25
Figure 2-13: major brain arteries supplying the brain.....	26
Figure 3-1: An overview of an EEG signal containing seizure for patient #4.....	36
Figure 3-2: An overview of an EEG signal containing seizure for FLE and TLE.....	34
Figure 3-3: The position of the strip electrode for patient #1.....	39
Figure 3-4: The position of the grid electrode for patient #1.....	39
Figure 3-5: The position of the depth electrode for patient #2.....	40
Figure 3-6: A screenshot of the information about a patient from the new database.....	41
Figure 3-7: An illustration of an EEG signal in time and frequency response.....	43
Figure 3-8: Typical rhythmic activities of an EEG signal	45
Figure 3-9: The time and frequency domain representation of the not clean ECG	46
Figure 3-10: The time and frequency domain representation of the clean ECG	46
Figure 4-1: An illustration of a simple binary classification problem with two informative features	56
Figure 4-2: Hyperplane separating the two classes	57
Figure 4-3: Hard Margin Classification	57

Figure 4-4: Soft Margin Classification makes a classifier more robust	58
Figure 4-5: An illustration of the double cross-validation on the data set	60
Figure 4-6: Flowchart of the proposed algorithm for seizure prediction	62
Figure 4-7: Non-overlapping 4 seconds window for extracting the feature	63
Figure 4-8: A 4 seconds window with 50% overlapping for extracting the feature	63
Figure 4-9: Probability of seizure happening for patient #16 with a window of 20 seconds	64
Figure 4-10: Probability of seizure happening for patient #16 with a window of 10 seconds	66
Figure 4-11: Probability of seizure happening for patient #16 with a window of 4 seconds	66
Figure 4-12: Probability of seizure happening for patient #16 with a window of 2 seconds	67
Figure 4-13: Probability of seizure happening for patient #16 with a window of 1 second	68
Figure 4-14: Probability of seizure happening for patient #16 overlapping 2 seconds	68
Figure 4-15: Multi-layer perceptron with one hidden layer with a Relu activation function	71
Figure 4-16: Illustration of different activation functions.....	71
Figure 5-1: An overview of the proposed work in analyzing the discriminative features	74
Figure 5-2: Investigation of dividing the EEG signal into various sub-bands for both lobes.....	81
Figure 5-3: An overview of the features ranked by Kruskal-Wallis as a winner method for both lobes.....	86
Figure 5-4: Heat-map of the correlation matrix between the top 30 subset of features for both lobes.....	87
Figure 5-5: The result of mRMR feature ranking for temporal lobe.....	89
Figure 6-1: Durations of seizure occurrence period (SOP) and prediction horizon (SPH) are demonstrated.....	95
Figure 6-2. Flowchart of the work to select the best length of the window.....	97
Figure 6-3: MCC for all the classifiers in a graph.....	100
Figure 6-4: MCC of XGB for various window size	101
Figure 6-5: Precision and recall (PR) curve for 2 s window	101
Figure 6-6: The Receiver Operating Characteristic (ROC) curve for 2 s window	102
Figure 6-7: Flowchart of the proposed work	105
Figure 6-8: The histogram of Platt scaling for the whole test set (PIN).....	107
Figure 6-9: P_{IN} related to preictal and proceeding of the interictal portions for Patient #10.....	112
Figure 6-10: P_{IN} related to preictal and proceeding the interictal portions for Patient #12.....	113
Figure 6-11: Comparison of sensitivity, false positive rate and the prediction time.....	116
Figure 7-1: Global coherence for various frequencies of a frontal lobe epilepsy patient.....	120

Figure 7-2: Global coherence for various frequencies of a temporal lobe epilepsy patient.....	121
Figure 7-3: The interaction of EEG and ECG.....	123
Figure 7-4: The cortical electrode array and the Lateral implantation of a WIMAGINE implant.....	124
Figure 7-5: An illustration of an autonomous medical device with a closed loop seizure suppression system.....	125

List of Tables

Table 3-1: The information of patients in the dataset.....	37
Table 3-2: Different kinds of artifacts and their sources.....	42
Table 3-3: The parameters of the filters.....	44
Table 3-4: Categorizing various features employed in seizure prediction in different domain.....	47
Table 4-1: Detection Results for Different window size.....	64
Table 4-2: Prediction Results for Different window size.....	64
Table 4-3: The results for 2s-window with different overlapping window.....	61
Table 4-4: The results for 2 seconds window with different overlapping.....	69
Table 4-5: The overall results for 2 seconds window with 50% overlapping	69
Table 5-1: Features extracted during a 2 seconds sliding window.....	76
Table 5-2: The confusion matrix for a binary classification task	79
Table 5-3: The comparison of computational cost of three feature selection methods applied on the train set.....	82
Table 5-4: The performance of the various feature selection methods for both lobes applied on the test set.....	82
Table 5-5: Top 30 feature-subset ranked by three types of approaches for temporal lobe epilepsy.....	84
Table 5-6: Top 30 feature-subset ranked by three types of approaches for frontal lobe epilepsy.....	85
Table 5-7: Features with the strongest linear relationship among top 30 features with the highest KW scores.....	88
Table 6-1: The confusion matrix for a classifier.....	96
Table 6-2: Various measures for multiple classifiers in an imbalanced dataset.....	96
Table 6-3: The performance of various classifiers for different window sizes.....	99
Table 6-4: The comparison of our study with existing works.....	104
Table 6-5: The implemented thresholds after numerous experiments for various patients.....	110
Table 6-6: The result of two kinds of prediction time for 6 patients.....	113
Table 6-7: The overall Forecasting results using for 6 patients	114
Table 6-8: Our forecasting results using for 6 patients compared to the published works.....	115

List of symbols

Acc	Accuracy
β	AR coefficient
C	Regularization parameter
c	Number of output group
cov	Covariance
Cv	Coefficient of Variation
D	Dimensional of a feature vector
E	Energy
En	Entropy
G	Kruskal-Wallis ranking value
hi	Hidden layer in NN
Kurt	Kurtosis
N	Total number of observations
n	Number of samples
ni	Number of observations in group i
Wi	Weights in NN
S^2	Variance
$P(x)$	Probability mass function
P_{IN}	Platt scaling based on the XGB model
P_{MAX}	Maximum probability of P_{IN}
P_{MEAN}	Average of the probability during a period
PNS	Probability of the non-seizure
P_{OUT}	Output of the probabilistic framework
PA1	First prediction alarm
PA2	Second prediction alarm
PSOP	Probability of the seizure occurrence period
R^d	Input pattern space
Ri	Mean rank of the group i
τ	Threshold
τ_N	Threshold based on non-seizure period
τ_S	Threshold based on seizure period
T_1	Time of prediction for PA1

T_2	Time of prediction for PA2
W_i	Weight parameters
$\bar{x}$	Mean of a signal
X	Input data point for a classifier
$x(t)$	Signal
ε_t	Error term
ε	Non-negative slack variable
μ_{Pns}	Mean of the probability of the non-seizure
Skew	Skewness
S	Standard deviation
y	Target of a classifier

List of Acronyms

AdaBoost	Adaptive Boosting classifier
AECR	Average Energy Concentration Ratio
AIS	Artificial Immunity System
ALP	Adaptive Linear discriminant- analysis-based Prediction
ALV	Amplitude Lock Value
ANBP	Adaptive Naive Bayes-based Prediction
ANS	Autonomic Nervous System
ANOVA	Analysis of variance
APP	Adaptive Probabilistic Prediction
AR	Autoregressive
AUC	Area under curve
BCI	Brain Computer Interface
CFC	Cross-Frequency Coupling
CNS	Central nervous system
COV	Coefficient of Variation
CPU	Central Processing Unit
DPSS	Discrete Prolate Spheroidal Sequence
DT	Decision Tree
ECG	Electrocardiogram
ECoG	Electrocorticography
EEG	Electroencephalogram
EMD	Empirical Mode Decomposition
EMG	Electromyogram
EOG	Electrooculogram
EP	Electrophysiology
FLE	Frontal Lobe Epilepsy
FN	False Negative
FP	False positive
FP/h	False Positives per hour
FPR	False Positive Rate
GPDC	Generalized Partial Directed Coherence
HF	High-frequency

HDD	Hard disk drive
HR	Heart Rate
Ictal	Period of the seizure
iEEG	Intracranial electroencephalography
IQR	Interquartile Range
Interictal	Period free seizure
Kaggle	World's largest data science community
k-NN	k-Nearest Neighbors
KW	Kruskal Wallis
LDA	Linear Discriminant Analysis
LF	Low-frequency
LR	Logistic Regression
MAD	Mean Absolute Deviation
MCC	Matthews correlation coefficient
MLP	Multi-Layer Perceptron
MVAR	Multivariate autoregressive
NC	Not Clear
NB	Naive Bayes classifier
NN	Neural Networks
NN50	Number of pairs of NN intervals that is longer than 50 ms
NS	Non-seizure period
NumPy	Python library for working with arrays
PCA	Principal Component Analysis
PLS	Partial Least Squares
PLV	Phase lock value
PNS	Peripheral nervous system
PR	Precision-recall
Preictal	Period before the seizure
postictal	Period after the seizure
PP	Preictal Period
RAM	Random Access Memory
RBFNN	RBF neural network
RF	Random Forest classifier
RKHS	Reproducing Kernel Hilbert Space

ROC	Receiver Operating Characteristics
RMS	Root Mean Square
RMSSD	Root Mean Square of the Successive Differences
RR	Time elapsed between two successive R-waves of the QRS complex
RRI	RR Interval
Rij	Rijswijk rat
scikit-learn	Python module for machine learning
SD1	Standard deviation of an instant beat-to-beat variability of an ECG signal
SD2	Standard deviation of a continuous long- term RRI
SDNN	Standard deviation of the NN (R-R) intervals
SOP	Seizure occurrence period
SVM	Support Vector Machine
SVM-RFE	Support Vector Machine - Recursive Feature Elimination
TLE	Temporal Lobe Epilepsy
TN	True Negative
TP	True Positive
TPR	True Positive Rate
WAG	Wistar-Albino-Glaxo rat
XGBoost	eXtreme Gradient Boost
ZCR	Zero Cross Rating

Chapter 1

Introduction

1.1 Motivations

As a chronic neurological disorder, epilepsy is the third most common neurological disease; it affects about 1% of the world population [1], [2]. Patients with epilepsy suffer from recurrent seizures that affect their quality of life with loss of consciousness, depression, mental illness, strange sensations, and convulsions. If seizures are not controlled, these patients will have to deal with serious limitations in terms of family, social, educational, and vocational activities [1].

Epileptic seizures happen when there is a sudden excessive electrical discharge in the entire brain. During the epileptic seizure, the normal activity of the brain will be interrupted, which alters an individual's behaviour and functioning [3]. Patients with epilepsy sometimes are treated with medication or surgery, but these approaches are not completely effective and poor response to medication remains a serious limitation in the treatment of epileptic seizures [4]. Unfortunately, the seizures that cannot be completely cured have a negative influence on the patient's lifetime. In this scenario, a patient cannot autonomously work and do some activities [5], [6], since he/she is susceptible to face severe issues such as injuries and sudden death, limited independence, driving restrictions and difficulties in finding and keeping a job.

A seizure forecasting tool can therefore significantly enrich the patient's life. Prediction of an epileptic seizure at an early stage is vital for such people [1], [7] and, consequently, the objective of an ever-increasing numbers of research teams, as demonstrated by the number of papers published in this field (Figure 1.1 [8]).

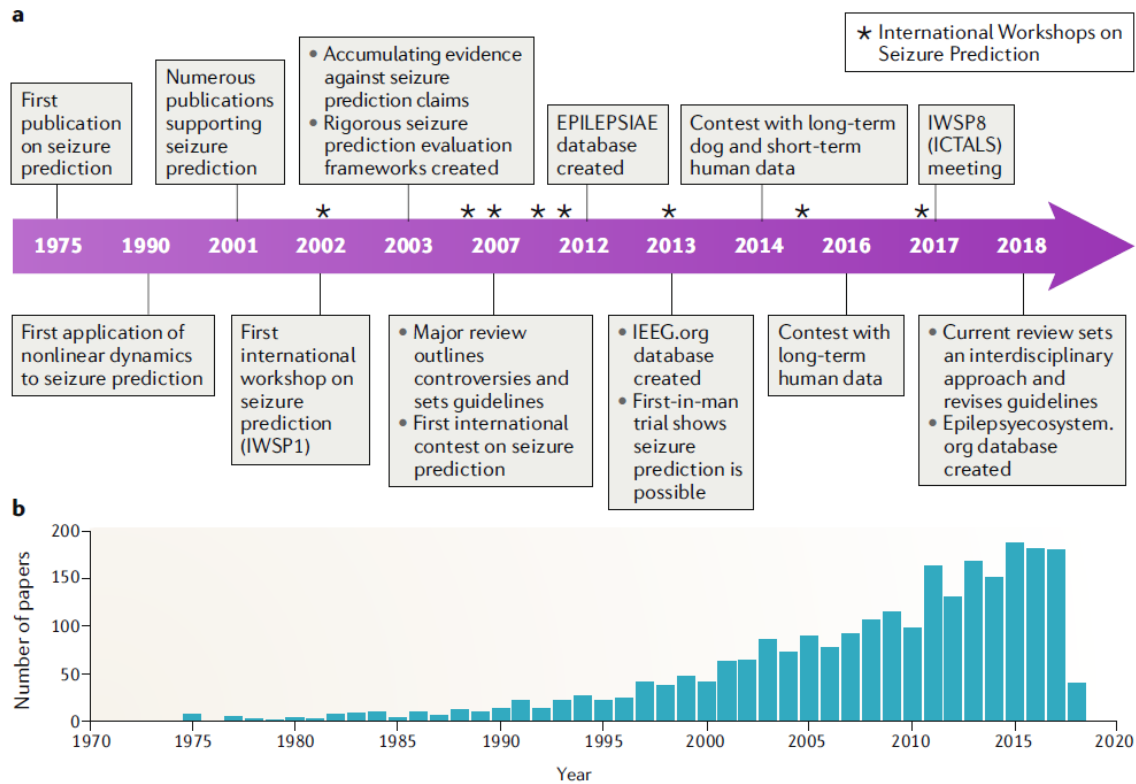


Figure 1-1 The evolution of research in seizure prediction [8].

To this aim, non-invasive/scalp electroencephalogram (EEG) and intracranial electroencephalography (iEEG) tests are seen as the major sources of data that can help efficient diagnostic of different neurological disorders namely, epilepsy, and to study the function and behaviour of the brain [9]–[13].

Since iEEG is an invasive method that records by implanting the electrodes in the brain during surgery, there are generally many fewer artifacts than in scalp recordings (EEG), which can be concluded from the higher signal-to-noise ratio (20 to above 100 times) of iEEG compared to non-invasive EEG [8], [11]–[15]. Therefore, iEEG is more valuable for epilepsy research. Figure 1-2 illustrates the comparison between the non-invasive EEG compared to the invasive one.

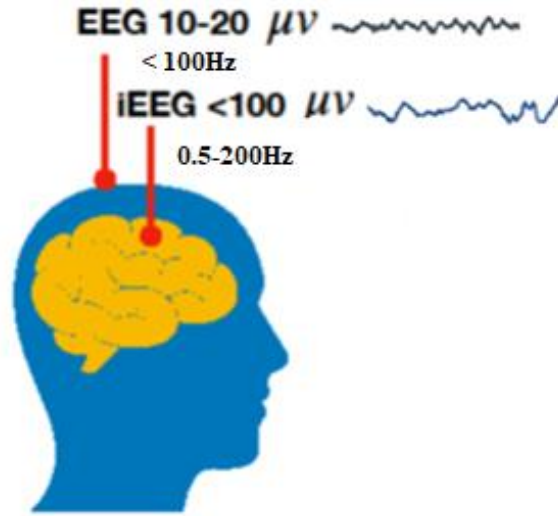


Figure 1-2 Signal characteristics between EEG and iEEG signals.

The International League Against Epilepsy (ILAE) divided epileptic seizures into partial or focal and generalized seizures. Focal seizures originate in a limited region of the brain and may spread to other regions. On the other hand, generalized seizures are initiated in bilateral hemispheric areas and quickly propagate to all cortical areas [16]. Though there are different forms of seizures, we focused on those that are focal, mainly in the temporal and frontal lobes, entitled to Temporal Lobe Epilepsy (TLE) and Frontal Lobe Epilepsy (FLE), respectively.

TLE is one of the most prevalent form of focal epilepsy. It has received a significant amount of attention from the neurologists due to its high likelihood of clinical occurrence [17]–[20]. This kind of epilepsy may be treated medically at the onset of the disease with different antiepileptic drugs [20]. However, investigation on predicting of temporal lobe epilepsy is still open since in the last decade, various approaches have been employed to attain to increase the sensitivity and decrease the low False-Positive rate [7], [21]–[24].

FLE is the second-most common form of focal epilepsy after TLE accounting for 25% of epilepsy [25]–[28]. Instead, FLEs, as compared to TLEs, tend to be brief, drug resistant, more problematic, and to occur during sleep. Furthermore, the surgery for FLE has poorer outcomes than for TLE; as a result, the surgical workup of FLE is even more demanding [25], [29], [30]. Diagnosis of the

FLE is rather hard due to having similar symptoms as sleep disorder, night terror, and psychiatric diseases [30].

Consequently, designing and implementing a reliable forecasting and early warning tool that can help epileptic individuals to take appropriate drugs during an early warning period is also vital [31]–[33]. Therefore, the challenge is to reach a higher sensitivity with lower false positive rate and longer anticipation time. To do so, the characteristics of EEG signals need to be investigated to select the most appropriate features that reliably reflect the variations in the signals, the proper bands of frequencies to consider, the window size to apply, and the type of classifiers to use in order to better predict seizures.

Efficient feature selection is indeed the key in prediction. Some studies already touched this aim but in a relative limitative way, like in [34], in which the authors just used a filter method to select features or in [35], where the authors based their approach on accuracy, which, due to imbalanced data, may be seen as not so accurate. In fact, even if their model can predict the majority class, normal cases, it might be not so adequate for the minority class, seizure cases. Also, in their work, EEG signal was filtered by a band-pass filter 0.5-70 Hz, so an important part of the seizure information was removed (frequencies higher than 70 Hz). In fact, one of the limitations of existing works is filtering the EEG signal with a pass-band filter, which removes the high-frequency sub-bands that are very important in the prediction of the seizure [35], [36]. It is then important to use a wide range of frequencies (up to 120 Hz) and consider the whole available data for the nominated patients. Another notable limitation of published works in this field is about employing limited data in their studies. In [37], the authors attempted to study just one minute of data or in [38], they used a limited amount of data: 5 min preictal and 10 min interictal.

As for which kind of features to consider to best predict seizures, several works have been published in this field. In [39], [40] the authors compared various feature ranking methods but here also, only accuracy was considered as solely measure. In [38], [41] the authors implemented few features and applied just a classifier so that the results may not be generalizable to a larger set of data. Also, in [13], [34], [42] the authors tried to combine and investigate the frontal lobe and temporal lobe epilepsy cases and provide a general result for all of the patients. However, there

are some neuropsychological differences between FLE and TLE and also the spreading of the TLE is different from that of the FLE [43], [44].

Currently, there are three main approaches in analyzing EEG signal for prediction of seizures. In one group, researchers divided the EEG signal into various sub-bands and extracted the features for each correspondent sub-band [45], [46] while in a second group, researchers applied the feature extraction on just the EEG electrodes [37], [47]–[49]. In the last group, people intended to combine those two early mentioned groups in their works [34], [41]. It will be then beneficial to examine the effect of considering sub-bands in processing of such complicated, non-linear, and uncertain signals.

Among the several feature extraction techniques that have been introduced in the last few years, the time-domain ones are the earliest recommended methods. Time-domain features are employed to achieve discriminative information at a low computational cost. The obtained features are then fed to feature selection methods [50]. High-quality features can be defined as those that produce maximum class separability, robustness, and less computational complexity, key parameters to consider when targeting their use in everyday devices such as wearable devices. It is therefore crucial to explore the impact of iEEG signal variations on a reduced set of commonly used features and compare their performance. However, there is little evidence on the effectiveness of various feature selection methods on iEEG data of epilepsy patients. So, it is important to explore their differences and find which ones may perform better in the perspective to deploy an algorithm in an implantable medical device that uses linear features, allowing rapid calculation with less complexity in prediction of the seizure.

In addition, any selection implies setting guidelines in terms of threshold. In fact, decision making is one of the unmissable steps in prediction of the seizure [10], [11], [15]. In [2], [38] the authors did not define properly the threshold rules, while in [13], [51] just the preictal portion of the signal has been used in generating the threshold. Therefore, setting a reliable threshold approach is an important step to address while extracting the signal features.

Seizure prediction is also strongly dependent on the window size used in processing data. Different works have been published in selecting the most suitable window size. However, some of them

used accuracy as the plain evaluation metric, while in other studies, the performance measure of the feature classifier was not reported [15], [38], [52], [53]. Also, in some published works, the authors just employed one feature to optimize the classifier and find the best window size [38], [54] while redundant information may be extracted from the electrodes during a given period of time, which may be useless until be combined with other discriminative features. Therefore, selecting the best window size can highly improve the prediction performance.

1.2 Methodology

This work intends to design and implement a reliable forecasting and early warning tool that can help epileptic individuals to take appropriate drugs during an early warning period [2], [55], [56]. Furthermore, since portable devices are so present in our daily life, targeting tools that could be easily implemented in such devices is the objective of this study. However, knowing that:

- for new customers, the application will have to be frequently updated during its first uses to be able to efficiently integrate the new patient data,
- some patients may not have regular access to wireless connections and/or computers/tablets,

we opted for a strategy of dual-mode operation: the training/update should be performed on the portable device itself while the application is still working on prediction mode. So, in order to ensure an efficient online training/prediction, it is crucial to shorten the training CPU time while making the tool operation as simple as possible.

As starting point, we will try to avoid the limitations mentioned earlier in existing works by taking into account a large range of frequencies up to 120 Hz in processing of EEG signals and considering various classifiers with different performance metrics. In fact, accuracy cannot be good enough when dealing with imbalanced datasets in which the vast majority of a class comes from one class. Thus, we will utilize various classifiers and evaluation metrics other than accuracy to make sure our results are not just based on one measurement and a single classifier.

Also, in order to avoid any confusion whether the division of EEG signals to various sub-bands is beneficial in processing or not, we will run two scenarios: on one side, we will divide the EEG signal into 6 sub-bands and extract the energy accumulated for each sub-band of EEG electrode while on another side, the energy will be extracted from an electrode without diving it to various sub-bands. Finally both scenarios will be compared together with a distribution free test, Kruskal Wallis, to show which approach extracts more meaningful and informative biomarkers from an EEG signal.

Furthermore, we will investigate the importance of 16 time-domain features with three types of feature ranking methods (filter, wrapper, and embedded approach); the aim being to reduce the dimensionality from 630 to 30 by removing the irrelevant features and then, reducing the computational time and cost [57]. Since the morphology of the signal for TLE and FLE is different, they will be studied and investigated separately in this work instead of combining the patients with various origin of epilepsy.

After determining the ability of a number of measures to discriminate between the preictal and interictal period utilizing a non-parametric statistical technique, which does not rely on any distribution and require any tuning parameters, we will look for the best window size, evaluating the measure for performance of a classifier, and developing a framework for efficient epileptic seizures prediction with introducing a novel threshold (by considering the information from both interictal and ictal stages). This can lead us to higher sensitivity, less false positive rate, and longer anticipation time. All in all, we will investigate the main parameters that can affect the prediction of the seizure such the choice of the features, window size, and the classifier.

Note that because of the limited number of patients available in the database we used, i.e., the Freiburg database [58], the obtained results should be seen much more as a proof of concept and promising step in epileptic seizures prediction rather than a broad validation of the approach we developed.

1.3 Contributions

The major original contributions of this thesis are as follows:

- To the best of our knowledge, this is the first study that combines the three known types of feature selection (filter, wrapper, and embedded) in a scenario of seizure prediction for both TLE and FLE (whereas evaluated separately).
- To the best of our knowledge, this is the first time the prediction of the frontal lobe epilepsy for human subjects was effectively investigated. Therefore, we enhanced that by dividing the patients based on the origin of their seizure, utilizing efficient performance metrics and applying a comprehensive study on the discriminative features in seizure prediction.
- We demonstrated in this work how dividing the EEG signal into various sub-bands is important in prediction of the seizure while utilizing other performance time-domain metrics besides accuracy can further enhance the precision of the results.
- We considered a large panel of time-domain features to represent the changes of a series of non-stationary signals and, more specifically, the ones that are most informative and more recommended to achieve discriminative information at a low computational cost.
- We found out that a 2-second window is the best choice knowing that, according to [59], by choosing a window size longer than 1s and smaller than 4s, a non-stationary signal like EEG can be assumed as stationary.
- We developed an efficient algorithm with a novel threshold method to predict the seizure with an invasive approach, so that patients can be warned sufficiently in advance with high sensitivity and very low zero false positive rate. The anticipation time can be up to about 75 minutes and varies from patient to patient, which is enough to provide adequate clinical treatment time prior to a seizure [56]. Interestingly, our novel prediction system does not rely just on an early warning since the medical decision making continues to inform the patient for the upcoming seizure.

1.4 Outlines

In the next Chapter, we will conduct a literature survey to support the fundamental concepts of the proposed approach in prediction of the seizure.

Chapter Three describes the dataset used in this work, the preprocessing stage, and the feature engineering while the preliminary results of SVM for prediction will be explained in chapter Four.

In chapter Five, we compare various feature selections methods based on the time domain extracted features.

In chapter Six, we will discuss the proposed prediction approach. Finally, we will conclude the work in the last Chapter, Seven.

1.5 Publications

- [1] B. Abbaszadeh, C.A.D. Teixeira, and M.C.E. Yagoub, "Towards an online seizure prediction system: A novel probabilistic approach for efficient prediction of epileptic seizure with iEEG signal," accepted in PLOS ONE.
- [2] B. Abbaszadeh, C. A. D. Teixeira, M. C.E. Yagoub, "Feature selection techniques for the analysis of discriminative features in temporal and frontal lobe epilepsy: A comparative study, " The Open Biomedical Engineering Journal (In press).
- [3] B. Abbaszadeh, R.S. Fard, and M.C.E. Yagoub, "Application of global coherence measure to characterize coordinated neural activity during frontal and temporal lobe epilepsy," In 2020 42nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), pp. 3699-3702, Montréal, QC, Canada, July 20-24, 2020.
- [4] B. Abbaszadeh, T. Haddad, and M.C.E. Yagoub, "Probabilistic prediction of epileptic seizures using SVM," IEEE Annual International Conference of the Engineering in Medicine and Biology Society, Berlin, Germany, July 23-27, 2019.

- [5] B. Abbaszadeh and M.C.E. Yagoub, "Optimum window size and overlapping for efficient seizure prediction," International Conference on Computational Intelligence in Bioinformatics and Computational Biology, Siena, Italy, July 9-11, 2019.
- [6] B. Abbaszadeh and M.C.E. Yagoub, "Feature extracting and ranking for an efficient epileptic seizure prediction based on iEEG signal," Ottawa Artificial Intelligence Alliance Annual Workshop (AI-Alliance 2019), Ottawa, Canada, Nov. 28, 2019.

Chapter 2

Related concepts and State of the Art

This chapter is devoted to presenting the background information and related works in the field of automated iEEG-based, EEG-based, ECG-based, and EEG with ECG-based epileptic seizure prediction.

2.1 Background

2.1.1 The central nervous system

The nervous system has two different components namely, central and peripheral nervous systems [11]. The peripheral nervous system (PNS) is the section of the nervous system that contains the nerves outside the brain and spinal cord.

The main role of the PNS is to connect the central nervous system (CNS) to the variety of organ systems. On the other hand, CNS is the part of the nervous system that includes the brain and spinal cord. This part of the nervous system analyses information received from the PNS, and harmonizes the activity of all parts of the body [12]. Based on its function and structure, the nerve cell (neuron) can be separated into three main parts (figure 2-1): (1) the cell body named the soma, (2) the various short processes of the soma, termed dendrites and (3) the axon, which is the single long nerve fiber [12].

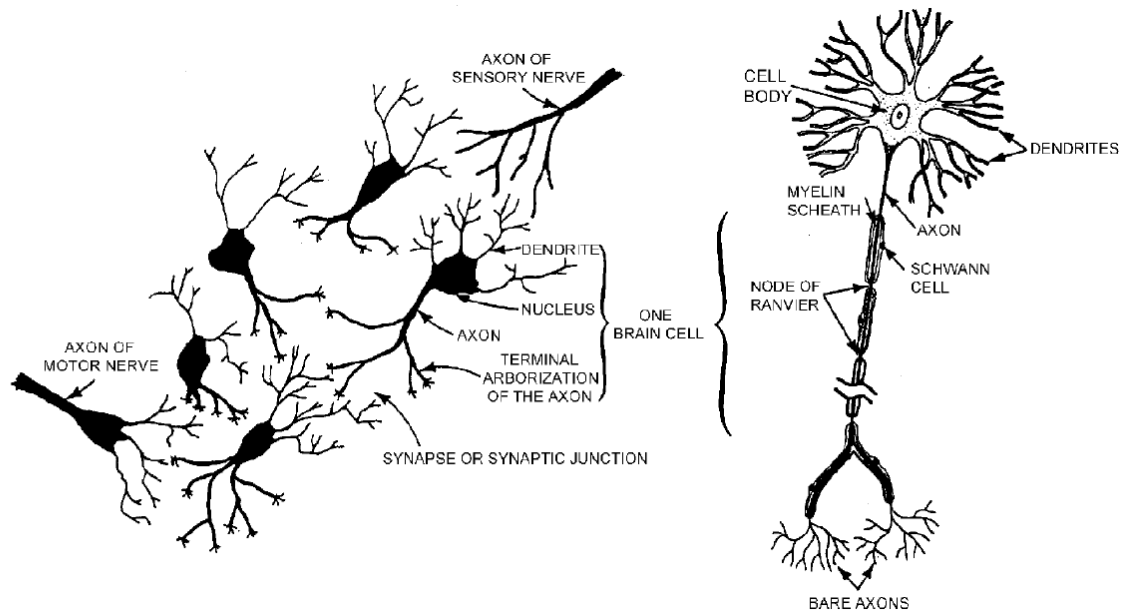


Figure 2-1: Brain cell-interconnection [12].

2.1.2 The brain and its parts

The brain has a complex structure that contains a huge number of cells and is divided into two hemispheres. These cells have many interconnections that enable them to receive data from various organs. Briefly, the brain can be seen as a supervisory control unit.

The brain is protected by several layers; from outside to the surface of the brain (cortex) they are scalp (skin), skull (bone), and meninges (a membrane that covers and protects the brain) [11], [12].

The outer layer of the brain is named cortex, which contains cell bodies of the neurons. The brain cortex has a significant role in memory, attention, perception, consciousness, thinking, language, and awareness.

Our brain cortex is divided up into four lobes (note that since each lobe has a right and left side, there are in fact 8 lobes) [13], [60], [61], which are (figure 2-2):

- The frontal lobe placed at the front of the brain and responsible for higher mental processes like thinking, decision making, and planning.
- The parietal lobe located above the occipital lobe and behind the frontal lobe.
- The temporal lobe located below the lateral fissure (one of the structures of the human brain that separates frontal lobe and parietal lobe and has a major role in processing sensory input, auditory perception, language and speech production, memory).
- The occipital lobe located in the back of the brain and acting as a visual processing center.

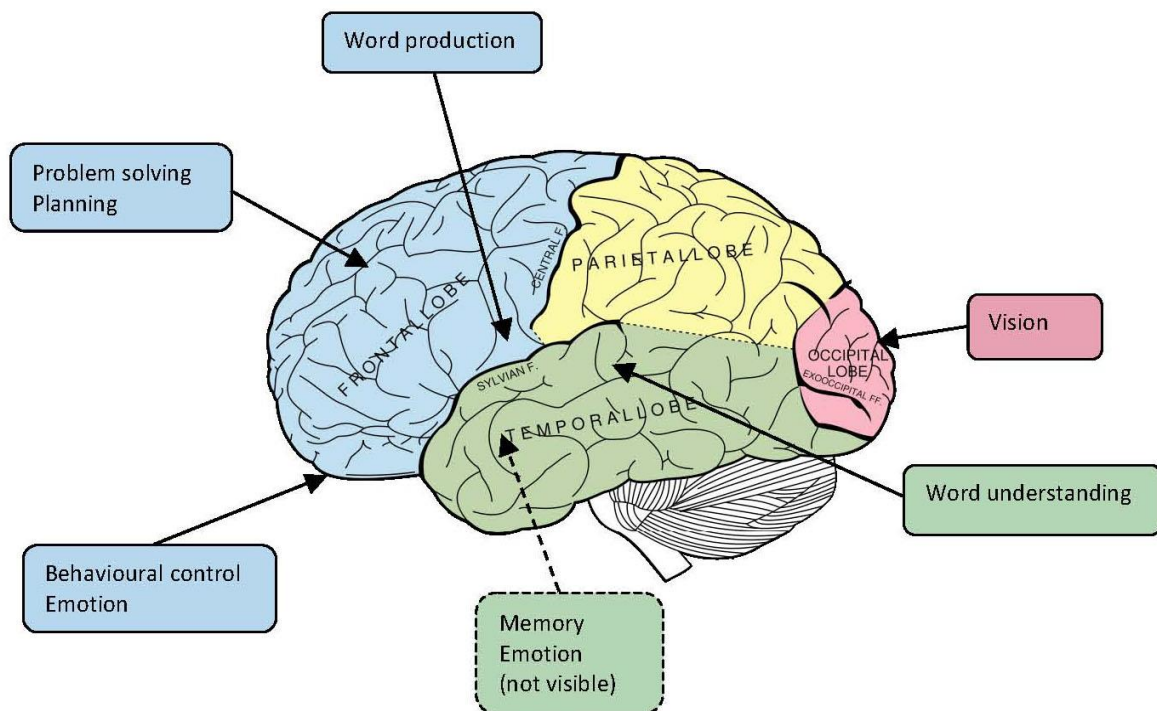


Figure 2-2: Illustration of the different lobes of the brain [61].

2.1.3. Neurotransmission

The nervous system has developed an exclusive ability in intracellular signalling (communicating inside of a cell) and intercellular signalling (communicating among the cells).

To send an electrical signal (action potential) very fast and over a long distance along the axon, neurons have developed special capability, called conduction. This mechanism is how a cell body of a neuron communicates with its own terminals through the axon. The interaction among the neurons is attained at synapses via the process of neurotransmission [62].

The conduction in the nerve starts when an action potential is produced around the cell body. The action potential is an electrical impulse and is quite similar to the electrical signals in electronic systems. An electrical impulse in a neuron happens owing to the movement of the ions along the neuronal membrane like the flow of electrons through a wire in electrical devices. Ions are electrically charged particles (positive or negative) and the neuronal membrane performs like a barrier to ions. The ions transfer from the membrane to ion channels that open and close owing to the existence of neurotransmitters. Once the relative concentration of the positive ions inside the neuron changes, the electrical potential across the membrane (the transmembrane potential) changes. This is because the action potential depends on Na^+ flowing in and then K^+ flowing in or out (it is the charges on the ions and the relative concentration of positive ions that matter). In general, the resting membrane potential of a neuron is around -70 millivolts (i.e., the membrane is to be polarized). The ions influx and outflux will turn the inside of the neuron more positive than the resting potential (means depolarization). Once the depolarization reaches a certain value of no return, named a threshold, a large amplitude is produced. In total, we say that an action potential was generated (Figure 2-3) [63], [64].

There are two types of intercellular transmission of information in the nervous system, electrically and chemically [65]:

Electrically: neurons are connected to each other. An action potential reaches the end of the axon through small holes (named gap junctions) in an individual cell membrane and the depolarization will continue through the membrane to the postsynaptic neuron.

Chemically: there is a gap (the synaptic cleft) between the axon terminal and the nearby neuron. When an action potential comes to the end of an axon, a chemical (neurotransmitter) is produced that stimulates the next neuron to alter its electric potential [66].

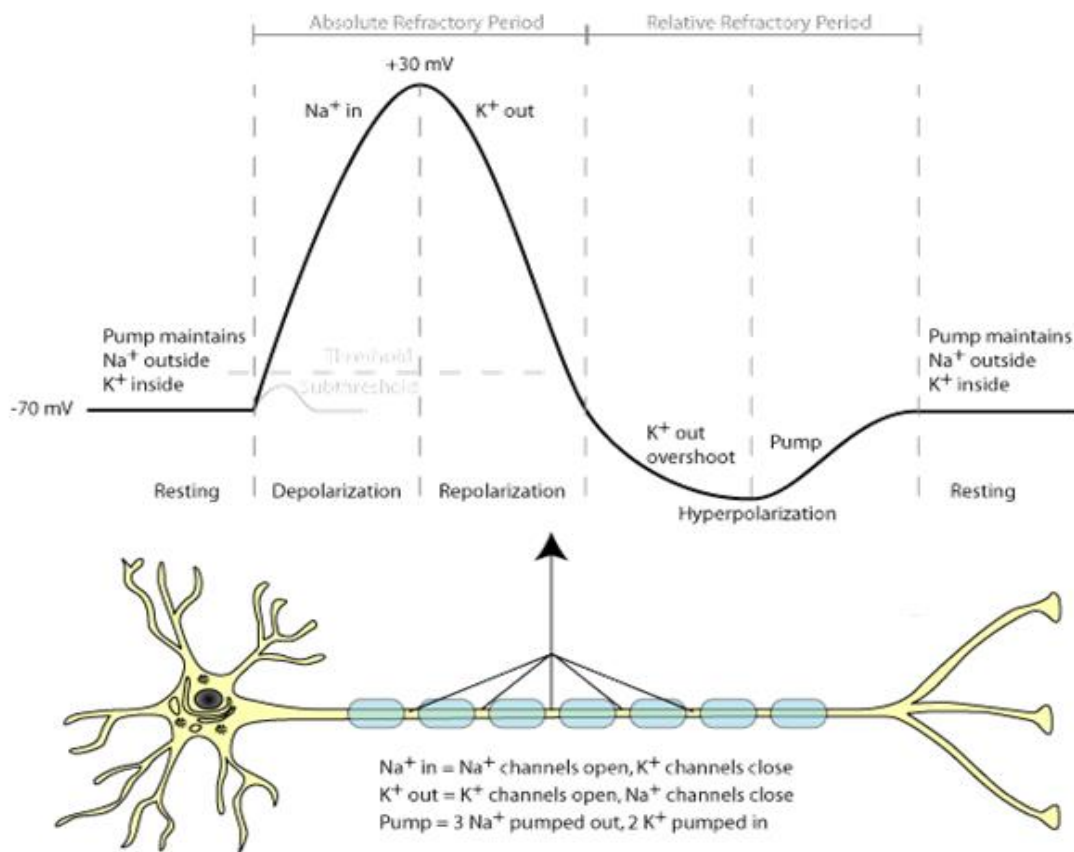


Figure 2-3: Illustration of the general form of an action potential in a neuron [64].

2.2 EEG recordings

EEG records can be made non-invasively from the scalp or invasively via surgical implantation of invasive electrodes in the intracranial structures.

2.2.1 Scalp EEG

The scalp EEG measures the difference of potentials between pairs of electrodes attached to the scalp and finds the smallest electrical charges that come from the activity of the neurons. The 10-20 system, an internationally recognized method, proposed by Jasper in 1958 [67], has been a long-

established convention for electrode placement and labelling in clinical settings. The system states the position of 21 EEG electrodes related to specific anatomical landmarks such that they are separated by 10% or 20% of the space between the nasion and the inion and between the preauricular points (figure 2-4). In order to have a higher-resolution EEG, higher density systems such as 128, 256, and 345 electrodes have been introduced and used by different centres [68].

Scalp EEG recordings are limited by the extracranial artifacts caused by the scalp muscle and heart activities, eye movement, external electromagnetic field, etc. Therefore, implementing an automatic removal of the artifacts from EEG signal stays a key challenge for finding valuable information from brain activities.

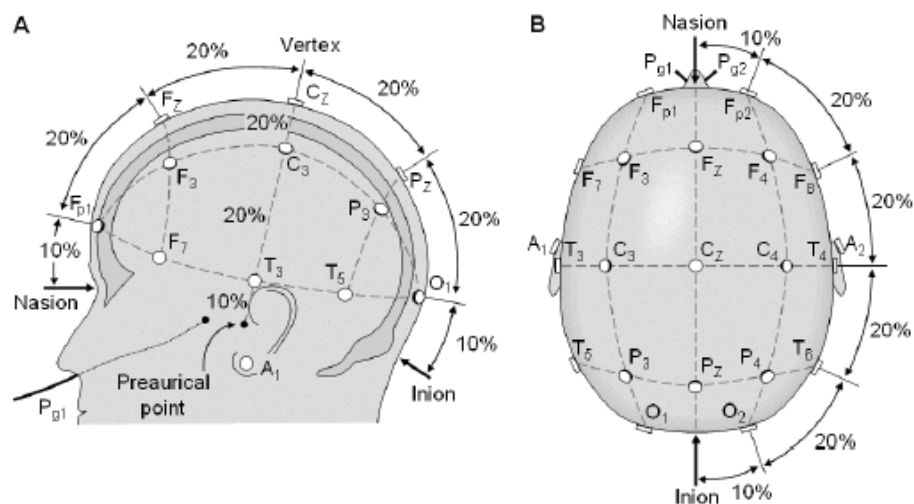


Figure 2-4: An illustration of EEG recording in 10-20 system from a left sagittal (A) and top (B) view of the head.

Nomenclature: A= Ear lobe, C= central, P= parietal, F= frontal, Fp= frontal polar, O= occipital. Odd numbers are employed on the left hemisphere, even numbers on the right, and Z (zero), in the midline [69].

2.2.2 Intracranial EEG

Another type of EEG called Electrocorticography (ECoGs) or intracranial EEG (iEEGs) is typically recorded with the help of subdural depth electrodes and/or grid electrodes. A subdural electrode is a small metal disc implanted in Teflon or silastic sheath material and categorized in

various arrangements (Figure 2-5). These electrodes, connected to a thin metal wire, are individually insulated and ultimately bundled and covered with plastic material (tail). When the configuration of electrodes is in a single column, it commonly alludes to a subdural strip electrode. When it includes rows and columns, it is known as a subdural grid electrode and is available in many different sizes and configurations. These grids or strips are embedded subdurally on the surface of the brain, and their tails exit the meninges, skull and eventually over the scalp to protrude on the scalp (Figure 2-6).

In order to attach the electrodes on the cortex of the brain to record electrical activity, a neurosurgeon should first do a craniotomy (remove part of the skull to access to the brain). This process may be done either under general or local anesthesia [69], [70].

Compared to the scalp EEG, electrode placement is not standardized in the intracranial EEG. Once the intracranial EEG is employed to record the brain activity, brain wave is not weakened or changed by the skull/scalp tissue which perform like a low-pass filter. Moreover, the brain wave is not affected by Electromyography (EMG) artifacts, regularly observed in the scalp EEG. Furthermore, the invasive electrode is able to record signals from the small-population neurons, which is non-recordable with the scalp one. Then, the seizures can be identified typically earlier employing the intracranial electrodes compared to the scalp electrodes (Figure 2-7) [70]–[73].

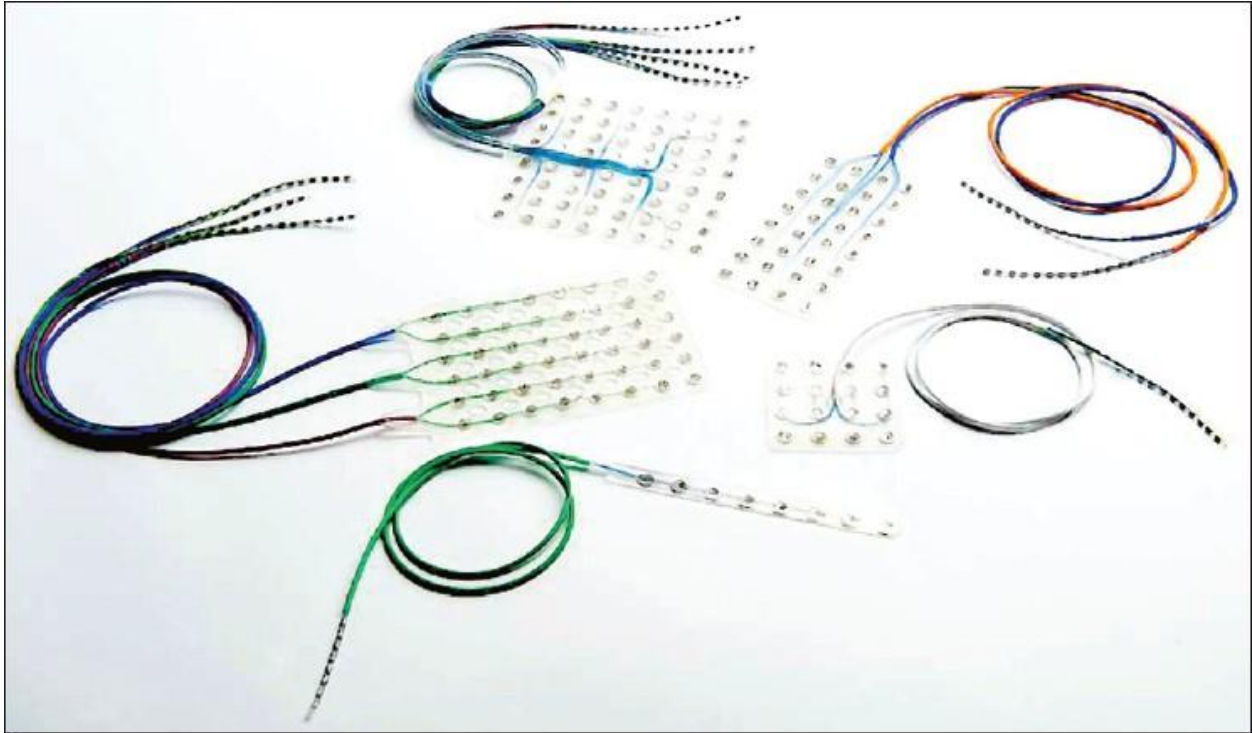


Fig 2-5: Different size and arrangement of subdural grid and strip electrodes fabricated by a manufacturer (PMT Corporation, Chanhassen, MN, USA) [72].

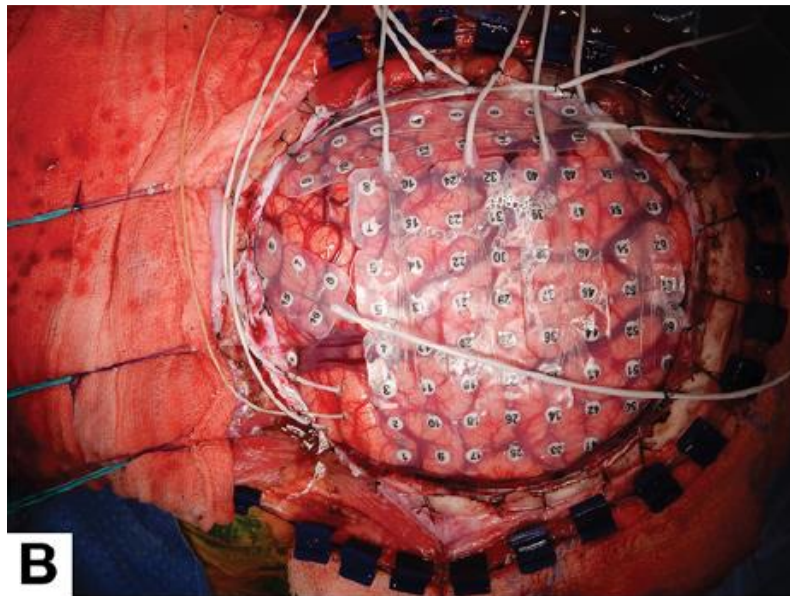


Figure 2-6. An intraoperative photograph implanted on the left cerebral hemisphere of a patient covered by various subdural grid electrode arrays [72].

2.3 Epilepsy

For many years, people thought epilepsy was because of the attack of demons and spirits [74]. Today, epilepsy is known as the third most prevalent neurological condition characterized by recurrent seizures, a sudden uncontrolled electrical discharge in the brain, and people with frequent, and unprovoked seizures are diagnosed with epilepsy.

Neurons firing seems to be random at first. Normally, brain activity is non-synchronous. After a neuron fires, it becomes more resistant to produce new spikes, as seen in Figure 2-8. These mechanisms are broken during epilepsy seizures [75]. As a result of these abnormalities, a collection of nerve cells starts firing excessively and synchronously. Actually, based on the offered definition by the International Bureau for Epilepsy and the International League Against Epilepsy (ILAE), the epilepsy seizure is “a temporary incident of the abnormal and simultaneous neuronal activity in the brain” [16], [76], [77].

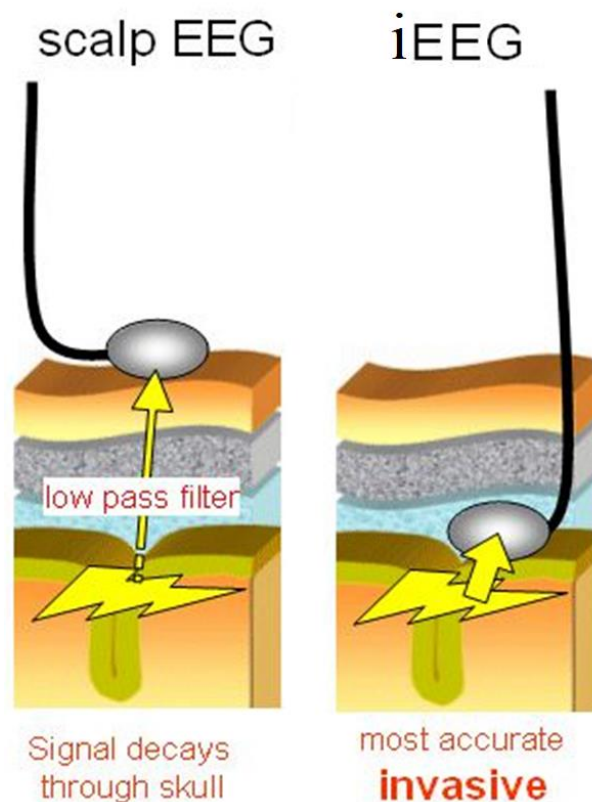


Figure 2-7. Non-invasive EEG compared to invasive EEG [78].

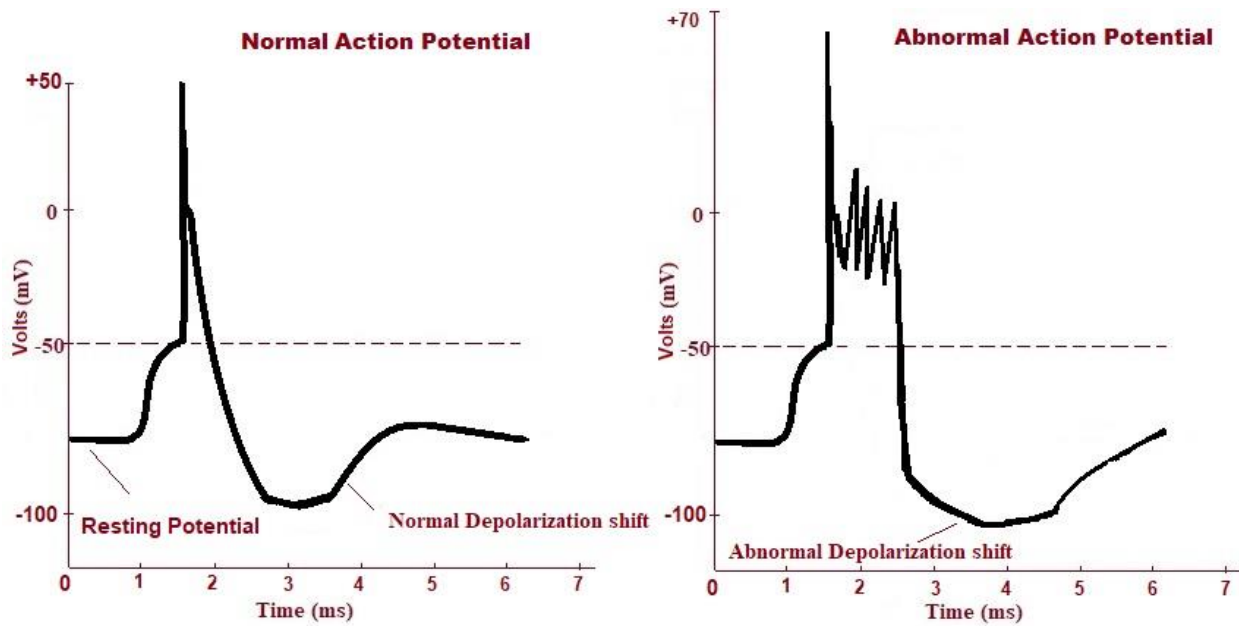


Figure 2-8. Action potential in normal and during seizure activity.

Genetic features increase the risk of developing the seizure, the abnormal electric discharge in the brain [74]. Other causes of seizure are brain tumours, head injuries, infectious illnesses, fever, lead poisoning, and imperfect development of the brain. Even for some patients with seizures, the cause of their seizure is still unknown [79], [80].

The International League against Epilepsy (ILAE) divides epileptic seizures into partial or focal and generalized seizures. Focal seizures originate in a limited region of the brain and may spread to other regions. Depending on whether or not consciousness is impaired during the attack, a partial seizure is categorized as simple or complex [16].

On the other hand, generalized seizures are initiated in bilateral hemispheric areas which appear to be simultaneously involved. Consciousness may be altered and motor behaviours are bilateral. Generalized seizures are often separated into convulsive and non-convulsive types.

Epilepsies based on their origin can be classified into four major classes: temporal, frontal, occipital, and parietal lobe epilepsy. Temporal lobe epilepsy is the most common form of focal seizures [81], [82]. It can be divided into different regions in the temporal lobe structures. Frontal lobe epilepsy is the second-most common kind of epilepsy. Frontal lobe seizures are usually brief

and tend to occur during sleep. The clinical expressions of this seizure can vary depending on the area of the frontal lobe affection. The symptoms are involuntary and repetitive movements, abnormal body posture, paresthesia, and difficulty in speaking. Finally, because of the complex functions of the parietal lobe, its epilepsy is infrequent. Also, seizures of parietal origin have no typical symptomatology and no particular EEG signature [16], [81], [82].

2.4 Brain and heart interactions in epilepsy

2.4.1 The autonomic nervous system

In 1628, for the first time, William Harvey hinted at a connection between the heart and brain and after that, in 1947, Bayer *et al.* were the first to report that cerebral vascular disease can generate myocardial infarction and arrhythmia [83]. Now, neuroradiology is an emerging specialty that deals with the interaction between brain and heart, i.e., the cardiac injury can affect the brain, and the brain injury can affect the heart [84].

Also, for many years, various anatomic and physiological works of research of the cardiac autonomic nervous system (ANS) have been done to prove the relation between the brain and heart and found out its complexity [85]–[89].

As briefly explained in section 2.1.1, the brain and spinal cord form the central nervous system (CNS) while the system extended to outlying or peripheral parts of the body is called peripheral nervous system (PNS). An overview of the CNS and PNS is illustrated in Figure 2-9. A subdivision of the peripheral nervous system, the autonomic nervous system (ANS), controls the autonomic functions that happen below the level of consciousness. In other words, ANS controls the body's autonomic or involuntary functions, namely the blood pressure, body temperature, heart rate, hunger, thirst, etc. [48], [90], [91] (Figure 2-10).

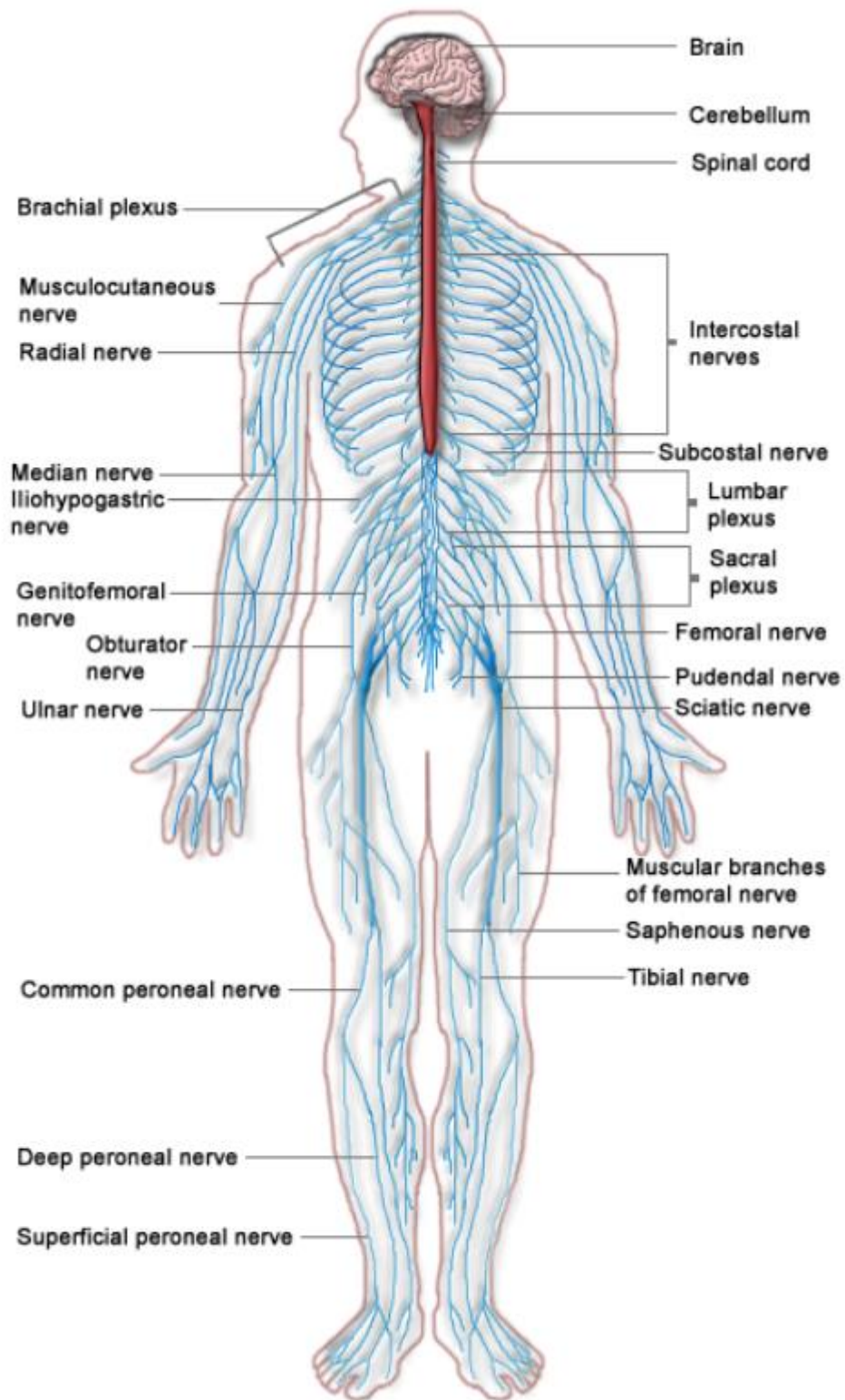


Figure 2-9. An illustration of CNS (in red and purple color) and PNS (in blue color) [49].

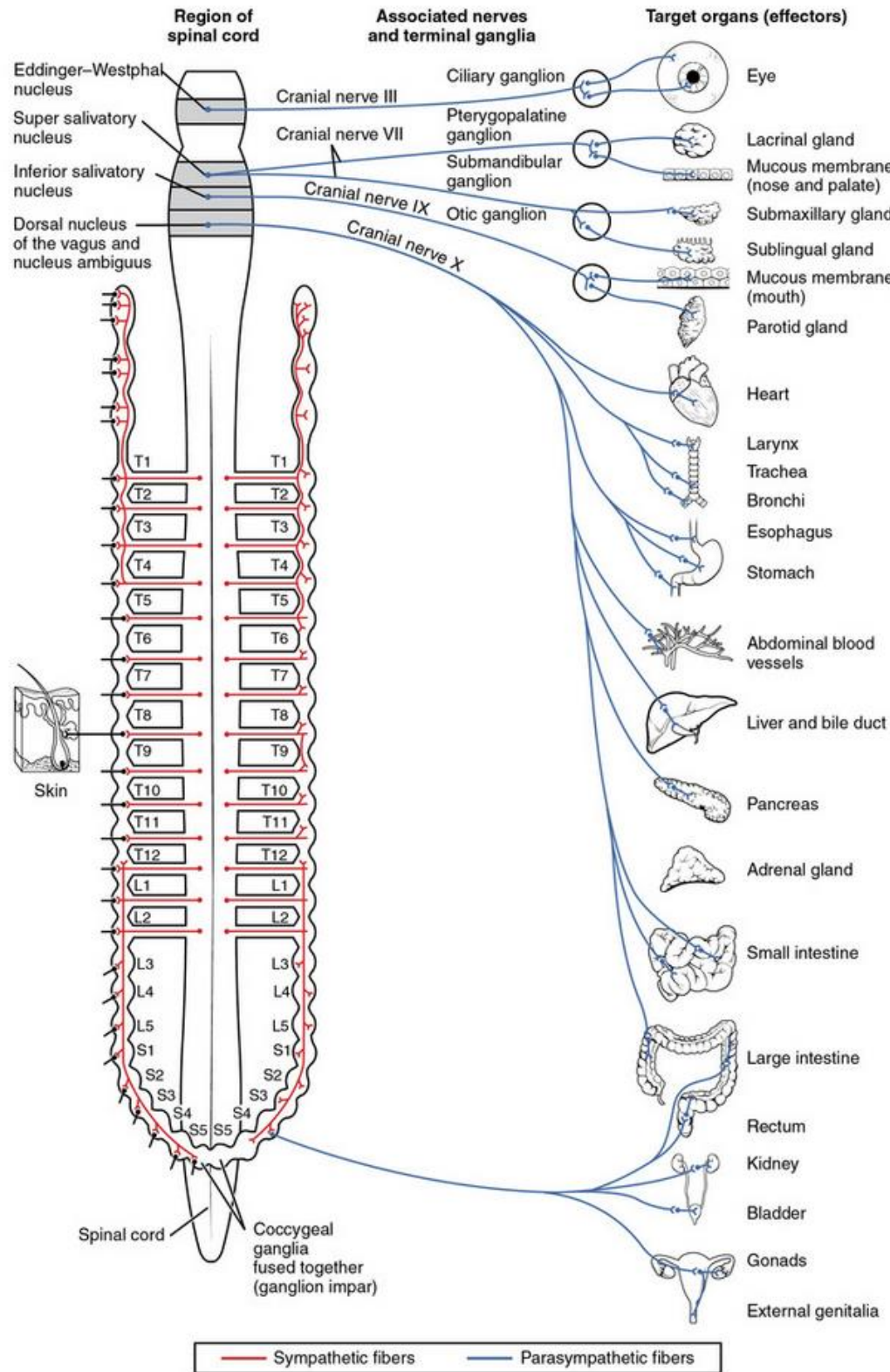


Figure 2-10. A picture of the autonomic nervous system and its connections [91].

The region of CNS in charge of the ANS is the hypothalamus (about the size of an almond in Figure 2-11) and located at the base of the brain [11], [85]. The Hypothalamus is also involved in long term memory and emotion. Interestingly, the hypothalamus and thalamus are located together as one regulates automatic survival behaviours (hypothalamus), whereas the other controls conscious behaviours (thalamus) [49].

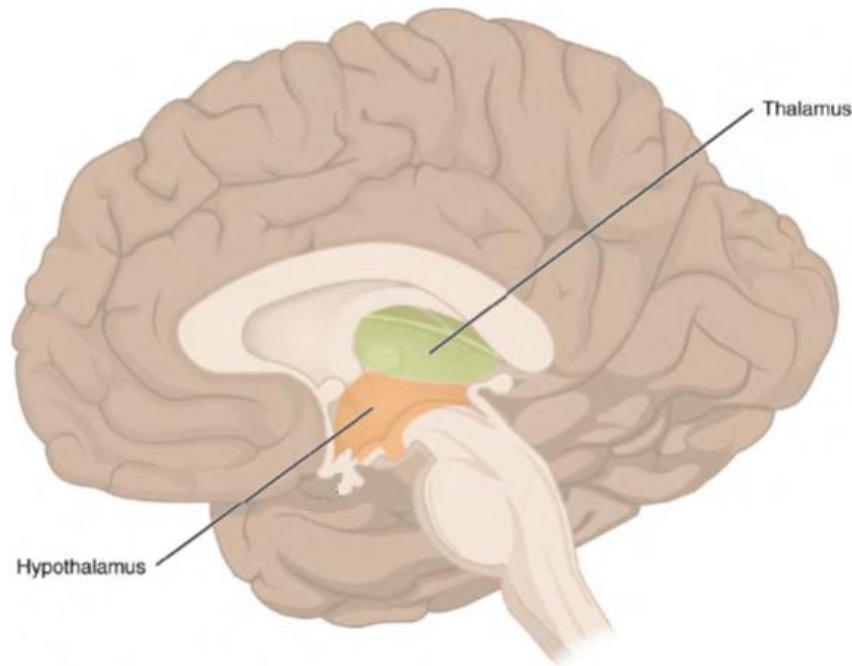


Figure 2-11. The location of the hypothalamus near the brainstem [11].

2.4.2 Influences of ANS on Cardiac Electrophysiology

The ANS consists of the parasympathetic and sympathetic nervous systems by which many organs are innervated by the two. Since the effects of the two systems on a given organ often have opposing actions, with one decreasing the activity of the organ and the other increasing it, this results in control of the organ's function. For example, sympathetic activation increases heart rate and heartbeat strength while the parasympathetic system does the opposite [48], [90], [91].

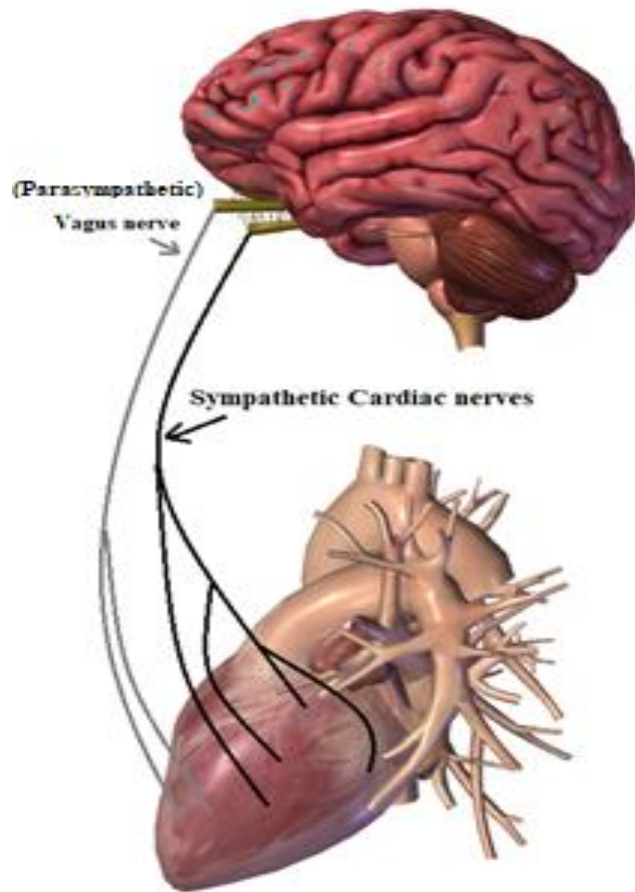


Figure 2-12. The innervation of the heart by parasympathetic and sympathetic systems.

While the heartbeat basically originates within the heart itself, the heart rate can be affected by the ANS and also by other external factors namely, emotions and drugs [90]. At a given time, just one system predominates under certain circumstances. Generally speaking, the sympathetic system dominates during an emergency while during resting, the other one, parasympathetic, prevails [48]. Actually, during physical activity or emotional issues, the sympathetic system stimulates the AV (the atrioventricular node) and SA (sinoatrial node) nodes and consequently, the heartbeats increase. On the other hand, the parasympathetic system through essentially vagus nerve fibers, tries to slow down the heart rate [85], [92].

2.4.3 Heart/Brain interaction in seizure period

Heart Rate Variability (HRV) is an assessment tool for the autonomic nervous system so that we can measure the variation of the heartbeat interval, the time elapsed between two successive R waves in the QRS complex of ECG [84], [93]. HRV reflects the instant variation in the heart rate and considers a complicated relationship between the heart and brain [93]–[95].

HRV is classified into four frequency sub-bands: HF (high frequency), LF (low frequency), VLF (very low frequency) and ULF (ultra-low frequency). Each sub-band can be interpreted to reflect different parasympathetic and sympathetic activities [84], [96].

Heart and brain health are interrelated. On the one hand, the autonomic nervous system (hypothalamus) regulates the cardiac involuntary actions, but on the other hand, the heart as a biological pump is responsible to deliver oxygen and other nutrients to the body as well as to the brain (Figure 2-13).

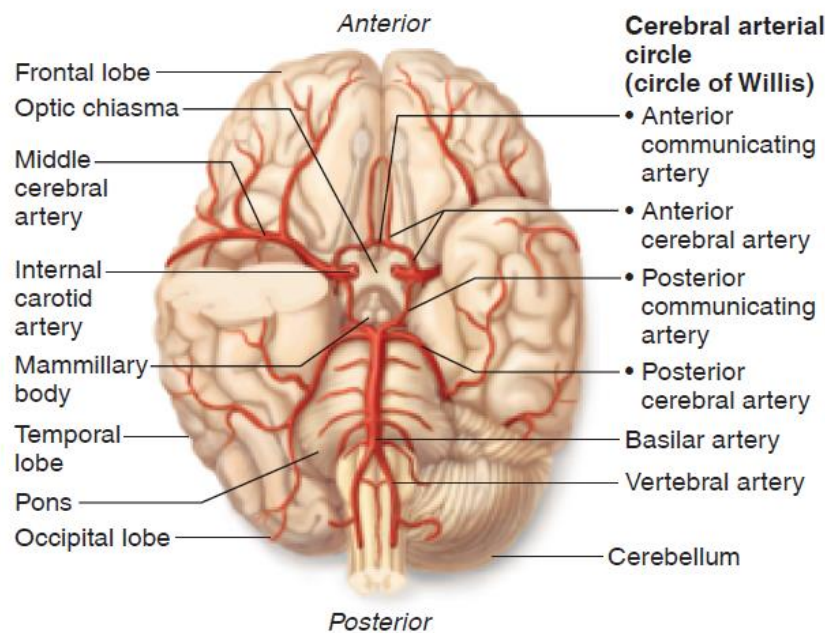


Figure 2-13 major brain arteries supplying the brain [97].

Investigating the pre-ictal changes on non-neurological signals can be helpful in forecasting the seizure. There is some evidence that shows changes in the human respiratory or heart rate 30 to 45 minutes before the seizure onset [98]–[100]. In [101], it was reported that the seizure warning in the ECG signal preceded the EEG onset for 23% of the seizures. Also in another study, it was reported that the heart rate rose from the preictal to the ictal interval for 74% of the participants [102].

The evidence shows that the autonomic nervous system, the cerebral blood flow (might be interpreted as variation in neural activity that cannot be detected with EEG recording), and the neurovascular coupling can be altered during the preictal period [21], [103], [104].

Then, working on the interaction between the brain and heart and the correlation between their signals becomes significantly important.

There is a trending interest in the development of algorithms that not only detect the onset of seizures but also predict the seizures before they occur. The analysis could involve Time-domain, Frequency-domain, Wavelet-domain, Empirical mode decomposition, Singular Value Decomposition, Principal component analysis, Independent component analysis, as well as non-linear approaches (e.g. chaos theory, correlation dimension, and entropy). All those methods have been used on their own or with the combination of one of the machine learning algorithms.

2.5 Prediction of the seizure by analyzing electroencephalogram

Research on seizure prediction via an electroencephalograph (EEG) recording started in the 1960s [105]. Since then, research on predicting epileptic seizures using an electroencephalogram (EEG) continued.

The authors in [2], first the 23 scalp EEG channels converted into a surrogate channel in order to enhance the SNR and after that they employed empirical mode decomposition (EMD) to the single channel to enhance the SNR. Then in the next stage, four statistical and three spectral moments were extracted and finally the feature vector was fed into three classifiers, k-NN (k-Nearest

Neighbors), NB (Naive Bayes), and SVM (Support Vector Machine). After comparing the performance of the classifiers, SVM outperformed among the other classifiers due to having the average sensitivity of 92.23% in prediction. The authors claimed that the maximum anticipation time was 33.46 minutes and the specificity was reported 93.38%.

The proposed work in [15], 7 subjects (5 dogs and 2 humans) with iEEG signal for two databases were required. The framework employed the Bernoulli-Gaussian Mixture model which considered as an active feature selection method and after that an adaptive seizure prediction framework was employed to minimize the number of labelled data and reduce the complexity of the framework. Finally, a SVM classifier coupled with the active learner utilized in the prediction of the seizure. The imbalanced ratio was reported from 2:1 to 20:1 and accuracy was just used to measure the performance of the classifier. They reported the performance of the prediction with $FP/h = 0.03-0.6$, a sensitivity of 40-97% for the dogs as well as a sensitivity between 40 and 74% and a $FP/h = 0.26-0.6$ for humans.

In [41], an SVM-based system to predict the seizures via iEEG signals from human and dog brains has been described. The continuous stream of records was segmented and labelled. The preictal period was defined as one hour before the seizure. Features were extracted from 20 s windows and one-hour segments were used for prediction. Three spectral characteristics were used as features in this work. These were spectral power in six Berger frequency bands, the signal in time-domain from six bands, and cross-correlation matrix. Features from 20 s windows were classified by SVM based system and accuracy was employed as the final performance index. Their sensitivity was about 90–100%, and their false-positive rate was about 0–0.3 times per day.

In [106], the authors used the k-nearest neighbours (k-NN) algorithm for the classification of seizure and non-seizure. Statistical parameters such as approximate entropy, standard deviation, mean absolute value, standard error etc. were employed followed by regression analysis for predicting seizures. Their algorithm was applied to patients of different age groups. Among all of the methods of classification like support vector machine (SVM), artificial neural network (ANN), linear discriminant analysis (LDA), naive Bayes (NB), and RBF neural network (RBFNN), k-NN was found the best classifier.

The approach proposed in [51] shows a high correlation level among pairs of electrodes for the delta sub-band and a reduction in the Lyapunov index for the gamma sub-band. They employed power spectral density and statistical analysis to define threshold levels for the delta sub-band. The result of this work showed that the seizure signature can be taken out from two frequency sub-bands: the lowest and the highest. For the sake of validation, six participants from both genders and different age groups with temporal epilepsy were studied using the Freiburg database with an accuracy of 72%, and a false-positive rate of 0%.

In [107], the authors proposed a non-invasive wearable EEG and they tested the algorithm on 16 subjects of Children's Hospital Boston based on scalp EEG recordings from 22 pediatric patient database. This paper not only proposed a novel approach in detection of the seizure but also in prediction of the system. The system remove the artifact from the EEG by the Cauchy-based filter and after that the system employs AIS (Artificial Immunity System) architecture which is the main forecasting mechanism of the designed work. The state-of-the-art architecture could predict seizures with accuracy of 72% and 14.57 second average anticipation time.

In [40], the authors utilized the Genetic algorithm as an effective and extensive feature extraction approach in classification of EEG signal of epileptic patients. In this work the frequency domain features along with the nonlinear ones were extracted and then an optimal feature subset was selected. After that the feature subset were applied to various classifiers such as k-NN, decision tree, AdaBoost, MLP (Multi-Layer Perceptron), and NB and accuracy was used in evaluating and comparing the performance of the work which has been claimed to be up to 99% for the two classification problem.

The authors in [108], introduced an adaptive pattern learning framework in prediction of ten epileptic patients. The features were extracted at the first-level were the short-term largest Lyapunov exponent, averaged pairwise Euclidean distance, T-statistic, and Pearson correlation and then in the second- level feature extraction four measures were captured from the temporal fluctuation of the first-level features. Three prediction rules were implemented and assessed based on the continuously-updated patient-specific pattern library for each individual, containing the adaptive probabilistic prediction (APP), adaptive linear discriminant- analysis-based prediction (ALP), and adaptive Naive Bayes-based prediction (ANBP). The highest testing prediction

accuracy (PA) were 79%, 78%, and 82% for the APP, ALP, and ANBP prediction system, correspondingly which PA defined as $(\text{sensitivity} + \text{specificity}) / 2$.

In [38], 21 patients from the Epilepsy Centre of the University Hospital of Freiburg were retained and 5 minutes prior to the seizure onset and 10 minutes of the interictal segments were investigated. They reported that a 10s window outperformed among 5s and 15s due to being more consistent with respect to the AECR (Average Energy Concentration Ratio) values of preictal and interictal segments. Three SVM kernels were compared together and finally RBF (Radial Basis Function) kernel SVM overtook the rest with the sensitivity of 91.95% and false positive rate of 2.14.

In [53], the posterior probability of SVM classifier employed and later this value was compared for various window sizes and then, the 90 s window (from 10- 600 seconds) was chosen due to having the highest probability. In this work, the dataset of Kaggle with iEEG information and contained 61 training and 196 test segments.

In [3], Panichev *et al.* required two patients from Kaggle website along with 5 epileptic dogs. Various classifiers namely Naïve Bayes (NB), Logistic Regression (LR), Support Vector Machine (SVM), Decision Tree (DT), k-Nearest Neighbor (k-NN) and Discriminant Analysis (DA) were employed to investigate the effect of the cross correlation of iEEG channels. The area under the curve (AUC) was the only performance characterises of the classifiers in this study. The window length (T_w) varied from 0.5 to 300 seconds and they found out in humans, best classification was showed by SVM classifier for a time window $T_w = 60$ s ($\text{AUC} = 0.9349$) while for seizure prediction in dogs the highest obtained AUC was 0.9432 with SVM and $T_w = 30$ s.

In [35], Priyanka *et al.* employed epilepsy patients from Neurology & Sleep Centre, New Delhi dataset. The signals from the scalp EEG were filtered between 0.5 to 70 Hz and then divided into preictal, interictal and ictal stages. At the next step, various features were extracted namely, Mean, Maximum, Variance, and Entropy. Then, in order to reduce the dimensionality of the database and increase the accuracy, three feature ranking were applied. A comparison was performed with respect to the normal signal over other stages such as ictal, preictal, and interictal. Finally an artificial neural network model was implemented with sensitivity of 95.8% and specificity of 97.2%. During the whole process, accuracy was considered as the only evaluation metrics.

2.6 Prediction of the seizure based on ECG signal

As for the prediction based on ECG signals, many works have been conducted [95], [96]. The study in [109] examined 8 heart rate variability features of 7 patients in the prediction of epileptic seizures from clinical data. The results showed that HRV features, such as mean Heart Rate (HR), Standard Deviation of the instantaneous beat-to-beat variability of data (SD1), Standard Deviation of the continuous long-term R-R intervals (SD2), and low frequency (LF)/high frequency (HF) ratio, changed 5–10 minutes before seizure onset in all seizure episodes. Their proposed seizure prediction algorithm achieved a sensitivity of 88.3% and specificity of 86.2%. They claimed that interpreting the HRV parameter changes is more difficult than EEG signal and needs further investigation. Finally, they concluded that a combination of EEG and ECG signals can help to enhance the results and efficiency of the algorithm.

Also, in [95] a novel HRV based seizure prediction algorithm was introduced. In this method, the RR interval of epileptic patients was converted into 8 HRV features such as mean NN, SDNN (Standard deviation of the NN (R-R) intervals), RMSSD (Root Mean Square of the Successive Differences), total power, NN50 (number of pairs of NN intervals that is longer than 50 ms), LF, HF, and LF/HF. The result of the work showed a 91% sensitivity with a false positive rate of 0.7 times per hour. It is reported that the authors could predict ten out of eleven seizures successfully. The proposed approach can be considered as a useful tool in daily life application.

Therefore, parameters like mean NN, SDNN, RMSSD, total power, NN50, LF, HF, and LF/HF can be employed as valuable features to distinguish a seizure from a non-seizure.

2.7 Prediction of the seizure based on EEG and ECG signals

In [110], an approach to predict epileptic seizures from EEG and ECG signals was introduced. It showed that an ECG signal would be a useful approach in the prediction of the seizures as the electrical activity of the ECG is easier to be analyzed compared to the EEG signal. These bio-

signals were individually examined for about 15 minutes before the seizure onset and non-seizure sections. They found a promising pattern with consideration of both brain and heart signals, which can forecast the seizure onset with a 94% average accuracy. The EEG was classified with a linear Bayes classifier and the ECG with a k-Nearest Neighbors (k-NN) classifier. k-NN classifier showed an effective classifier for prediction based on the sensitivity and accuracy measurements.

In [111], the authors investigated the functional interactions between EEG and ECG on a data of a mice in with periods with and without seizures in the frequency domain from 0 to 200 Hz. They found out in the high frequency from 130 to 170 Hz the Generalized Partial Directed Coherence (GPDC) shown the highest value during seizures in the epileptic animal. To measure the GPDC value they fitted the 5-dimensional (4 iEEG electrodes and 1 ECG electrode) with a 7th order multivariate autoregressive (MVAR) models to 10s consecutive non overlapping EEG and ECG sections.

2.8 Discussion

EEG data represent different neural activities of the brain which are of great value to the field of neuroscience. Using EEG, the neurological state of the subject or patient can be determined, which can further be used for clinical, therapeutic, medicinal and physiological applications.

From a clinical perspective, EEG diagnoses brain malfunction like attention deficit hyperactivity disorder, Alzheimer's illness, Schizophrenia, Epilepsy, etc. It is also capable of monitoring the state of coma, alertness and brain death. EEG analysis also helps in testing the effect of drugs and for detecting the origin of seizures, brain injuries, and lesions [76], [112], [113].

For the above-mentioned applications of EEG data, representing EEG signals is a crucial issue. However, one of the biggest problems is the non-stationary and multidimensional nature of the EEG data subject to frequent and non-predictable changes. Besides the presence of errors and artifacts of various types during recording, scalp EEG signals require extensive pre-processing and filtering mechanisms [6], [25], [114].

For patients with infrequent seizures, the frequency of seizures becomes difficult to discover. Even if a patient has frequent seizures, the types of seizures (focal, generalized, etc.) can be different [76].

Assuming the algorithms based on EEG can successfully predict epileptic seizures, the patients who have epilepsy would need to wear a predesigned cap with recording electrodes for online EEG monitoring. Obviously, such a method threatens their privacy and people may tease them. Too many false alarms are another purpose for patients' reluctance to employ such approaches as they lose their confidence.

Addressing these issues and their complexities will require the continuation of investigation in finding an algorithm that can provide a sufficient compromise of performance, reliability, and privacy. As discussed in [8], [115], the prediction of epileptic seizures using HRV can significantly resolve the problems of permanent recording, noise, and privacy. A lot of research has been done to support the hypothesis that epileptic seizures can be anticipated by employing the HRV. However, the results obtained in this field are still not comparable with EEG-based algorithms and still have great potential for optimization.

Nevertheless we should not forget that if the brain is the source of epileptic seizures, the heart is the second organ affected by seizures, and its behaviour cannot be anticipated through results obtained by a method based on data from the brain. Even though the results for ECG may be better than EEG results, EEG is an important complement for ECG because the ECG methodology is based on HRV, which can be affected by any daily activities for the patient; here is where EEG becomes relevant for this task. Therefore, the aim will be devoted to improving the performance of seizure prediction in a combination of both signals [95], [110], [116], [117].

2.9 Conclusion

After providing some background information, reviewing existing research works, and possible methods to predict the seizure, applying machine learning and employing different algorithms to

predict seizure via iEEG signals should enhance the performance of the prediction. However, before the implementation step, we need to work on other biological signals such as ECG to better understand the mechanism behind the prediction of the seizure. Therefore, we will employ a method that can forecast the seizure at an adequate time with higher sensitivity and that can be used in practice for any patient.

Chapter 3

Data Pre-Processing and Feature Engineering

In this chapter, the database used in this study, the preprocessing, and the feature engineering of the proposed approach, will be introduced.

3.1 Freiburg seizure prediction EEG database

The Freiburg EEG Database is one of the most cited resources employed in predicting and detecting experiments. The interictal and preictal intracranial electroencephalogram (iEEG) recordings of the Freiburg database (FSPEEG) was offered in the early 2000s as an EEG database [58]. The database consists of intracerebral (strips, grid and depth electrodes) EEG recordings from 21 epileptic patients. It contains six intracranial electroencephalography (iEEG) electrodes with a sampling frequency of 256 Hz and a 16-bit A/D converter.

The database contains 24 hours of continuous and discontinuous interictal recordings for 13 patients and eight patients, respectively. Each participant had 2 to 5 preictal recordings with about 50mn preictal recordings. This database contains 582 hours of EEG data, including preictal recordings of 88 seizures. Since 2012, the FSPEEG has been complemented and substituted by the larger EPILEPSIAE database which covers datasets of the marked long-term invasive and non-invasive EEG recordings from 31 participants [118]. An overview of the dataset is inserted in Table 3-1. However, even if such database is very useful, we have to note that due to the limited number of patients available, the results we will obtain should be relativized.

Figure 3-1 shows an illustration of epileptic EEG signals (from the database) for a patient with temporal lobe epilepsy. In this figure, the voltage of a channel for patient #4 for each sample of time is depicted. The period in which the seizure takes place is ictal, and the time before the seizure

onset happens is preictal. The postictal is the period after an epilepsy seizure, which can last more than an hour and interictal refers to the period between seizures. Also, in Figure 3-2, one-hour data from the patient with frontal lobe epilepsy (FLE) and temporal lobe epilepsy (TLE) is depicted. As we can see, epilepsy from the frontal lobe (which takes about 7 seconds) is shorter than the temporal one (which takes about 91 seconds). In addition, the morphology of the signal over a period of time for each type of epilepsy is completely different. In fact, we can observe some hints in the preictal stage of the temporal one, but not the same for the other one.

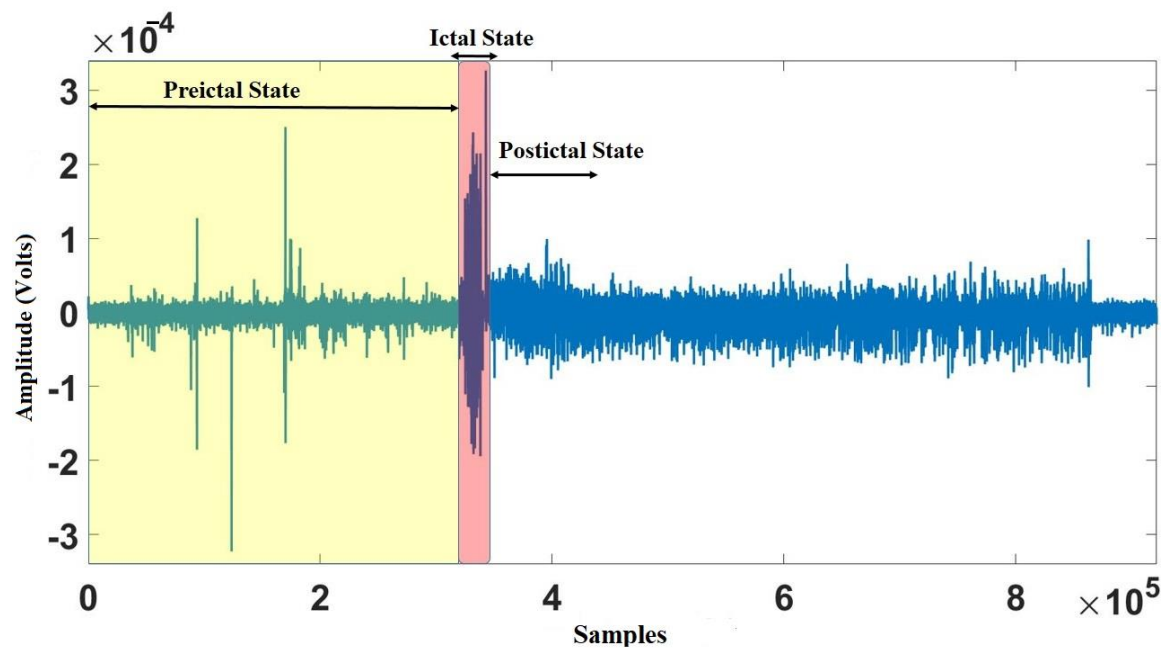


Figure 3-1. An overview of an EEG signal containing seizure for patient #4, suffering from temporal lobe epilepsy. Red bar indicates the duration of the seizure.

Table 3-1: The information of patients in the dataset. SP = simple partial, CP = complex partial, GTC = generalized tonic-clonic; H = hippocampal origin, NC = neocortical origin; d = depth electrode, g = grid electrode, s = strip electrode

Patient #	Sex	Age	Seizure type	H/NC	Origin	Electrodes	Seizures analyzed
1	F	15	SP, CP	NC	Frontal	g, s	4
2	M	38	SP, CP, GTC	H	Temporal	d	3
3	M	14	SP, CP	NC	Frontal	g, s	5
4	F	26	SP, CP, GTC	H	Temporal	d, g, s	5
5	F	16	SP, CP, GTC	NC	Frontal	g, s	5
6	F	31	CP, GTC	H	Temporo/Occipital	d, g, s	3
7	F	42	SP, CP, GTC	H	Temporal	d	3
8	F	32	SP, CP	NC	Frontal	g, s	2
9	M	44	CP, GTC	NC	Temporo/Occipital	g, s	5
10	M	47	SP, CP, GTC	H	Temporal	d	5
11	F	10	SP, CP, GTC	NC	Parietal	g, s	4
12	F	42	SP, CP, GTC	H	Temporal	d, g, s	4
13	F	22	SP, CP, GTC	H	Temporo/Occipital	d, s	2
14	F	41	CP, GTC	H, NC	Fronto/Temporal	d, s	4
15	M	31	SP, CP, GTC	H, NC	Temporal	d, s	4
16	F	50	SP, CP, GTC	H	Temporal	d, s	5
17	M	28	SP, CP, GTC	NC	Temporal	s	5
18	F	25	SP, CP	NC	Frontal	s	5
19	F	28	SP, CP, GTC	NC	Frontal	s	4
20	M	33	SP, CP, GTC	NC	Tempo/Parietal	d, g, s	5
21	M	13	SP, CP	NC	Temporal	g, s	5

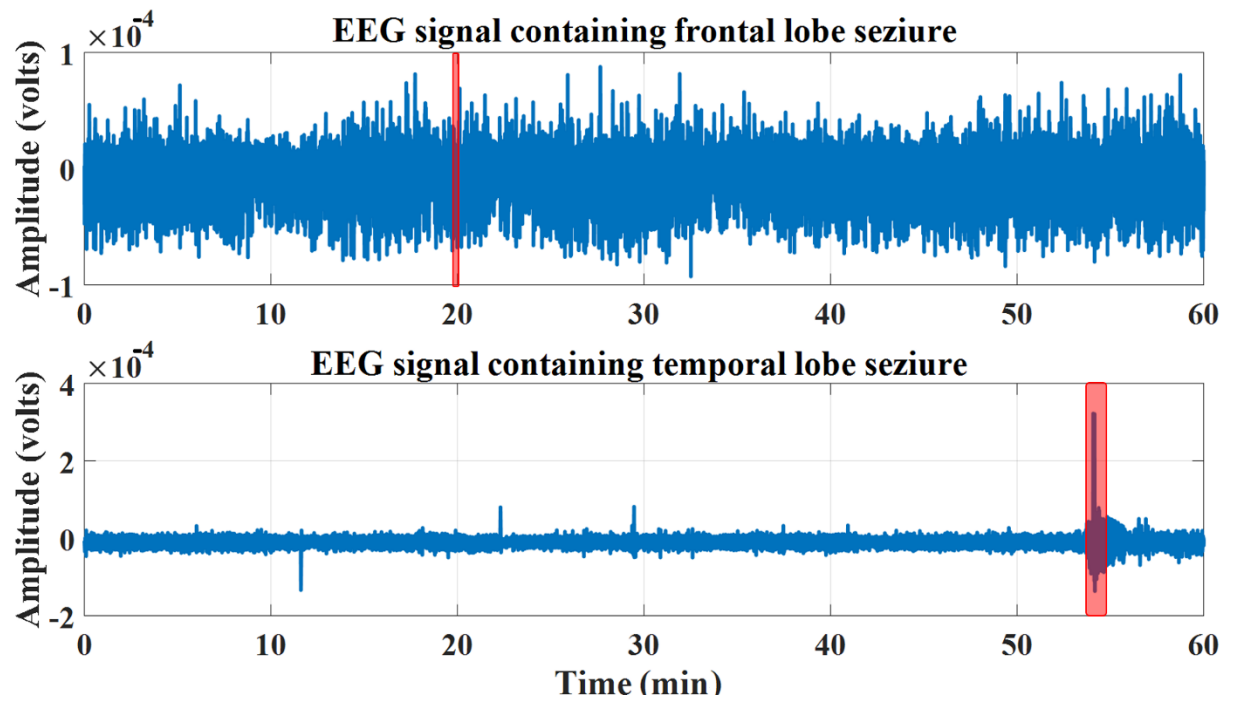


Figure 3-2. An overview of an EEG signal containing seizure for patient #1 suffering from frontal lobe epilepsy (on the top) and patient #2, suffering from temporal lobe epilepsy (at the bottom). Red bar indicates the duration of the seizure.

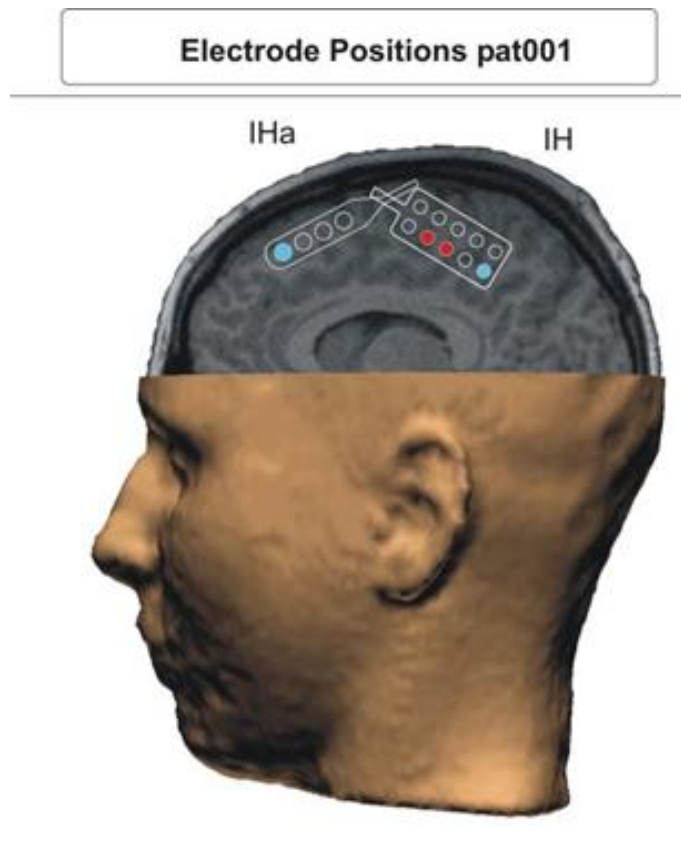


Figure 3-3. The position of the strip electrode for patient #1 (pat001) [58].

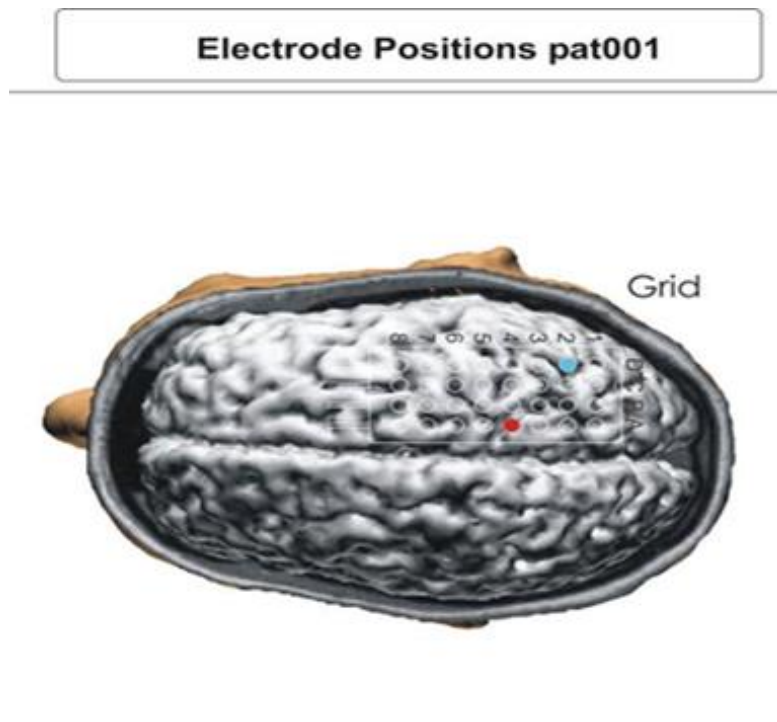


Figure 3-4. The position of the grid electrode for patient #1 (pat001) [58].

Electrode Positions pat002

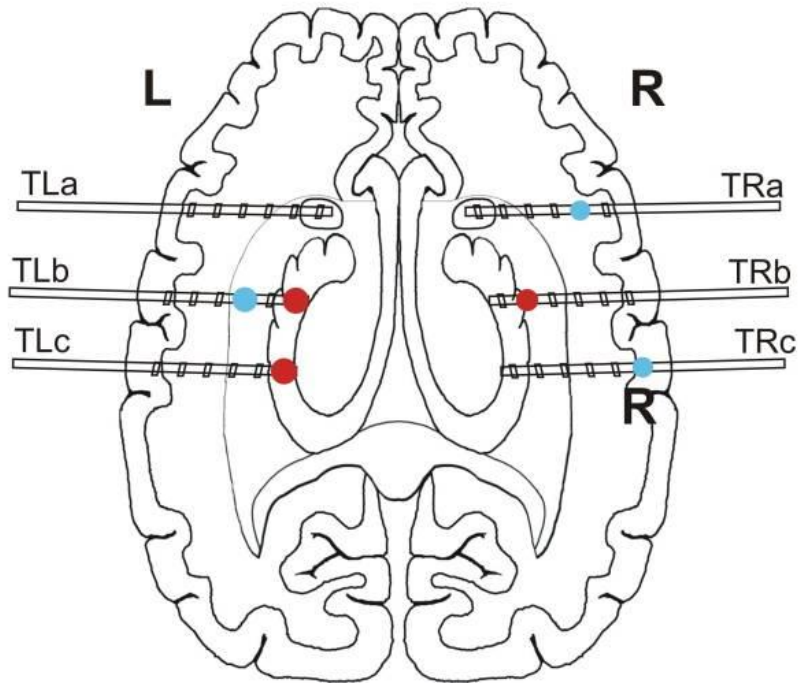


Figure 3-5. The position of the depth electrode for patient #2 (pat002) [58].

Figure 3-6, shows a screenshot of a patient from the newest version of the Freiburg database, Patient FR_264, contains EEG, iEEG, ECG, and EOG.

3.2 Artifacts and Noise Removal

The artifacts in the EEG recordings are of different kinds which can originate from interior and exterior sources.

recordings				duration	blocks	net dur.	chan.	sampl.rate
1:	050125da	25.01.'2005	- 25.01.'2005	(1:03:09)	2	~1h	67	256Hz
2:	050125db	25.01.'2005	- 27.01.'2005	(1 day, 15:05:18)	40	~37h	67	256Hz
3:	050127da	27.01.'2005	- 31.01.'2005	(4 days, 0:05:09)	98	~96h	91	256Hz
					140	~134h		

Surface electrodes:
 2 x **10-10**: FT9, FT10
 19 x **10-20**: FP1, F7, FP2, F3, F4, C4, P3, P4, O1, O2, T7, T8, P7, P8, FZ, CZ, C3, F8, PZ
 1 x **ECG**: ECG
 2 x **EOG**: EOG1, EOG2
 1 x **EMG**: EMG

Invasive electrodes:
TLA (1x4 strip), **TLB** (1x4 strip), **TLC** (1x6 strip), **TRA** (1x4 strip), **TRB** (1x4 strip), **TRC** (1x6 strip), **BRA** (1x4 strip), **BRB** (1x4 strip), **BRC** (1x6 strip), **BLA** (1x4 strip), **BLB** (1x4 strip), **BLC** (1x6 strip), **HR** (1x10 depth),

[view details ...](#)

seizures [text-file with seizures](#)

#	classif.	onset	pattern	vigilance	origin
1:	CP	eeg: 26.01.'05 23:48:52.945313	rhythmic beta waves	unclear	HR6,BRA1,BRA2,BRA3,BRA4,BRB1,BRB2,BRB3,BRB4,HR1,HR2,HR3,HR4,HR5,HR7,HR8,HR9,HR10
2:	CP	eeg: 27.01.'05 11:06:21.800781	rhythmic beta waves	unclear	BLA1,BLB1,BLB2
3:	CP	eeg: 27.01.'05 19:12:30.074219	rhythmic sharp waves	sleep stage II	TRA1,TRA2,TRA3,TRA4,BRA1,BRA2,BRA3,BRA4,BRB1,BRB2,BRB3,HR1,HR2,HR3,HR4,HR5,H
4:	UC	eeg: 27.01.'05 23:32:13.699219	rhythmic beta waves	sleep stage II	BLC1,BLC2
5:	UC	eeg: 27.01.'05 23:44:08.398438	rhythmic beta waves	sleep stage II	BLC1,BLC2,BLC3,BLC4
6:	UC	eeg: 27.01.'05 23:55:16.351563	repetitive spiking	sleep stage II	TRA1,TRA2,TRA3,BRA3
7:	CP	eeg: 28.01.'05 00:03:10.425781	rhythmic sharp waves	sleep stage II	TRA2,BRA1,TRA1,TRA3,TRA4,HR4,HR5,HR6,HR7
8:	CP	eeg: 28.01.'05 09:50:29.683594	repetitive spiking	sleep stage II	TRA3

Patient FR_264

female epilepsy onset: 22 years

commentary: None

admission [\[details...\]](#)

Jan. 25, 2005 @ UKLFR
 patient age: 48
 presurgical: surgery is not offered
 surface & invasive EEG

SeizureFrequencies: complex partial: 1.5 / month
 simple partial: 5.0 / month

complications:

EEG-foci: inv_1: temporal, none, bilateral
 surf_2: temporal, none, bilateral

etiology idiopathic

Spect -

Pet hypometabolism

MRI:

[mri_264102.img](#)

[mri_264102.hdr](#)

[Details...](#)

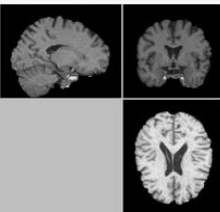


Figure 3-6. A screenshot of the information about a patient from the new database.

We can categorize the sources of the artifacts in EEG as

- Internal/Physiological: Ocular Artifacts, Muscle Artifact (EMG), Cardiac Artifacts, Respiration, Sweat, ...
- External: Movement (Head, body, limbs movement, ...)
- Environmental: Power line Noise, Baseline Noise due to the loose electrode ...

A summary of various artifact types and their sources show in Table 3-2 [119].

Table 3-2: Different kinds of artifacts and their sources [119].

Physiological / Internal				Extra-physiological / External		
Ocular	Cardiac	Muscle	Others	Instrumental	Interference	Movement
Eye blink, Eye movement, Eye flutter, REM sleep, etc.	ECG; Pulse.	Chewing; Swallowing; Clenching; Sniffing; Talking; Scalp contraction; etc.	Glossokinetic; Skin; Respiration; etc.	Electrode displacement and pop-up; Cable motion; Poor ground; etc.	Electrical; Magnetic; Sound; Optical; EM waves; etc.	Head movement; Body movement; Limbs movement; Tremor; Other movements, etc.

Most of the above artifacts are related to the scalp EEG. However, for the iEEG signal, due to employing the specific electrodes and implanting on the cortex surface, we would just need to eliminate the power line noise [14], [15], [41]. In fact, biological recordings are frequently contaminated by the power line frequency (50/60 Hz) and such interferences, that can come from any conductor, lamp light and/or electronic device, are received by the electrodes and acquisition system. This artifact is spread in different iEEG channels and can generate errors during the analysis of the record. Figure 3-7 shows that only the first odd harmonic of 50 Hz (frequency of the power line in Germany) can affect the signal. The normalized frequency should be about 0.39π rad/sample for 50 Hz based on the $2\pi f/f_s$ formula.

3.3 Brain wave sub-bands

For processing the brain wave, we recorded the brain activity with an EEG electrode. It can be divided into different sub-bands. Amongst the simple waveforms are the alpha, beta, theta, delta, and gamma rhythms [120]:

- Delta waves have a frequency between 0.5 and 4 Hz. These waves are prominent during sleep and involve in triggering and releasing the hormones that help the body settle and recuperate. Delta waves play an important role in transporting new learning and memories into long-term memory storage during sleep time.

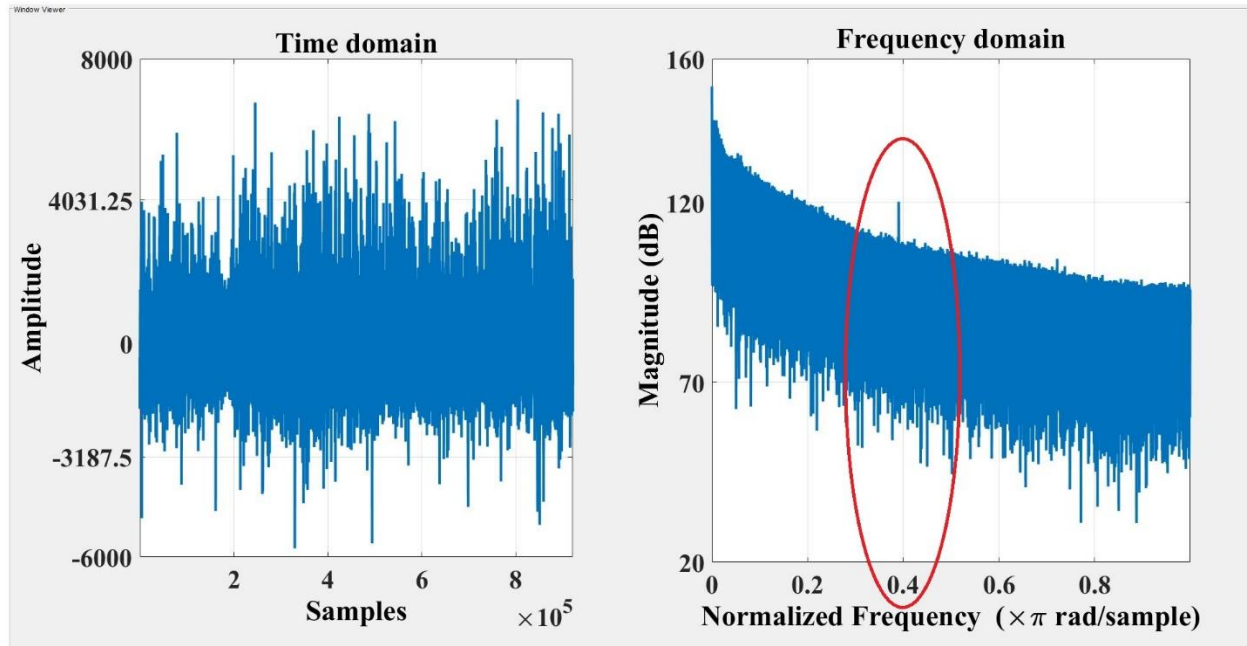


Figure 3-7. An illustration of an EEG signal in time and frequency response.

- Theta waves happen at a frequency range of 4 to 8 Hz, mostly observed in children and young adults.
- Alpha waves present in the frequency range of 8 to 13 Hz in a regular rhythm and are strong around the occipital and frontal lobes. They are present when we are awake, typically with closed eyes and do not actively process the information. Usually, they disappear when we open our eyes or begin to mentally concentrate [63], [121].
- Beta waves range of frequency are between 13 and 30 Hz. They appear when we are anxious, depressed, or when we use sedatives.
- Gamma waves have a frequency higher than 30 Hz. They involve higher processing tasks and cognitive functioning. These waves are the fundamental waves for learning, memory and information processing [120], [122].

Figure 3-8 shows different brainwave frequencies [120]. FIR (finite impulse response) filter is one of the most basic elements in digital signal processing. Among different FIR windowing methods such as Rectangular, Hamming, Hanning, Blackman, and Kaiser with different beta parameters (5,

8, 12) the latest one, Kaiser with beta of 12 showed the best result according to the main lobe, side lobe and SNR. Filters' order for different sub-bands varied from 158 to 226 [123].

We designed the FIR filters based on the “Filter Designer Toolbox” in Matlab. The complete characteristic of the filters are described in Table 3-3. Since the FIR windowing approach employed is symmetric, the filter can be considered as linear phase-frequency response [24], [124].

In this work, to eliminate the first odd harmonic of power line interference at 50 Hz (based on the database from Germany), we further divided the higher frequency sub-band (Gamma) into two parts instead of using a notch filter. Finally, EEG signals were split into the following sub-bands: Delta (0.5-4 Hz), Theta (4-8 Hz), Alpha (8-12 Hz), Beta (12-30 Hz), as well as two Gamma bands namely, low-Gamma (30-47 Hz), and high-Gamma (53-120 Hz) [23], [24], [125].

To decrease the effect of factors that cause baseline differences among different data sets, we also normalized the EEG signals with Z-scores (expressed in terms of standard deviations from their means). After pre-processing, the signal is ready to extract different features to feed for the classification.

Table 3-3: The parameters of the filters.

Band\frequency	Lower cut-off frequency (Hz)	Higher cut-off frequency (Hz)	Filter order
Delta	0.2	5	226
Theta	2.7	8.6	198
Alpha	4.6	16.6	158
Beta	10.1	33.4	100
Gamma 1	28	47	200
Gamma 2	53	97	200

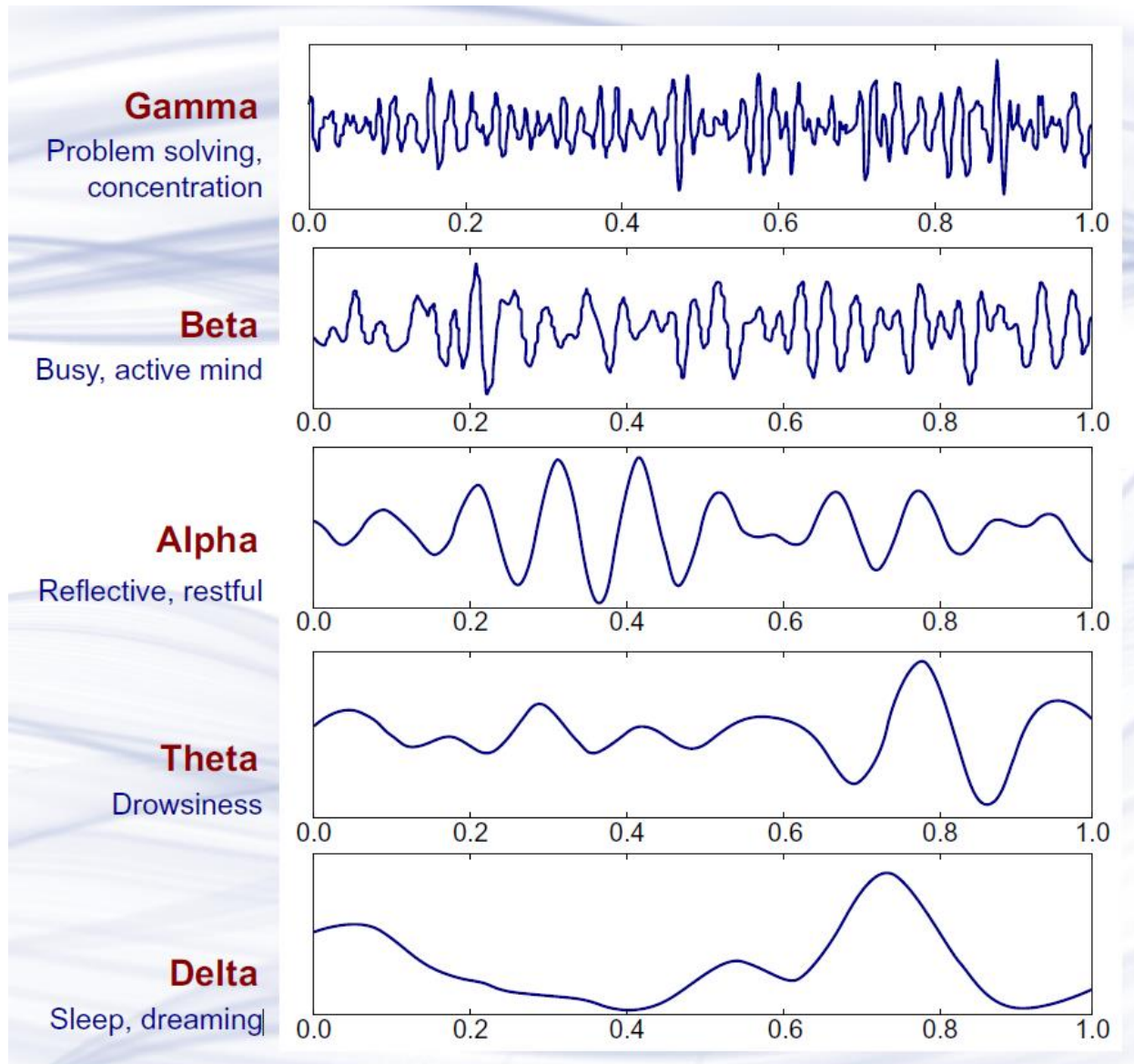


Figure 3-8: Typical rhythmic activities of an EEG signal [120].

3.4 Preprocessing of ECG signal

Figure 3-9 shows an illustration of a raw ECG from the new database for patient #265. The baseline drift was eliminated and the final smoothed signal depicted in Figure 3-10. After that, applying the Wavelet transform (db4) decomposed the signal into eight sub-bands. Then, by calculating a

threshold for R wave based on the max and mean of the wave, we can detect the QRS complex [126]. The PQRS complex detected is shown in Figure 3-11.

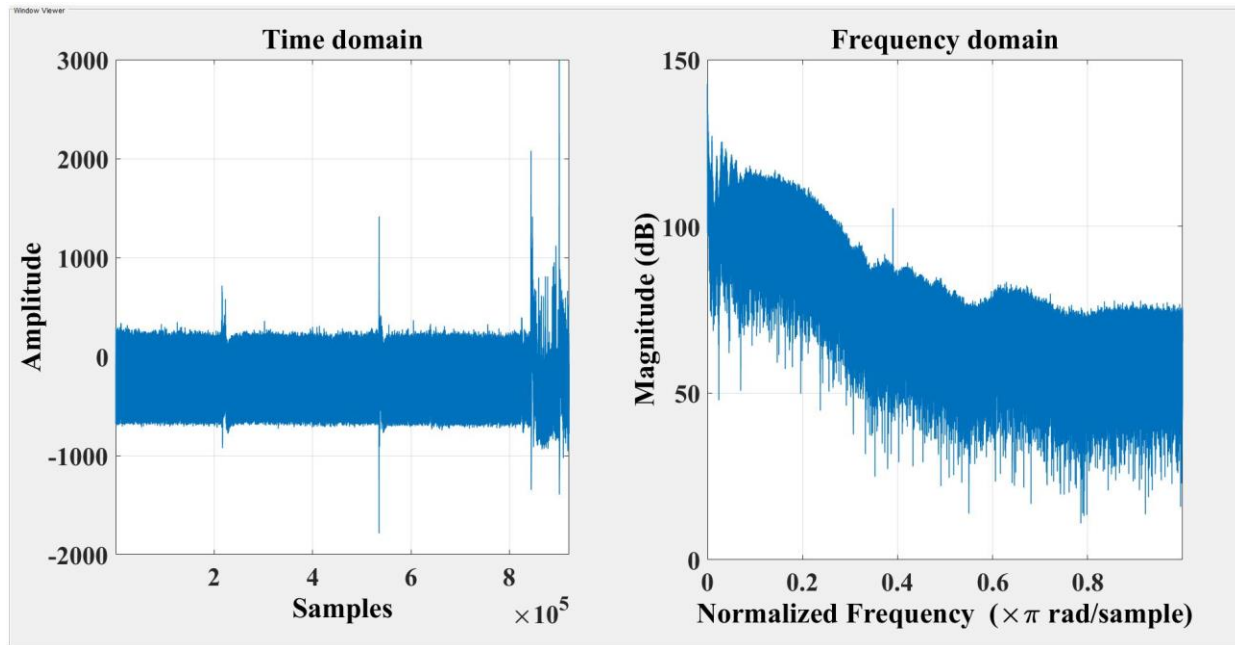


Figure 3-9: The time and frequency domain representation of the not clean ECG.

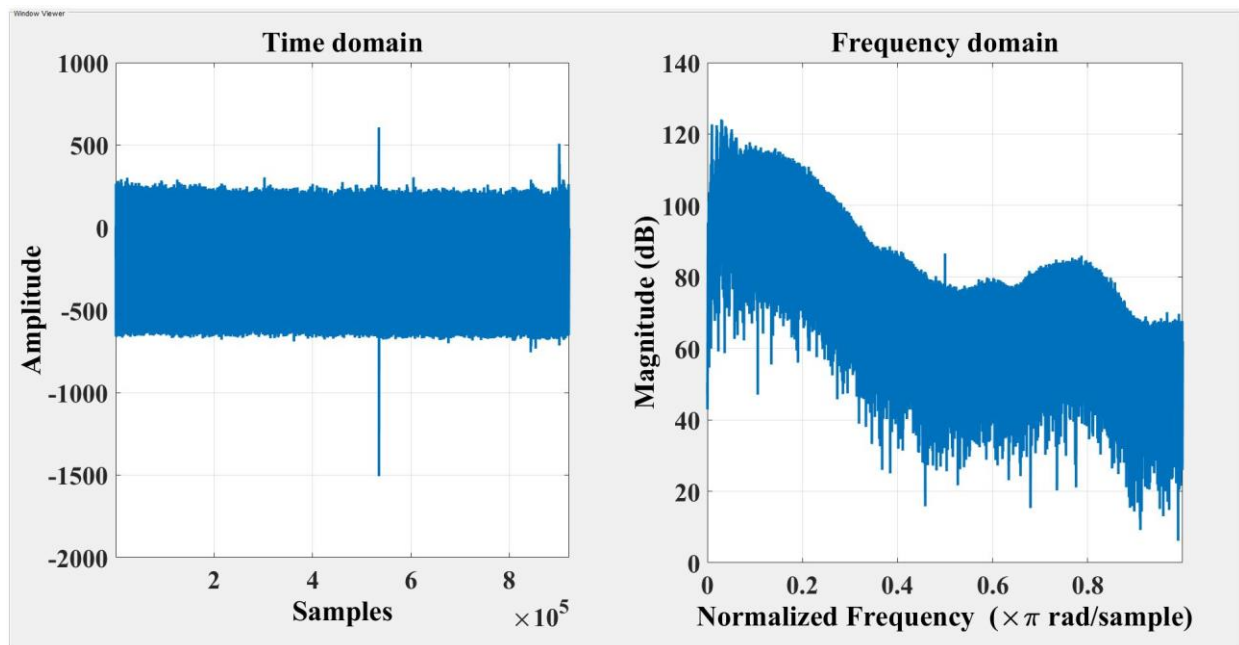


Fig 3-10: The time and frequency domain representation of the clean ECG.

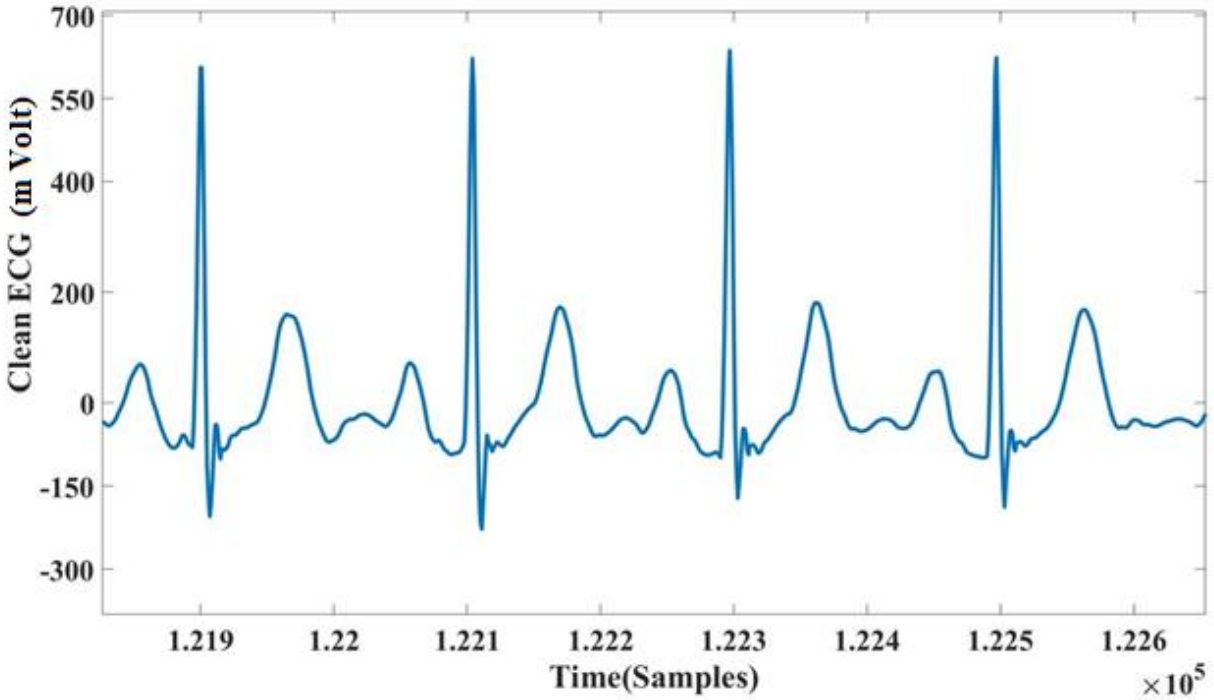


Fig 3-11: The time domain of the clean ECG.

3.5 Feature Engineering

3.5.1 Introduction

Several feature extraction techniques have been introduced in the last few years which can be classified as time, frequency, and hybrid (time-frequency) based approaches. Table 3-4 categorizes the various features that have been introduced in seizure prediction in these different domains [127]–[131].

Although the time-frequency domain features are most informative, the time-domain ones are more recommended to achieve discriminative information at a low computational cost. In fact, high-quality features can be defined as those that produce maximum class separability, robustness, and less computational complexity, e.g., less complex preprocessing – which do not need the burdensome task of framing, filtering, Fourier transform, and so forth [132], [133]. Thus, these

features, compared to other types, consume less processing power and time [134]. As a consequence, from the three above-mentioned domains, we retained, in this work, the features that belong to the time domain.

3.5.2 Feature Extraction

Feature is a variable that can represent the changes of a signal. In analyzing EEG signals, we selected common features that can discriminate between pre-ictal and interictal phases of the seizures. Note that, as will be discussed in Chapter 4, we first considered univariate features such as the Hjorth parameters (mobility and complexity), the average power in the considered frequency range, the mean of the power spectrum, the accumulated energy, the mean of the Fast Fourier Transform (FFT), and four statistical moments (mean, variance, skewness, and kurtosis) [41], [131]. Then, in chapter 5, we will concentrate more on time domain features that have been intensively employed by researchers in predicting seizures.

Table 3-4 Categorizing various features employed in seizure prediction in different domains.

Domain	NO.	Features
Time	1	Energy
	2	Mean
	3	Variance (VAR)
	4	Skewness
	5	Kurtosis
	6	Interquartile range (IQR)
	7	ZCR (zero cross rating)
	8	MAD (mean absolute deviation)
	9	Entropy
	10	Hjorth mobility
	11	Hjorth complexity
	12	Coefficient of Variation (COV)
	13	Root Mean Square (RMS)
	14	MAX Cross-correlation
	15	Autoregressive (AR) model
Frequency	16	power of frequency in sub 6 bands
	17	Mean of Absolute of FFT
	18	Spectral density estimation
	19	Global coherence
	20	SEP (Spectral edge power)
	21	the mean of the power spectrum
	22	coherence
	23	Spectral entropy
	24	SEF (Spectral edge frequency)
Time-Frequency	25	Energy wavelet coefficients (six levels)
	26	Hilbert transform (phase and magnitude)
	27	EMD (empirical mode decomposition)
	28	Hilbert-Huang Transform

3.5.3 Time domain features

3.5.3.1 Energy

This feature can be considered as a measure of the signal strength. Calculating the accumulated energy at a given time-point t , is a commonly used feature in finding abnormal behavior in the brain. For a given discrete signal, $x(n)$, the area under the squared of a signal is called energy and is expressed as [47], [135]

$$E = \sum_{n=-\infty}^{\infty} |x(n)|^2 \quad (3-1)$$

3.5.3.2 Mean

The mean of a discrete signal, $x(n)$, can be expressed as [136], [137]

$$\bar{x} = \left(\frac{1}{N}\right) \sum_{n=1}^N x(n) \quad (3-2)$$

where N is the number of the samples and $x(n)$ the discrete signal

3.5.3.3 Variance

The second moment of a signal is called the variance. The smaller its value, the higher the differences in frequencies across the samples [131]

$$S^2 = \left(\frac{1}{N-1}\right) \sum_{n=1}^N (x(n) - \bar{x})^2 \quad (3-3)$$

3.5.3.4 Skewness

The third statistical moment measures the asymmetry of the probability distribution about its mean.

$$\text{Skew} = \left(\frac{1}{N}\right) \sum_{n=1}^N \left[\frac{x(n) - \bar{x}}{S}\right]^3 \quad (3-4)$$

3.5.3.5 Kurtosis

The fourth statistical moment describes the flatness of a distribution real valued random variable.

$$\text{Kurt} = \left(\frac{1}{N}\right) \sum_{n=1}^N \left[\frac{x(n) - \bar{x}}{s}\right]^4 \quad (3-5)$$

3.5.3.6 Interquartile range

The Interquartile Range (IQR) feature is a measure of spread and variability based on dividing the data into four equal parts. The separated values Q_1 , Q_2 , and Q_3 for each part are named respectively first, second, and third quartiles. IQR is computed as the difference between the 75th and the 25th percentile or Q_3 and Q_1 as the following [47], [138]:

$$\text{IQR} = Q_3 - Q_1 \quad (3-6)$$

3.5.3.7 Zero crossing rate

The Zero Crossing Rate (ZCR) is the rate at which the signal changes signs or is the sum of all positive zero crossings into the EEG segment [139].

3.5.3.8 Mean Absolute Deviation

Mean Absolute Deviation (MAD) “as reported in [137], evaluates the robustness of the collected quantitative data.” In other words, it is the average distance between each data point and the mean. For a given dataset, $x = x_1, x_2, \dots, x_n$, MAD can be calculated as [137]:

$$\text{MAD} = \text{median} (|x - \text{median}(x)|) \quad (3-7)$$

3.5.3.9 Entropy

This feature is employed to quantify the degree of uncertainty and irregularity of a signal as well as the complexity of human brain dynamics. With $P(x)$ the probability mass function, “The signal uncertainty can be assessed in terms of repeatability of the signal amplitude”[137].

$$E_n = - \sum_{n=1}^N P(x(n)) \log P(x(n)) \quad (3-8)$$

3.5.3.10 Hjorth mobility

In [24], three parameters namely, Activity, Mobility and Complexity, were introduced as useful tools in indicating the statistical property of an EEG signal in time-domain. Since the activity parameter is the variance of a signal, this parameter can specify the origin of the power spectrum in the frequency domain. This means that if the Activity parameter is assigned a large value, then high-frequency components are present in the signal (and vice versa for a small value). The called Hjorth mobility parameter represents the square root of the variance of the first derivative of the signal, and it is proportional to the standard deviation of the power spectrum of a time series.

$$Mobility = \sqrt{var(x'(t))/var(x(t))} \quad (3-9)$$

In the above equation, $x(t)$ is a signal and $x'(t)$ its derivative. $var(-)$ is the variance of a signal over a period of time.

3.5.3.11 Hjorth complexity

The Hjorth complexity defines how the shape of a signal is analogous to an ideal sine curve. This parameter gives an estimation of the bandwidth of the signal.

$$Complexity = mobility(x'(t)) / mobility(x(t)) \quad (3-10)$$

We can have an estimation of the second and fourth statistical moment of the power spectrum in the frequency domain by employing the mobility and complexity, correspondingly. While Hjorth parameters are identified in time-domain, they can be useful for both time and frequency analysis. Interestingly, computation of the Hjorth parameters stands on variance, then the cost of their computation is significantly low [23], [24], [131].

3.5.3.12 Coefficient of Variation

The coefficient of variation (COV) is a measure is the division of the standard deviation to the mean of a signal [133]

$$Cv = \frac{s}{\bar{x}} \quad (3-11)$$

3.5.3.13 Root Mean Square

The Root Mean Square (RMS) of a signal can be calculated as [37]:

$$\text{RMS} = \sqrt{\frac{1}{N} \sum_{n=1}^N x(n)^2} \quad (3-12)$$

3.5.3.14 Maximum linear cross-correlation

As a simple bivariate measure, MAX cross-correlation calculates the linear association between two signals, which also yields fixed delays between two spatially distant EEG signals to accommodate potential signal propagation. This measure can also give us a similarity between two different channels.

Given an EEG signal containing N channels, one can compute the cross-correlation on $\frac{N \times (N-1)}{2}$ pairs of channels (e.g. 15 pairs for N=6 for the employed iEEG database). Calculating the MAX cross correlation for six channels and six sub-bands results in a 90-D vector.

3.5.3.15 Autoregressive (AR) model

A sequence of observations ordered in time or space is called time-series and, in the electrical engineering context, is titled as signal. An AR model can be described by modeling the existing value of the variable as a weighted sum of its own preceding values. Similarly, we can employ this concept to forecast the future based on the past behavior [137], [139], [140]. An AR model with order p can be described as the following formula:

$$y_t = \beta_1 y_{t-1} + \beta_2 y_{t-2} + \dots + \beta_p y_{t-p} + \varepsilon_t \quad (3-13)$$

where ε_t is the error term, usually specified as white noise and $\beta = (\beta_1, \beta_2 \dots \beta_p)$ is the AR coefficient. For a first order, an AR(1) model can be expressed as [141]–[143]:

$$y_t = \beta_1 y_{t-1} + \varepsilon_t \quad (3-14)$$

and its process can be considered as a stationary process once $|\beta_1| < 1$. The coefficient and the error term have been considered as features in the prediction of the seizure [144]–[146].

3.5 Conclusion

In this chapter, the patient database, the preprocessing, and the feature engineering section were introduced. We explained the use of filters for EEG and ECG signals and noises to be removed. Also, in this chapter, we introduced different features of EEG signal that can be considered in seizure prediction. In other words, we transformed the EEG samples to a proper feature space. Then, these features provide some measurements from the signal and can be fed to a classifier as feature vectors. The feature extraction part will be used as an input to the classifier. In the next chapter, we will employ the inputs provided in this section in different classifiers and then will investigate the results.

Chapter 4

Preliminary results for prediction and classification

4.1 Classification

4.1.1 Support vector machine (SVM)

Support Vector Machine (SVM) is a supervised learning model developed by Vladimir Vapnik who aimed to analyze data by employing the classification or regression approach [147]. For a binary classification, SVM maps an input data point X to an output class value y of +1 or -1. For a simple case where just two feature values X_1, X_2 are taken into consideration, we have $(W_1X_1+W_2X_2+b > 0$ or $W_1X_1+W_2X_2+b < 0)$ as in Figure 4-1. Here, $f(X, W, b) = W^T X + b$, X is the input pattern space (R^d), W_i the weight parameters of the input point's features, and b a bias term. The two classes are separated with a hyperplane ($W_1X_1+W_2X_2+b = 0$).

In fact, the SVM algorithm is looking for the points closest to the line (or hyperplane) from both classes and the points are named support vectors. SVM computes the distance between the support vectors and the hyperplane, i.e., the margin. In other words, the margin is the distance that SVM can go from the decision boundary perpendicular to the nearest data point. It can be expressed as [148]–[150]:

- An optimal hyperplane is chosen based on the maximum margin which means maximizing $|d|$ (Figure 4-2), where $d = \frac{W^T X + b}{\|W\|}$. So, in order to maximize d , we need to minimize W .
- If our data point is labelled as $y = +1$, then $W^T X + b \geq 1$
- and if $y = -1$, then $W^T X + b \leq -1$

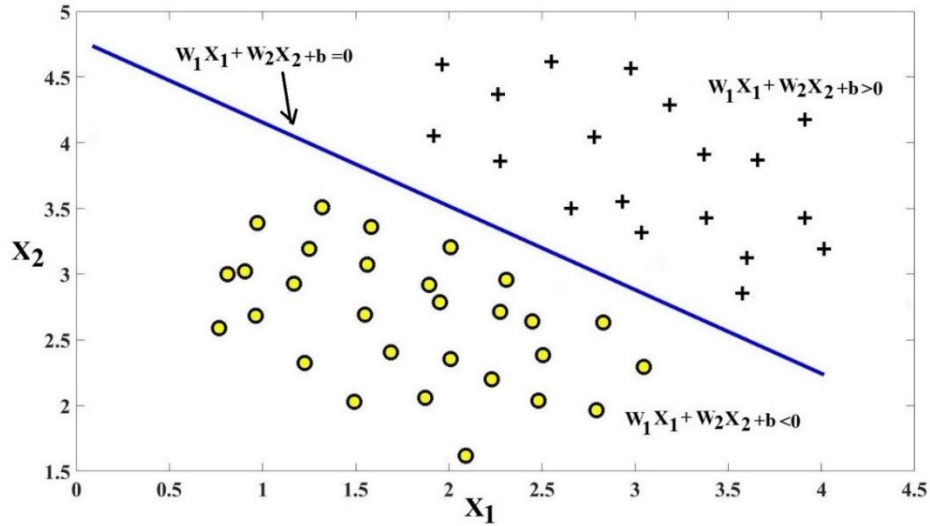


Figure 4-1: An illustration of a simple binary classification problem with two informative features.

It can be thus summarized as:

- For hard margin : $y (W^T X + b) \geq 1$
- For soft margin, allow some margin violation: $y (W^T X + b) \geq 1 - \epsilon$, where ϵ is a non-negative slack variable that allows points, support vectors, to be on the wrong side of their soft margin and decision boundary (Fig 4.3 and Fig 4.4).

Based on the primal form of the SVM:

- Minimize $\frac{1}{2} \|W\|^2 + C \sum_{i=1}^n \epsilon_i$ subject to: $y_i (W^T X_i + b) \geq 1 - \epsilon_i$, $\epsilon_i \geq 0$, $i = 1 \dots n$, where C is a regularization parameter that regulates the amount of misclassification. The larger the C value, the smaller regularization will apply. On the other hand, a small value of C employs more regularization that encourages the classifier to find a large margin on decision boundary. In other words, the smaller C (the stronger the regularization), the smaller the coefficients in the model and as a result, too many variables can be eliminated or too much coefficients can be shrunk to weaken the model [148]–[150].

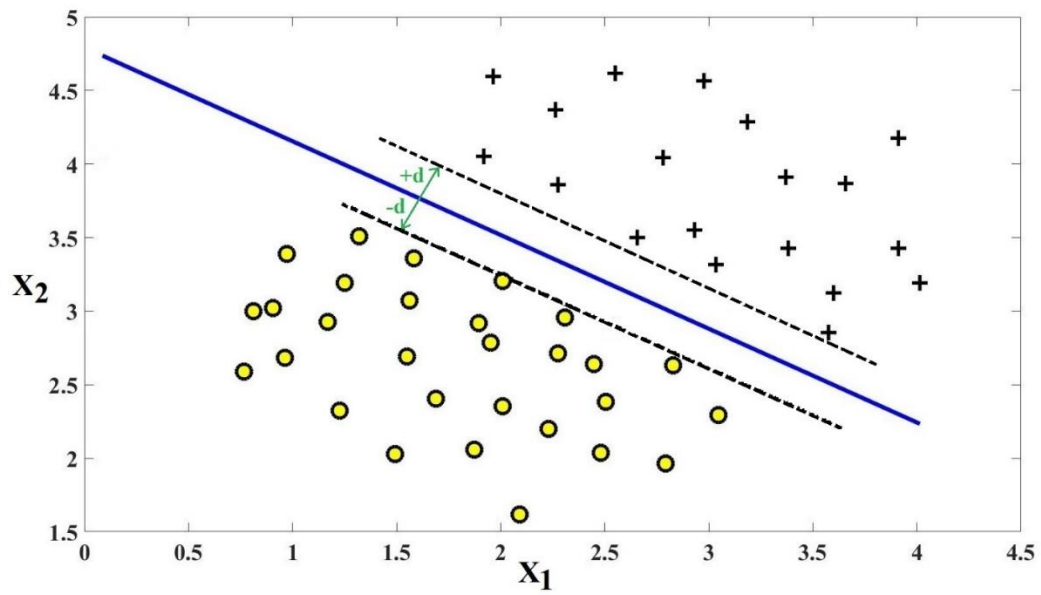


Figure 4-2: Hyperplane separating the two classes.

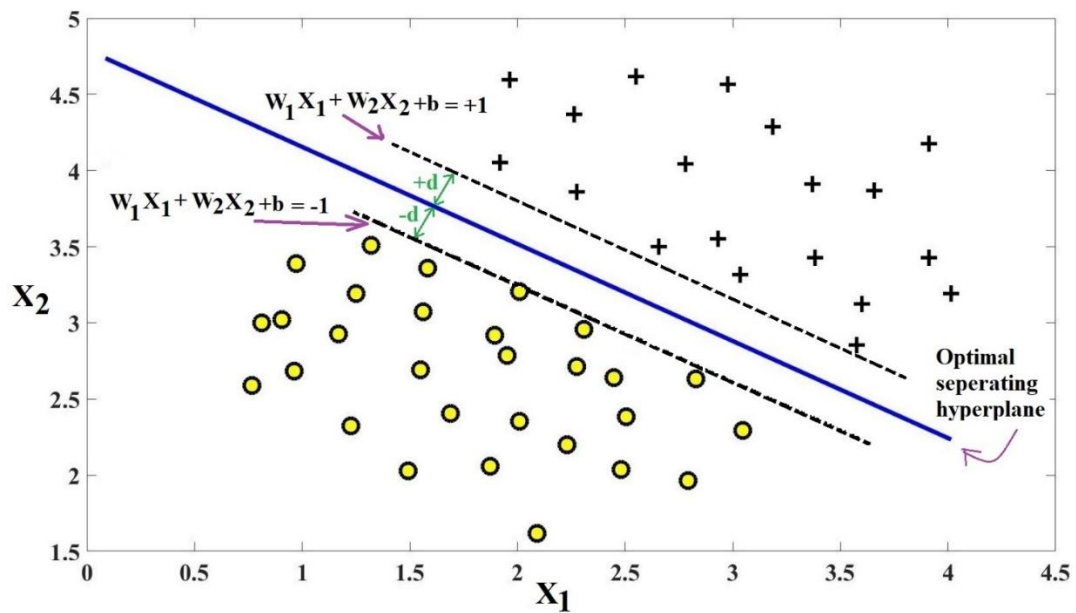


Figure 4-3: Hard Margin Classification.

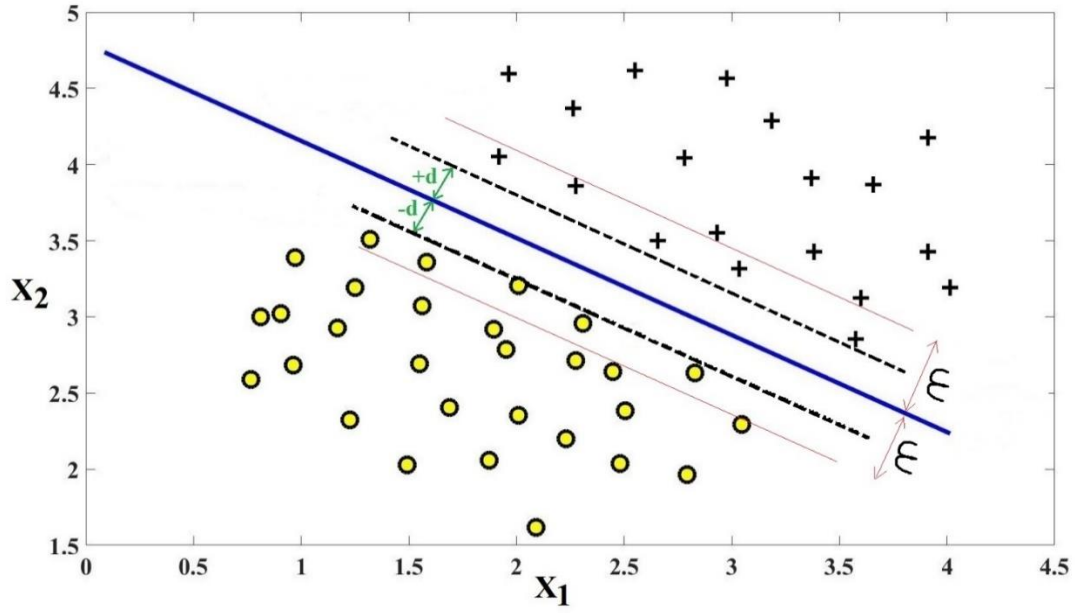


Figure 4-4: Soft Margin Classification makes a classifier more robust.

In this study, we applied the grid search to tune the cost value C in the training phase and to find the optimal parameter by using the Precision and Recall (PR) measure. In fact, based on the PR curve, we considered the higher F-score as the best result. F-score is the combination of precision and recall into a single performance measure that calculates the harmonic mean of the precision and recalls analysis and a measure of accuracy [151].

4.1.2 Constraints with overfitting

As explained in the previous sections, we considered binary classification, so that the proposed algorithm can be used effectively. SVM can exhibit high accuracy with the capability to handle high dimensional data sequences. Note that the SVM target is to find a mapping of the data into a new space where a linear hyperplane was used to separate the two classes (Figure 4-4).

Then, we considered in this study two main classes: “0” (the seizure-free period) and “1” (the seizure period). Since seizures are infrequent events, then we have access to a large amount of interictal data, while few preictal data that contain seizures are accessible. The asymmetric nature of the seizure set is typically named as an ‘imbalanced classification’ in analyzing the data. It should be noted that the ratio between interictal (105028) and ictal (2912) samples is 36:1, which shows a significant imbalance dataset. To overcome the overfitting problem (the model doesn’t generalize well for unseen data), dual cross-validation with regularization was applied in this study [152].

4.1.3 Double cross-validation

We used two approaches for validation namely, ‘Hold-out’ and ‘k-fold cross-validation’.

A. Hold-out

When the data are separated into two sets ‘training’ and ‘testing’, we have the called hold-out validation. The training set is used to train the model and, then, the model will be evaluated with unseen data, i.e., test set. We considered 70% of the data for training and the remaining 30% were employed for testing.

B. K-fold Cross-validation

To apply ‘k-fold cross-validation’ the dataset is first divided into ‘k’ folds. Each time, only one fold is employed for testing and the others are employed for training the model. Then, the model is trained on the training set and evaluated by the validating set. This process is repeated until each distinct fold is used as a validation set. For this study, we employed 10-fold cross-validation and ran each cross-validation experiment 10 times. Effectively, we run each model 100 times. Finally, the average of the obtained accuracy was considered [153].

In a nutshell, data were divided into two groups: 70% of them allocated for training and validating, and the remaining 30% (never-seen-before) for testing the model. Both training and testing data

segments have been numbered sequentially. Note that the validation set was used for estimating how precisely the classifier performs after being trained with the learning set, containing testing for over-fitting of the model. When the classifier was fully optimized with the learning and validation sets, via 10-fold cross-validation, it was applied to the testing set to evaluate the final performance. Figure 4-5 describes the double cross-validation approach.

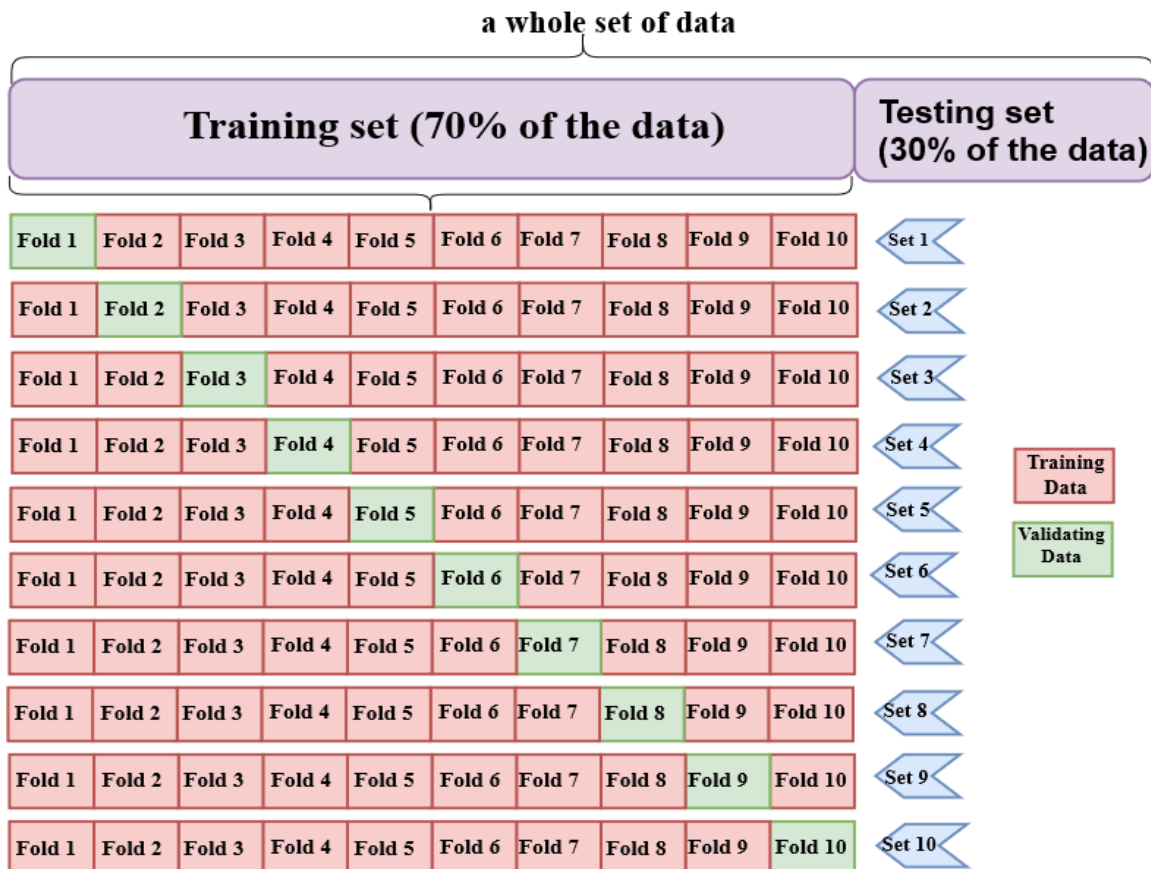


Figure 4-5 An illustration of the double cross-validation on the data set.

4.1.4 Performance analysis

The performance of a classifier must be generalized, i.e., it should perform well when submitted to data outside the training set. Since the ratio between the two classes is very high $105028/2912=36$ (strong imbalance data), accuracy is not the main criteria to use to assess the performance of

the classifier. Although accuracy remains the most intuitive performance metric, it is basically a ratio of correctly predicted observation over the total observations, so reliable only when we have balanced datasets.

The most common metrics employed for analysis are Sensitivity and Specificity and both of them focusing on the predictive model's ability to categorize the preictal class correctly [127]. Since in this work, it is critical not to skip seizure events (preictal), we intended to maximize the True Positive Rate (TPR) or Recall. Therefore, Sensitivity, which measures the percentage of the preictal class predictions that were correctly recognized, and the false positive rate, were measured to evaluate how effectively the proposed algorithm performs.

$$\text{Sensitivity} = \text{True Positive} / (\text{True Positive} + \text{False Negative}) \quad (5-7)$$

In our test, we considered zero false-positive rates, which means a specificity of 100%. In fact, because specificity means if the test result shows a flag or an alarm, we will be confident we found out the seizure onset. In other words, it is the probability of being test negative when there is no seizure.

$$\text{Specificity} = \text{True Negative} / (\text{False Positive} + \text{True Negative}) \quad (5-8)$$

As already noted, these results should be relativized because of the limited number of patients available in the used database.

4.1.5 Results for SVM

In this work, we considered the participants who have the same origin of the seizure, temporal lobe, and the proposed algorithm is summarized in Figure 4-6. Our aim here is to use the detection information and to show that we can predict the seizure.

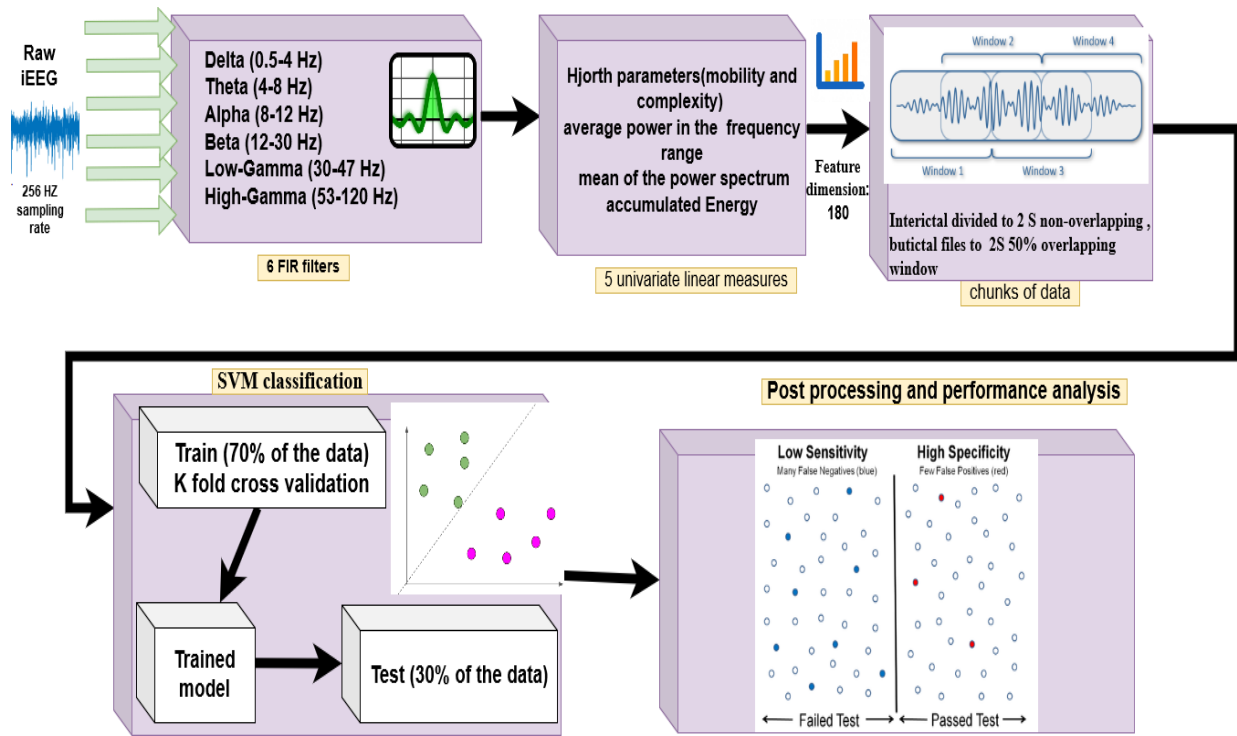


Figure 4-6. Flowchart of the proposed algorithm for seizure prediction.

To build a high dimensional matrix as an input to the classifier, we first considered all the possible features that are helpful in prediction of a seizure and then, let the classifier decide which of the weight of the features are less relevant in the model for forecasting, in order to set them to zero.

This allowed us to build 180 dimensions in the input space of each window (5 features, 6 electrodes, and power in the 6 sub-bands per electrode).

For the first group of patients, we considered the EEG dataset of six patients namely participants #2, #4, #7, #10, #12, and #16 with temporal seizure and 34 seizures of 154 hour-long interictal and preictal signals. The results of the detection and prediction of the algorithm are summarized in Tables 4-1 and 4-2, respectively. In both tables, we applied a 50% overlapping window for the files containing the ictal state and non-overlapping window for other files. These tables demonstrate that by decreasing the size of the window, the rate of detection remains constant, but the rate of prediction is dramatically improved.

Figures 4-7 and 4-8 illustrate the overlapping and non-overlapping windows used in this study. In the non-overlapping window, the number of windows is smaller and we have few samples for processing compared to the overlapping one. The feature extraction methods can be described as follows: An hour EEG signal was divided into various non-overlapping window sizes (20, 10, 4, 2 and 1 second) for interictal section, while the ictal section was split into chunks of same window sizes (i.e. from 20 seconds to 1 second) with 0%, 50%, and 75% overlapping.

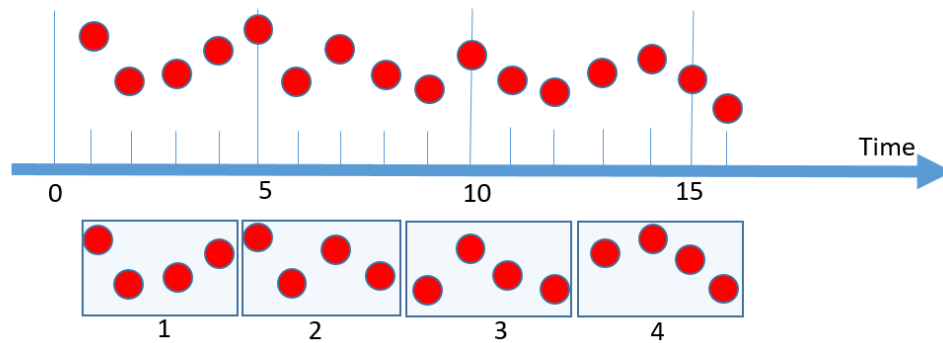


Figure 4-7 Non-overlapping 4 seconds window for extracting the feature (adapted from [152]).

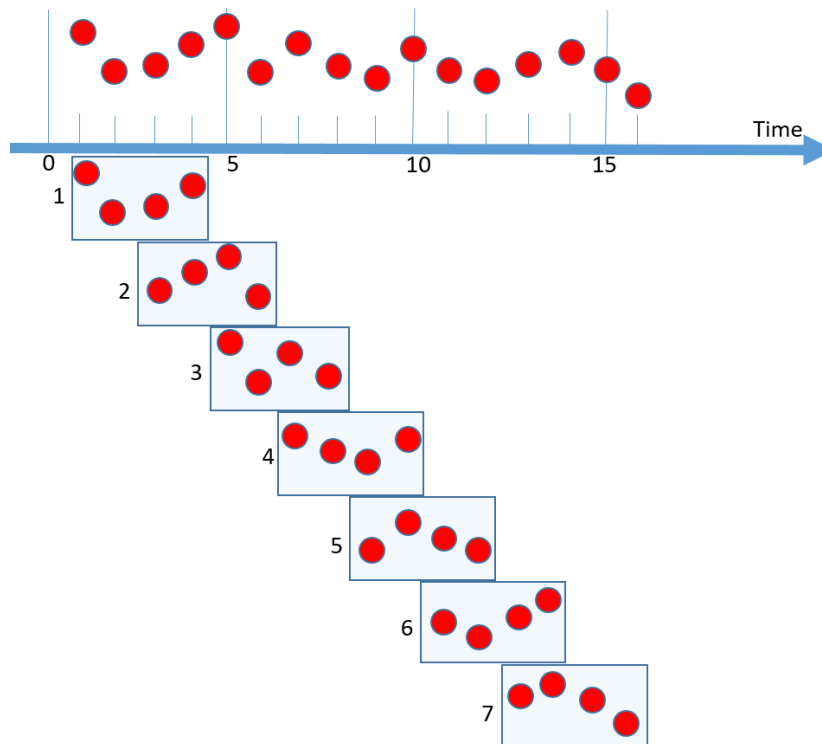


Figure 4-8. A 4 seconds window with 50% overlapping for extracting the feature.

Figures 4-9 to 4-13 show the probability of seizure happening for 20, 10, 4, 2, and 1 seconds, respectively. In all cases, we considered a non-overlapping window dividing the interictal files into different segments and a 50% overlapping window chunking the preictal files to various segments. The red lines in the graphs show the period of the seizure that has the highest peak in each graph.

The results of changing the window size clearly show that by decreasing the size of the window, the probability of seizure detection will be enhanced. Also, we found out that by decreasing the window size, the predictive model can detect some peaks before the onset of the seizure. Therefore, the lower the window size, the higher the probability of predicting the seizure. However, this trend does not continue up to 1-second window. While we spend nearly twice the time for computation, the prediction result does not improve. The accuracy of the classification for 2-second window reported 97.5%, while for 1-second window is 97.49%. Means by decreasing the window size we pay more for calculation but received quite less improvement.

It is worth mentioning that the two-second window was selected based on different empirical experiments from twenty to one second. Furthermore, based on the works of Islam [119] by choosing a very short window, e.g., less than one second, the seizure waveform may not be recognized properly in such a short duration. On the other hand, by defining a longer window length, e.g., more than 5 seconds, artifacts tend to be more transient than a seizure. Therefore, artifacts and seizure cannot be distinguished from each other. In conclusion, a non-stationary signal like EEG can be assumed as a stationary signal in a short duration epoch like a two-second window.

Table 4-1 Detection Results for Different window size.

measure\window size	20s	10s	4s	2s	1s
Sensitivity	83%	83%	83%	83%	83%

Table 4-2 Prediction Results for Different window size

measure\window size	20s	10s	4s	2s	1s
Sensitivity	16%	20%	50%	75%	75%

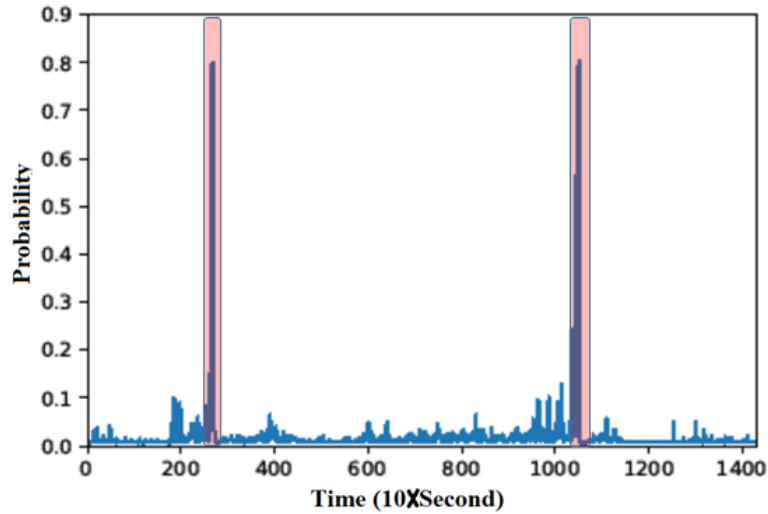


Figure 4-9. Probability of seizure happening for patient #16 with a window of 20 seconds. Red bar shows the duration of the seizure.

Table 4-3 shows the comparison of various overlapping windows ranging from 0 to 75%. The results clearly show that the best choice is 50% overlapping, and we can predict most of the seizures with a window of 2 seconds and an overlapping of 50%. As explained earlier accuracy cannot be a good measurement here due to having an imbalanced data.

Table 4-3. The results for 2s-window with different overlapping window

measures\	non-	25%	50%	75%
overlapping	overlapping	overlapping	overlapping	overlapping
Sensitivity	50%	41%	75%	58%
Accuracy	98.5%	97.9%	97.5%	96.6%

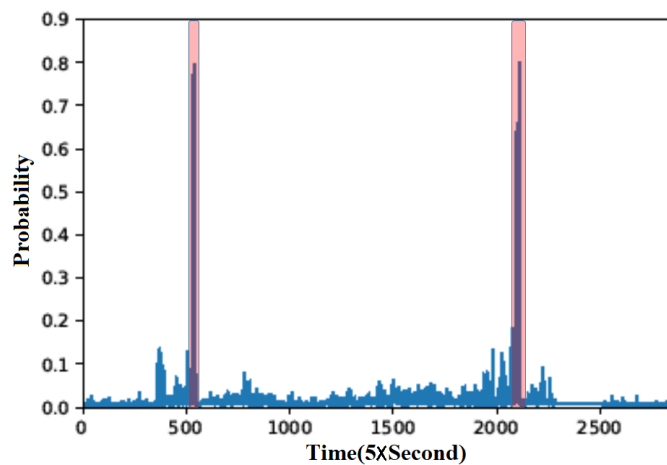


Figure 4-10. Probability of seizure happening for patient #16 with a window of 10 seconds. Red bar displays the duration of the seizure.

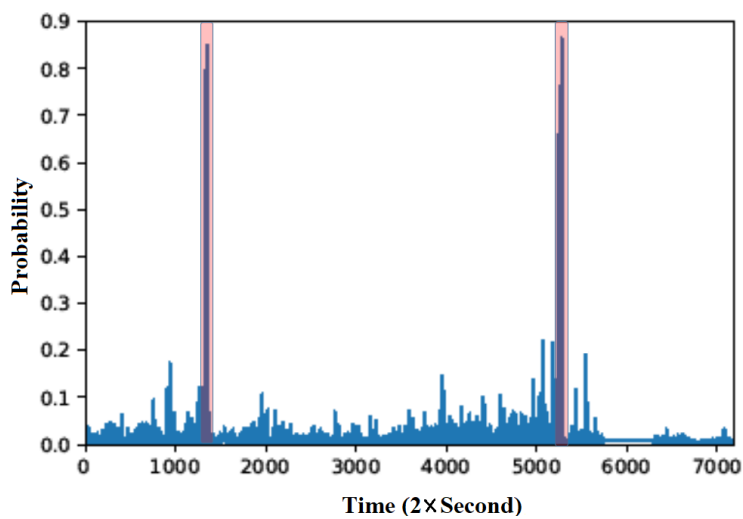


Figure 4-11. Probability of seizure happening for patient #16 with a window of 4 seconds. Red bar displays the duration of the seizure.

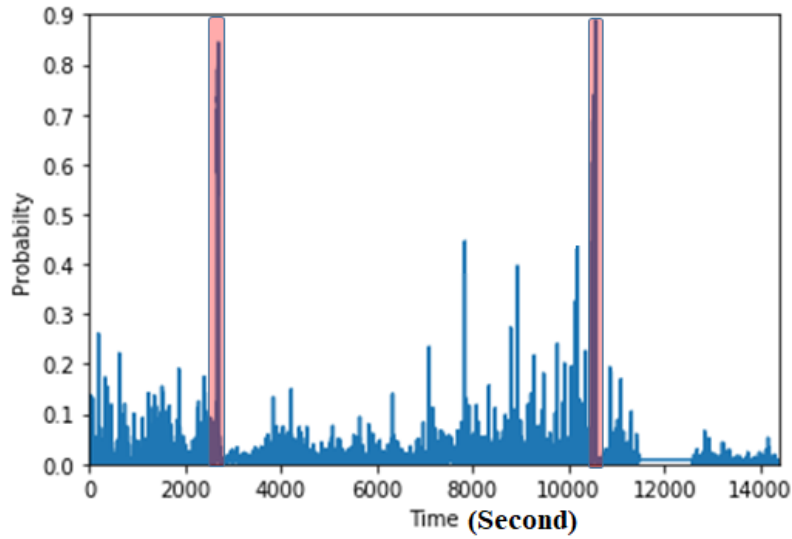


Figure 4-12 Probability of seizure happening for patient #16 with a window of 2 seconds. Red bar indicates the duration of the seizure.

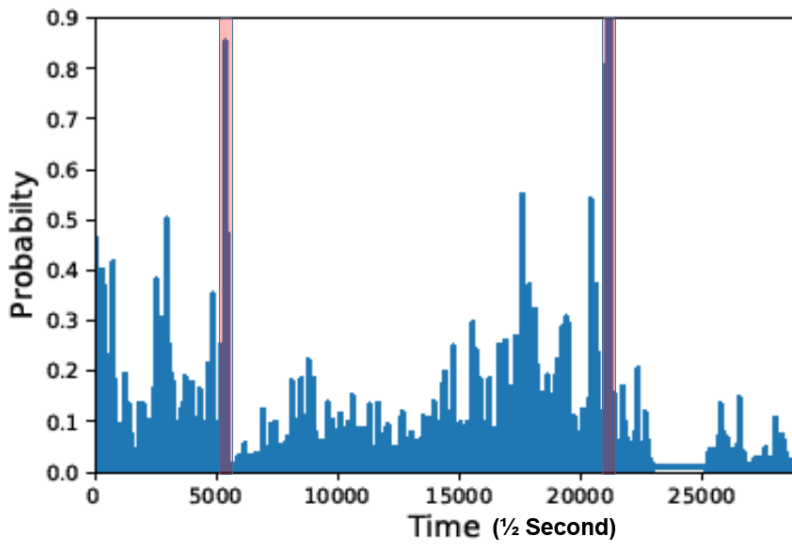


Figure 4-13 Probability of seizure happening for patient #16 with a window of 1 second. Red bar displays the duration of the seizure.

Figure 4-14 shows the results of the test data on the model for patient #16 (file #48), taken as illustration, with a window size of 2 seconds. In this figure, line B indicates the seizure happening (very good detection) and the point A, a peak of about 0.45 probability, which appears before the

seizure, can be considered as a prediction alarm. If we assign a threshold of about 0.45, we can warn the patient that a seizure is coming (about 45 minutes before the event).

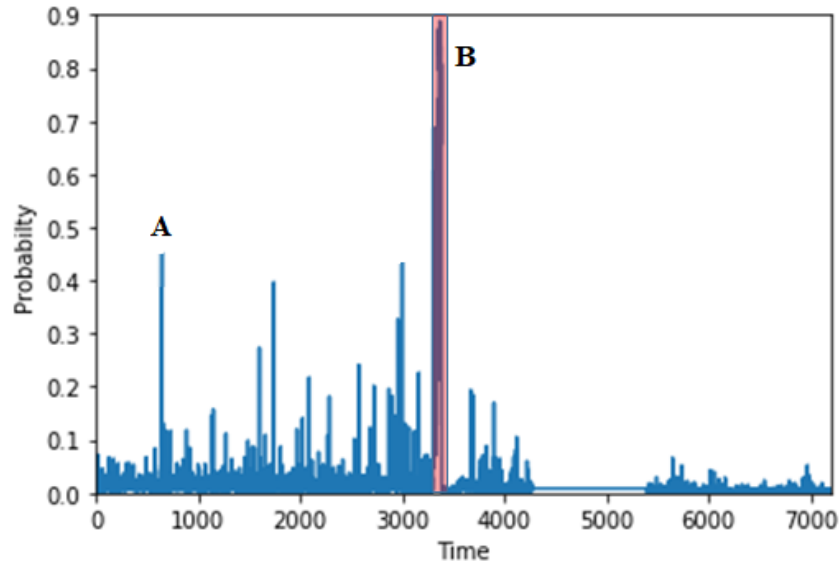


Figure 4-14 Probability of seizure happening for patient #16 file #48 with window 50% overlapping 2 seconds. Red bar shows the duration of the seizure.

Successive to learning data using the L1-regularized SVM, Platt's scaling [71], which employs logistic regression on the SVM scores, was used to extract the probabilities of each class, i.e. seizure vs. non-seizure. It should be noted that Platt scaling is an approach of converting the output of classification into a probability distribution. In other words, the output of binary 0 or 1 is converted into probability (values between 0 and 1).

In the next step, the mean probability of the non-seizure segments for each participant was created. Then, in order to predict an impending incident, the following simple model was retained:

$$\tau = c \cdot \mu_{Pns} \quad (5-9)$$

where τ is the threshold used to predict a future seizure, c is a constant empirically set to for each patient, and μ_{Pns} is the mean of the probability of the non-seizure (ns) segments for that particular subject.

For a specific patient input iEEG, if the output probability obtained using the trained iEEG segment is above τ , an impending seizure is reported. We applied this rule to our dataset and obtained the time (in minutes) expected to predict seizures before their occurrence. Table 4-4 states the parameters, c and μ_{Pns} , of this approach for the 6 participants in this study. For patient #16 we had two seizures for prediction and we could find an alarm for each seizure, one 6 minutes and the other 45 minutes before the seizure onset. The average time for prediction of this approach as well as the average of the probability of the non-seizure period is inserted in Table 4-5. We will elaborate on this approach for the rest of the database and try to increase the performance of the classifier by employing different strategies, as detailed in the next few sections.

Table 4-4. The results for 2 seconds window with 50% overlapping window.

	Pa#2	Pa#4	Pa#7	Pa#10	Pa#12	Pa#16
Time of prediction (min.)	59.6	80	5	44	22	6&45
Probability mean of non-seizure (μ_{Pns})	0.0096	0.0052	0.0079	0.0076	0.0187	0.0042
constant empirically (c)	41.6	76.9	50.6	52.6	21.39	95.2

Table 4-5. The overall results for 2 seconds window with 50% overlapping window.

Average of time prediction	37.4 minutes
The average probability of the non-seizure	0.0089

We were remarking that the conventional SVM with the best window size predicted 7 of 9 seizures, with a sensitivity rate of 78% and specificity of 100%. However, we were not able to predict 2 seizures because of the inadequate preictal database availability. Based on Ramachandran *et al.* [15], we might need at least 3 to 16 hours before the onset of the seizure to predict seizures. Another possible solution is employing non-linear measures such as phase synchronization to forecast unpredicted seizures.

In the above results, it is worth mentioning that the threshold was assigned in a way that there is no false alarm, but after considering the threshold, we found out very promising results.

The testing results based on the trained model and with the assumption that we “need at least 3 hours data to claim that there is an alarm for a seizure,” we concluded the following:

- In this work, we have 9 seizures for testing from 6 patients.
- We did not consider two seizures due to having just less than 2 hours preictal stage, and by discarding those two seizures, we could successfully predict the whole 9 remaining seizures.
- Besides linear kernels, we also investigated non-linear kernels such as Polynomial and Gaussian. However, due to taking a huge amount of time for the grid search in Python for tuning the hyper-parameters, we did not include them in this study.

4.2.1 Other Classifiers

We will employ various classifiers and compare their results in the next chapters.

4.2.2 Neural Networks (Multi-Layer Perceptron: MLP)

The early works on neural networks go back to the 50s and 60s. Recently, they have received huge interest as deep learning has given remarkable results [149]. Figure 4-15 shows an illustration of a binary classification scenario for a feed-forward neural network, entitled a Multi-Layer Perceptron (MLP).

In this method, a hidden layer is represented by an additional set of boxes (h_0 , h_1 , and h_2 in the figure). These boxes, hidden units, compute a nonlinear function of the weighted sums of the input features. Then, the results will be the intermediate output values namely, v_0 , v_1 , v_2 . After that, the neural network, usually a MLP (Multi-Layer Perceptron) network calculates a weighted sum of

these hidden unit outputs, to create the final output value, $\hat{Y}$. The nonlinear function employed in the hidden unit is named the activation function. In this figure, the activation function used is “Relu”, which stands for the Rectified Linear Unit and is a half rectified function.

Various activation functions are depicted in Figure 4-16. Nowadays, the Relu is one of the most used activation function [154], [155].

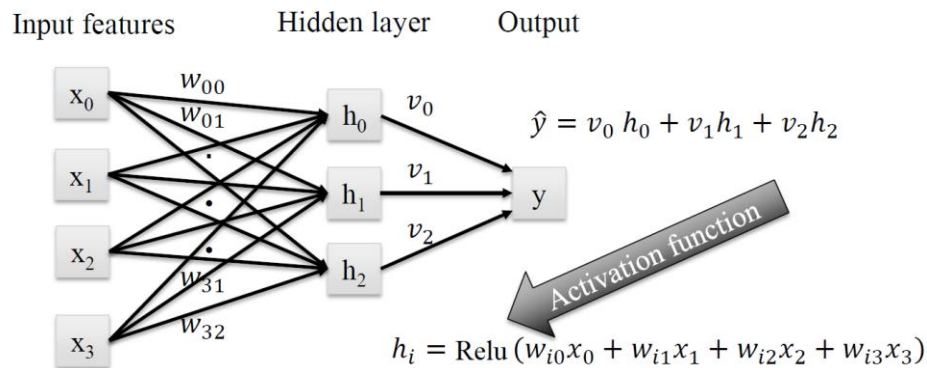


Figure 4-15 Multi-layer perceptron with one hidden layer with a Relu activation function [149].

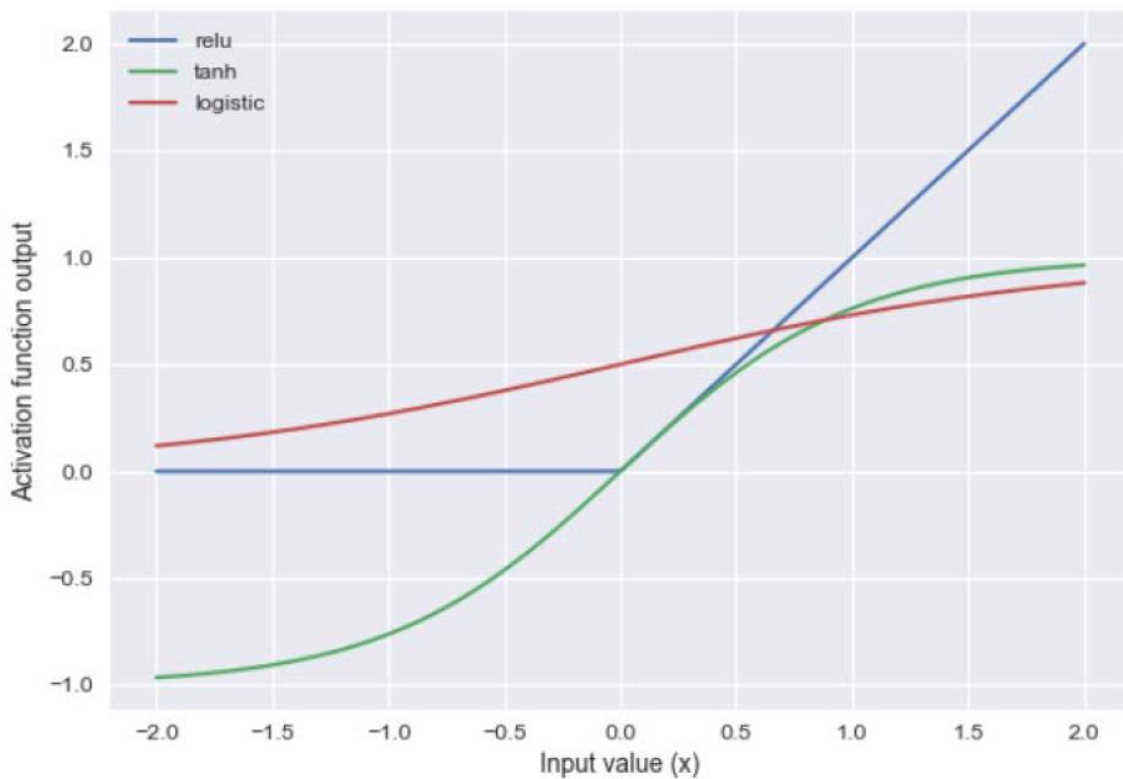


Figure 4-16. Illustration of different activation functions [149].

Indeed, the combination of the non-linear activation function with the weighted sum of the inputs enables MLP to learn more complicated models and allows neural networks to perform a more accurate prediction. The regularization parameter for MLPs is entitled alpha, which controls the model complexity by constraining the magnitude of model weights, as explained in SVM classifier [149].

4.2.3 Random Forest

Random Forest is one of the most effective methods in machine learning, consisting of creating models so-called as an ensemble [156]. An ensemble is a technique that combines several base models in order to create an optimal predictive model that is more powerful than any of its individual learning models alone.

After training the model, the model tries to predict every tree in the forest, then combine individual predictions based on a weighted vote. This means that each tree gives a probability for each possible target class label, then the probability for each class is averaged across all the trees and the class with the highest probability is the final predicted class [149].

Various parameters need to be set or tuned to build a random forest model such as:

- the number of decision trees to be built in a random forest,
- the maximum number of features permitted for the split in each decision tree,
- the depth of the trees.

4.2.4 XGBOOST

XGBoost stands for eXtreme Gradient Boosting which is the fast implementation of gradient boosting. The gradient descent algorithm accelerate the convergence by applying second order partial derivative of the loss function and Boosting basically means enhancing the performance.

Boosting is a sequential ensemble learning method to adapt a series of weak base learners to strong learner to increase the performance of the model [157]–[163]. It has much faster computation than the gradient boosting algorithms since it employs parallel processing. To further improve it, the XGBoost classifier has two regularization terms (inbuilt L1 and L2) to penalize the complexity of the model and avoid overfitting [157], [158], [161].

4.2.5 Conclusion

We introduced a probabilistic approach based on SVM and will extend the work for various classifiers and compare their results. Also, we will try to improve the performance classification by applying different strategies in the next chapter. In this work, we focus first on detection of the seizure and then used the information of the detection in prediction of the seizure. In this approach, F- score is low and for the rest of the studies the segments before the seizure along with the ictal stage were deemed as class one (1) the interictal with the postictal considered as zero (0) class.

Chapter 5

Comparative study of feature selection techniques

5.1 Methodology

To select the suitable features for predicting seizures, the proposed method steps are illustrated in Figure 5-1. In the first phase, six EEG signals are preprocessed and 16 features extracted, i.e., the first 14 mentioned in chapter 4 (time domain) plus 2 from the 15th one (i.e., coefficient and error term of the AR model). Then, the data are divided into train and test sets and three kinds of feature selection methods employed to reduce the data dimension, making the approach computationally efficient. Next, the obtained results are tested by a well-known judging classifier namely, Random Forest. The 30 top extracted important features are ranked by various methods and fetched into the Random Forest classifier, while the accuracy and the Mathew's correlation coefficient (MCC) are used as a performance analysis. Finally, the relevant features among the 30 top ones are retained for the winner feature selection method.

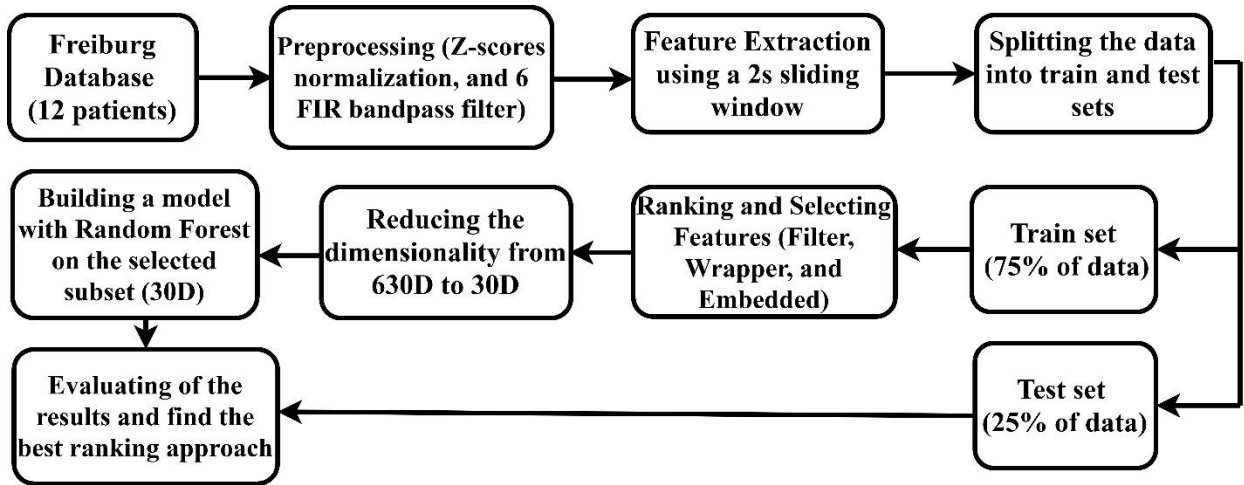


Figure 5-1. An overview of the proposed work in analyzing the discriminative features.

In this work, we divided the participants into two groups based on the origin of their seizures. We retained twelve epilepsy patients (six TLE with the hippocampal origin (134 hours) and six FLE with neocortical origin (137 hours)). The TLE patients were #2, #4, #7, #10, #12, and #16 and the FLE patients were #1, #3, #5, #8, #18, and #19 from Table 3-1 of the Freiburg database.

5.2 Feature Extraction

The preliminary stage in EEG signal analysis is the preprocessing. To decrease the effect of factors that cause baseline differences among the different recordings within the dataset and remove the signal DC component, iEEG signals were normalized using Z-scores (expressed in terms of standard deviations from their means). The only potential artifact that could be addressed was the harmonic power line interference at 50 Hz. We eliminated the 50 Hz interference indirectly by performing the sub-band filtering in which Gamma was divided into two sub-bands.

To deal with imbalanced dataset, an hour EEG signal was divided into 2 seconds non-overlapping windows for interictal section, while the ictal section was split into chunks of same window sizes with 50% overlapping. Various univariate linear measures were extracted at each epoch of two seconds along with a bivariate linear measure, as reported in Table 5-1 [47], [135], [137], [138], [147].

Prior to feature extraction, we utilized six band-pass FIR (Finite Impulse Response) filters to divide the iEEG signals into different frequency bands: Delta (0.5-4 Hz), Theta (4-8 Hz), Alpha (8-12 Hz), Beta (12-30 Hz), as well as two Gamma bands namely, low-Gamma (30-47 Hz), and high-Gamma (53-120 Hz) [125], [127], [136]. This led to an input space of 630 dimensions for each window. The first 13 univariate features make dimensions of $6 \times 6 = 36$, the bivariate feature, MAX cross correlation generates $6 \times 15 = 90$ and finally the AR parameters create a $2 \times 6 \times 6 = 72$.

Table 5-1. Features extracted during a 2 seconds sliding window.

NO.	Features	comments	Time of the computation for each channel (s)
1	Energy	One feature (36D) ¹	0.001593
2	Mean	One feature (36D)	0.018089
3	Variance	One feature (36D)	0.011741
4	Skewness	One feature (36D)	0.181114
5	Kurtosis	One feature (36D)	0.014463
6	Interquartile range	One feature (36D)	0.036152
7	Zero Crossing Rate	One feature (36D)	0.007410
8	Mean Absolute Deviation	One feature (36D)	0.010459
9	Entropy	One feature (36D)	1.168083
10	Hjorth mobility	One feature (36D)	0.007360
11	Hjorth complexity	One feature (36D)	0.003935
12	Coefficient of Variation	One feature (36D)	0.010528
13	Root Mean Square	One feature (36D)	0.171703
14	MAX of cross correlation	15 values for 6 channels, but we consider just one feature (90D)	0.087968 (between two channels)
15	AR model ²	Two features (72 D)	0.026372

¹ 36-dimensional (36D) feature vector. ² (coefficient and an error term).

5.3 Feature selection

Feature selection is a vital stage in analyzing the data to improve model performance and reduce mathematical computational complexity by projecting the existing features onto a lower dimensional space. This technique reduces the input dimensionality by removing irrelevant or redundant features from the entire feature set [35], [164]–[168]. In the machine learning literature, the approaches to feature subset selection are often categorized as filter, wrapper, and embedded strategies [165], [167]–[169].

Filter approaches are based on the statistical properties of explanatory variables (predictor variables) and their relationship to the outcome variable (response); basically, they are not computationally expensive. There are a great deal of filter methods such as PCA (Principal

Component Analysis), LDA (Linear Discriminant Analysis), and PLS (Partial Least Squares) which all of them are finding the linear combination of features to characterize two or more classes. However they are linear, simple, and relatively low cost to reduce the dimensionality of the data, we do not have any interpretation of the ranked features. Moreover, PCA as a famous feature reduction method is an unsupervised method which does not consider the dependant variable [170]–[172]. We employed the Kruskal Wallis (KW), a nonparametric test, without making prior assumptions about the data distribution, unlike the One Way ANOVA [173], [174].

The value of Kruskal-Wallis ranking can be calculated as the following equation:

$$G = \left[\frac{12}{N(N-1)} \sum_{i=1}^c \frac{R_i^2}{n_i} \right] - 3(N+1). \quad (5-1)$$

Here, N is the total number of observations across all classes, n_i is defined as the number of observations in group i , R_i is the mean rank of the group i , c is the number of output group [174].

Wrapper approaches try to find a predictive model by using various combinations of features, then select the set of features that offer the highest evaluation performance. These techniques can be time consuming and tend to be slow. Therefore, they are not appropriate for large-scale problems to select the subset of features. We used one of the most popular wrapper techniques, Support Vector Machine - Recursive Feature Elimination (SVM-RFE), which backward eliminate features [175], [176]. The backward elimination technique builds a model on the entire set of the all features and computes an importance score for each one. Then it removes the least significant features at each iteration which enhances the performance of the model. In other words, the top ranked variables are eliminated last [175]–[177].

Embedded techniques are inbuilt feature selection, allowing a classifier to build a model that automatically performs attribute selection as a part of model training (performs feature selection and model fitting simultaneously) [178], [179]. In this work we used XGBoost (Extreme Gradient Boosting) which has been broadly employed in many areas due to its parallel processing, high scalability and flexibility [157]–[161], [180]. This embedded technique is an optimized implementation of the Gradient Boosting framework. Boosting is building a strong learner with higher accuracy with combination of weaker classifiers and it is known as the Gradient Boosting once the weak classifiers in each phase are built based on the gradient descent to optimize the loss

function [157]–[161]. To further improve it, the XGBoost classifier has two regularization terms (inbuilt L1 and L2) to penalize the complexity of the model and avoid overfitting [157], [158], [161]. Interestingly, one of the reasons XGBoost outperforms other embedded feature selection methods such as Lasso as explained earlier XGBoost employs two regularization terms to combat overfitting problem. While Lasso is a linear approach which employs just one regularization parameter (L1) [181].

5.4 Evaluation and Performance analysis

We deemed 30 features that hold most informative and discriminative information. We extracted them from three approaches and then applied each new group of features to one of the most powerful ensemble method, Random Forest. This embedded approach belongs to a bagging for judging the performance of the other attribute selectors and it differs from boosting mechanism [162], [163], [178]. We selected bagging rather than boosting to avoid systematic bias in the comparison results.

A classifier must generalize, i.e., it should perform well when submitted to data outside the training set. Since we are dealing with the problem of class imbalance, accuracy seems inadequate metric to evaluate the performance of the classifier [182]–[185]. Although accuracy remains the most intuitive performance measure, it is simply a ratio of correctly predicted observations over the total observations, so reliable only when a dataset is symmetrical. However, this measure has been used exclusively by some researchers in analyzing seizures [52], [174], [175].

Numerous metrics have been developed to analyze the effectiveness and efficiency of the model in handling the imbalanced datasets such as F1 score, Cohen’s kappa, and Matthews’s correlation coefficient (MCC) [184], [185]. Among the above popular metrics, MCC is revealed as a robust and reliable evaluation metric in the binary classification tasks and, in addition, it was claimed that measures like F1 score and Cohen’s kappa should be avoided due to the over-optimism results especially on imbalanced data [176], [177], [184], [185].

To visualize and evaluate the performance of a classifier, we employed the confusion matrix (see Table 5-2, which represents the confusion matrix of a binary classification). After computation of the confusion matrices, we retained accuracy and MCC to compare the classification performance and effectiveness of the feature selection methods.

Table 5-2. The confusion matrix for a binary classification task.

Actual	Predicted	
	Positive	Negative
Positive	TP (True Positive)	FN (False Negative)
Negative	FP (False Positive)	TN (True Negative)

Accuracy (Acc):

Accuracy can be defined as the ratio between the number of correct predictions and the total number of correct predictions. Let TP be actual positives that are correctly predicted positives, TN be actual negatives which are correctly predicted negatives, FP actual negatives that are incorrectly predicted positives and FN actual positives that are incorrectly predicted negatives. Then, accuracy can be expressed as:

$$Acc = \frac{TP+TN}{TP+TN+FP+FN} \quad (6.2)$$

Matthews's Correlation Coefficient (MCC):

MCC takes into account all four quadrants of the confusion matrix, which gives a better summary of the performance of classification algorithms. The MCC can be considered as a discretization of the Pearson's correlation coefficient for two random variables due to taking a possible value in the interval between -1 and 1 [177], [179], [186]. A score of 1 is deemed to be a complete agreement, -1 a perfect misclassification, and 0 indicates that the prediction is no better than random guessing (or the expected value is based on the flipping of a fair coin).

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}} \quad (6.3)$$

5.5 Dividing signals into frequency sub-bands

Our aim is to analyse and rank the time-domain features introduced by different researchers and related to epileptic seizures forecasting with EEG signals by comparing the performance of three feature selection approaches. Before ranking the features and comparing them, we first investigated how much dividing EEG signal into various sub-bands can be important.

Therefore, we compared the accumulated energy for two cases, without and with dividing the signal into 6 sub-bands. The feature selection scores are represented by Figures. 5-2-a and 5-2-b for temporal and frontal lobe, respectively.

The results for both graphs show that dividing the EEG signal into various sub-bands can improve the performance of seizure forecasting because it contains much more discriminative information than the other case. Interestingly, we will increase the dimensionality of data for now but, later, we will focus on specific sub-bands and reduce the dimensional feature space to consume less memory at runtime.

5.6 Feature selection methods comparison

We employed three methods for feature ranking and, afterwards, independent train and test sets were defined to compare their performance using a Random Forest classifier. To have a better estimation of the generalization performance of the work, we evaluated the top 30 selected features on the testing dataset, which has not been used during the training process.

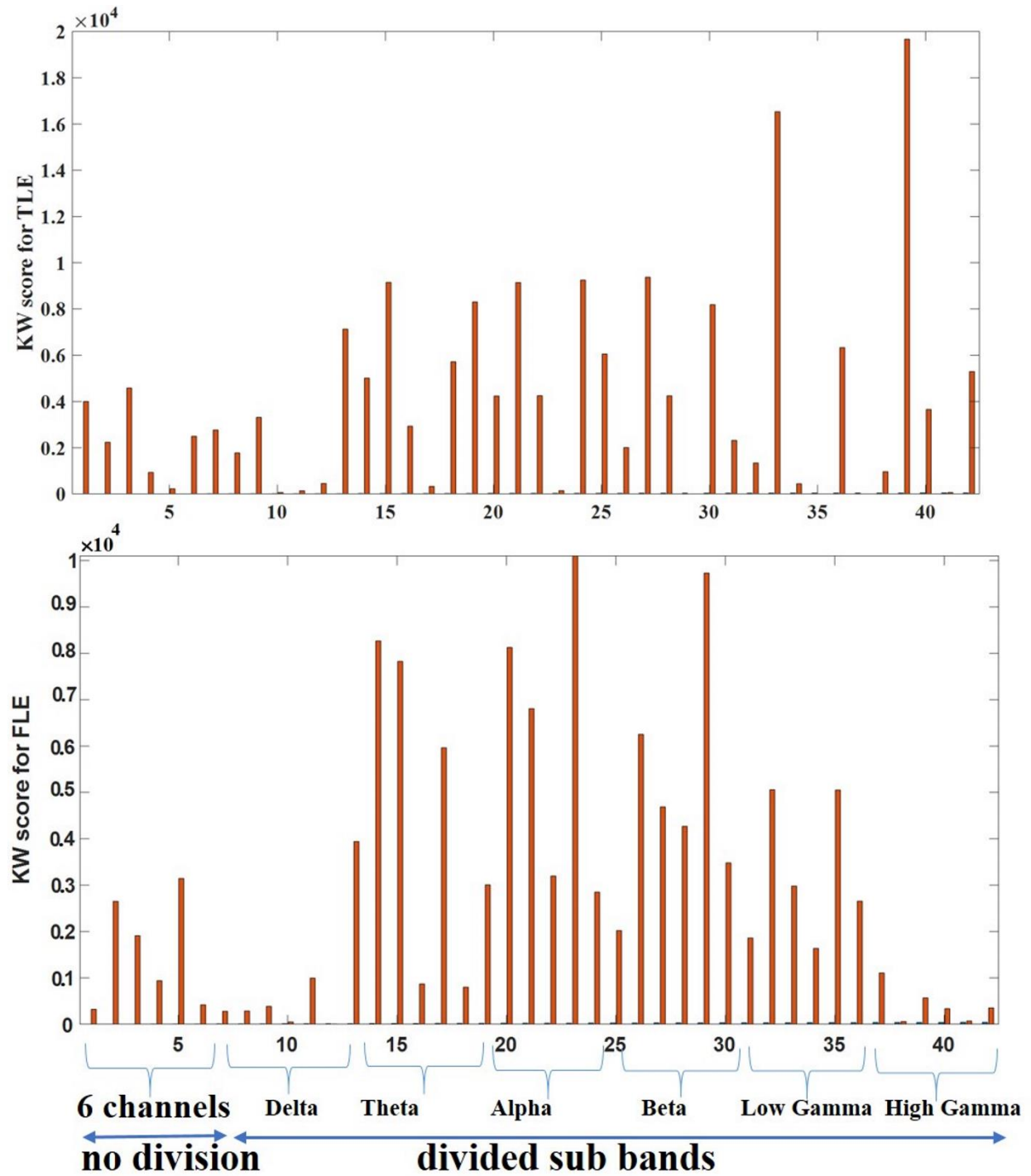


Figure 5-2. Investigation of dividing the EEG signal into various sub-bands a) for temporal lobe b) for frontal lobe.

Using Matlab, the calculations were made on an Intel(R) Core (TM) i7CPU 3.3GHz, and 16 GB RAM. Once the preprocessing stage was done in Matlab, MAT files were converted to NumPy

arrays and we developed the rest of the work in Python (3.7.6) programming language. The computation time for each feature selection method is listed in Table 5-3 and the performance of the various feature selection methods is listed in table 5-4.

Table 5-3. The comparison of computational cost of three feature selection methods applied on the train set.

Brain lobe/Selection method	Kruskal-Wallis	SVM-RFE	XGBOOST
Temporal lobe	12.4 min	3411 Min	18.7 Min
Frontal lobe	10.9 min	2375 Min	12 Min

Table 5-4. The performance of the various feature selection methods for both lobes applied on the test set.

Selection method performance	Temporal Acc. MCC	Frontal lobe Acc. MCC
Kruskal-Wallis	76% 0.55	59% 0.24
SVM-RFE	68% 0.40	55% 0.11
XGBOOST	72% 0.496	57% 0.12

If we compare the results in the two above tables, we can conclude that the computational complexity of filter-based method, Kruskal Wallis, has the highest MMC score and less computation time. While SVM-RFE has a longer computation time compared to the other approaches and shows the poorest performance.

The top 30 ranked subset, are listed in Tables 5-5 and 5-6 for TLE and FLE, respectively, based on the three feature selection approaches. The most popular feature ranked by the three feature selection methods is AR. The second most important feature is the MAX cross correlation and complexity is the next one. Interestingly, features like Mean, Skewness, Zero Crossing Rate, and Entropy are not deemed as top 30 ranked feature-subset as in Tables 5-5 and 5-6. Also the following remarks can be made:

- AR model is an interesting feature along with MAX cross correlation for all three feature selection methods and both lobes.
- Delta sub band is considered an important sub band for XGBOOST and SVM-RFE, but not the case for Kruskal Wallis.

- Error is more important than coefficient as AR parameters in discriminating feature between seizure and non-seizure for all three feature selection methods.

Some of the attributes are selected multiple times by the feature ranking approaches for both tables like AR High-Gamma Error which has been chosen by Kruskal Wallis as feature 1, 7, and 30 in table 5-5. The reason behind that is the feature has been selected and presented without considering the order of the electrodes.

Figures. 5-3-a and 5-3-b illustrate an overview of the features selected with the KW for temporal and frontal lobes, respectively. Figure 5-3-a confirms that, amongst the time-domain parameters that play important role in prediction of the seizure, AR is the top feature, followed by the cross correlation and the IQR. On the other hand, Figure 5-3-b shows that the AR model, a measure of complexity obtained with Hjorth's analysis, is an important feature besides IQR for patients with frontal lobe epilepsy.

In the last step, we applied another filter method to get the product-moment correlation coefficient, or Pearson correlation coefficient, in order to identify the linear relationship between the 30 top ranked features and thus, to eliminate any redundant information. This coefficient can be set as

$$\rho = \frac{Cov(X_1 X_2)}{\sigma_{X_1} \sigma_{X_2}} \quad (6-4)$$

where $Cov(X_1, X_2)$ is the covariance of two features and σ_{X_i} is the standard deviation of each variable. ρ can take a value in the range of $[-1, +1]$ with +1 the case of a perfect positive linear relationship between random variables and -1 a negative linear relationship between two features. That is to say, the larger X_1 values the smaller X_2 values and vice versa. $\rho = 0$ implies the independence between the variables. In other words, the higher absolute value of the correlation coefficient, the more similar they are [187]–[189].

We calculated the product-moment correlation matrix for the top 30 subset of features with the highest Kruskal-Wallis scores as shown in Figures. 5-4-a and 5-4-b (the darker the color, the stronger the correlation between two variables). We can see some features with a perfect positive linear relationship in both figures. We reported the features with the strongest linear pattern in Table 5-7. Interestingly, we can see a good relation between IQR and MAD in both lobes, which

has a strong linear relationship and it happens in Alpha and Beta sub-bands in frontal lobe and in low and high- Gamma sub-bands for temporal lobe. Also we can see a close relationship between variance and energy for both lobes.

Table 5-5. Top 30 feature-subset ranked by three types of approaches for temporal lobe epilepsy.

Top Features	KW	SVM-RFE	XGBOOST
1	AR High-Gamma error	cross correlation Beta	Beta error AR
2	AR Alpha error	AR Delta coefficient	AR Beta error
3	AR Beta coefficient	MAX cross High-Gamma	AR High-Gamma error
4	AR Beta error	MAX cross Beta	AR Low-Gamma error
5	MAX Cross Low-Gamma	MAX cross Beta	AR Theta error
6	MAX cross High-Gamma	MAX cross High-Gamma	AR Delta error
7	AR High-Gamma error	RMS High-Gamma	AR Delta error
8	IQR High-Gamma	AR Theta coefficient	AR Alpha coefficient
9	AR High-Gamma coefficient	AR Theta error	complexity High-Gamma
10	Cross High-Gamma	RMS Beta	AR Delta coefficient
11	IQR High-Gamma	AR Beta coefficient	AR High-Gamma error
12	MAD High-Gamma	cross Beta	AR Delta error
13	AR Beta error	cross Delta	complexity Beta
14	complexity High-Gamma	cross Delta	AR Alpha coefficient
15	Cross Low-Gamma	STDtoMEAN Low-Gamma	MAD High-Gamma
16	VAR High-Gamma	cross Alpha	AR Alpha error
17	Energy High-Gamma	RMS Delta	cross High-Gamma
18	RMS High-Gamma	cross Low-Gamma	AR High-Gamma error
19	Cross High-Gamma	RMS Theta	AR coefficient Alpha
20	MAD High-Gamma	cross Beta	AR High-Gamma coefficient
21	AR Low-Gamma error	RMS Delta	AR Beta coefficient
22	complexity Low-Gamma	RMS Low-Gamma	AR Beta coefficient
23	Cross High-Gamma	complexity High-Gamma	AR Beta error
24	VAR Low-Gamma	mobility Delta	AR Alpha coefficient
25	Energy Low-Gamma	mobility Low-Gamma	IQR Beta
26	RMS Low-Gamma	cross Delta	AR Delta coefficient
27	IQR Low-Gamma	complexity Alpha	AR Delta error
28	Cross Low-Gamma	complexity High-Gamma	AR Alpha error
29	AR coefficient Alpha coefficient	complexity Low-Gamma	AR Low-Gamma
30	AR High-Gamma error	complexity Delta	mobility Delta

Table 5-6. Top 30 feature-subset ranked by three types of approaches for frontal lobe epilepsy.

Top Features	KW	SVM-RFE	XGBOOST
1	AR Low-Gamma error	AR High-Gamma error	AR delta coefficient
2	AR Alpha coefficient	AR Alpha error	AR Theta error
3	AR Theta error	cross High-Gamma	AR Delta coefficient
4	AR Low-Gamma error	AR Theta coefficient	AR Theta error
5	AR Low-Gamma error	cross correlation Beta	AR Theta error
6	AR Theta error	AR Delta error	AR Delta error
7	Complexity Alpha	cross Alpha	AR Low-Gamma error
8	IQR Alpha	cross Beta	AR Alpha coefficient
9	mobility Beta	AR Delta error	AR Alpha error
10	IQR Alpha	RMS Low-Gamma	AR Alpha error
11	IQR Theta	cross delta	AR Theta coefficient
12	IQR Beta	RMS High-Gamma	AR Beta coefficient
13	IQR Low-Gamma	MAX cross Alpha	AR Beta error
14	IQR Alpha	complexity Alpha	kurtosis High-Gamma
15	kurtosis Beta	MAX cross High-Gamma	AR Beta error
16	AR Beta coefficient	MAX cross Beta	AR Low-Gamma coefficient
17	MAD Alpha	MAX cross Theta	AR Beta coefficient
18	MAD Beta error	MAX cross Low-Gamma	AR High-Gamma
19	Complexity Alpha	MAX cross Beta	AR Beta coefficient
20	IQR Theta	RMS Delta	AR Beta error
21	AR Beta error	MAX cross Alpha	AR High-Gamma error
22	IQR High-Gamma	MAX cross Theta	AR Delta coefficient
23	AR Theta coefficient	complexity Delta	AR High-Gamma error
24	Energy Alpha	RMS Low-Gamma coefficient	AR Delta error
25	RMS Theta	mobility Alpha	AR High-Gamma coefficient
26	VAR Alpha	complexity High-Gamma	AR Delta error
27	MAD Alpha	mobility Alpha	AR High-Gamma error
28	Complexity Theta	complexity Low-Gamma	AR Theta coefficient
29	Complexity Theta	complexity Delta	MAD Beta
30	Energy Beta	mobility High-Gamma	AR Delta coefficient

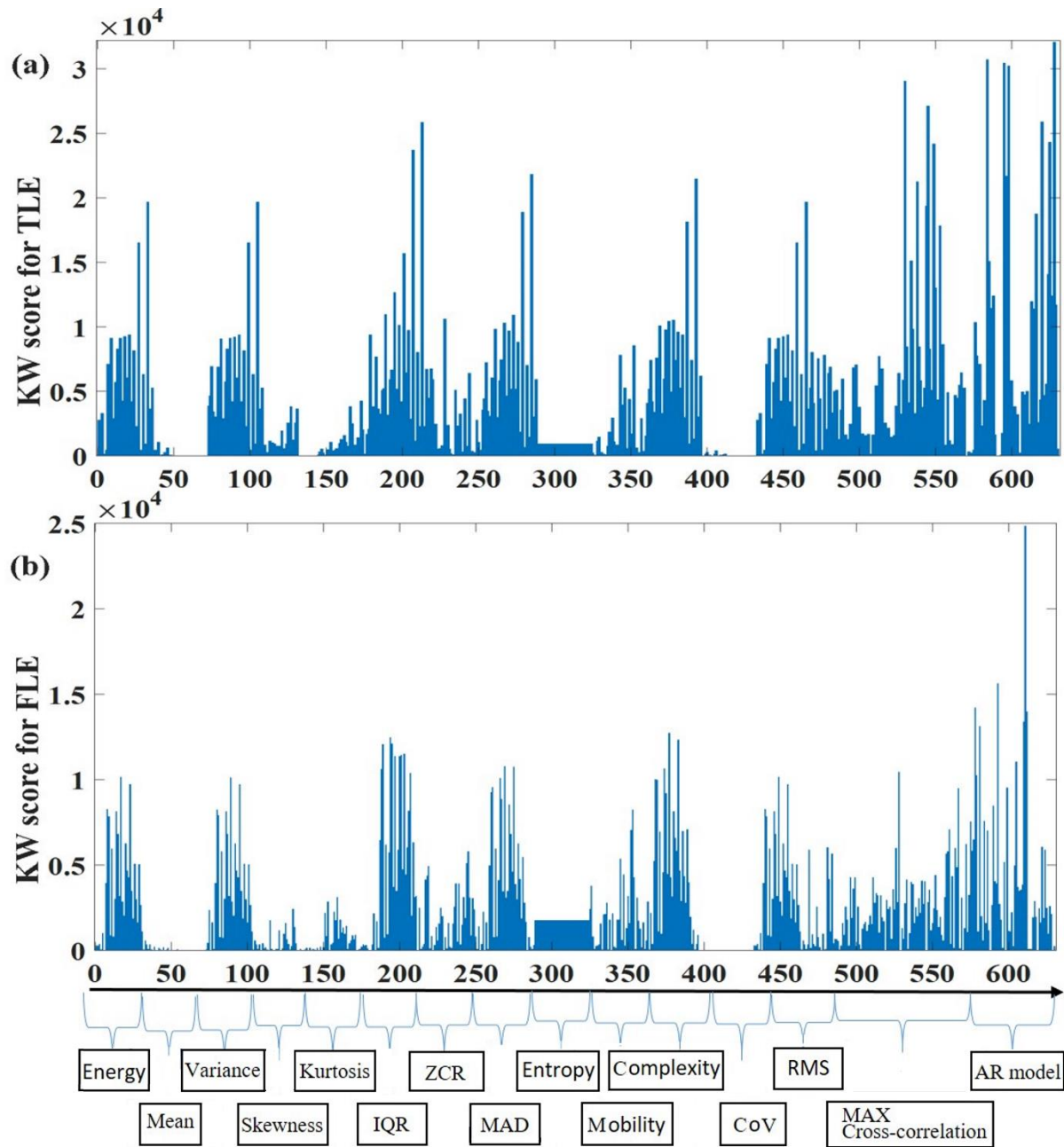


Figure 5-3. An overview of the features ranked by Kruskal-Wallis as a winner method. : (a) for temporal lobe; (b) for frontal lobe.

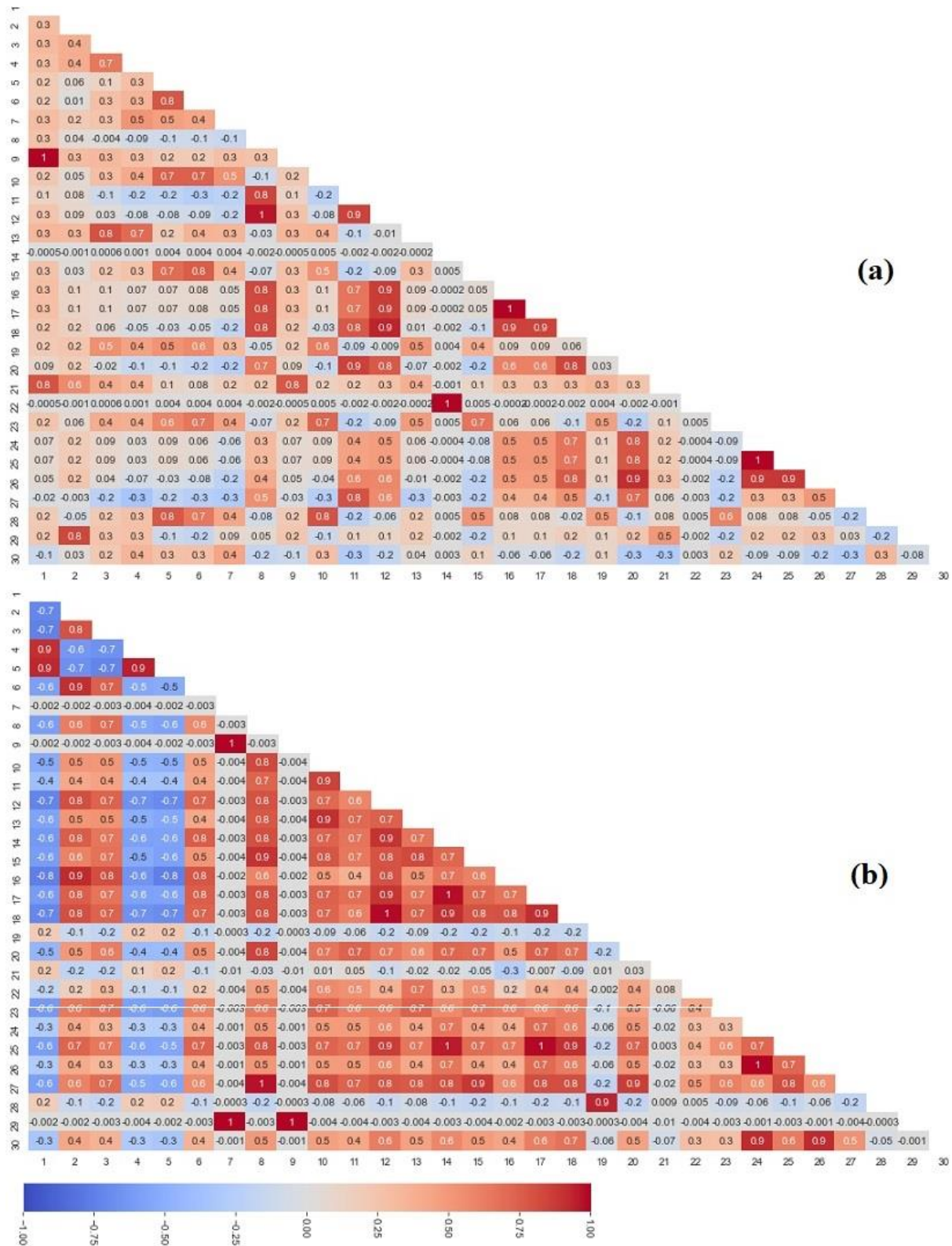


Figure 5-4. Heat-map of the correlation matrix between the top 30 subset of features with the highest Kruskal-Wallis scores: (a) for the temporal lobe; (b) for the frontal lobe. The dark red and dark blue show a high correlation magnitudes which indicate those features contain similar information while gray color identifies a low correlation.

Table 5-7. Features with the strongest linear relationship among top 30 features with the highest Kruskal-Wallis scores.

Lobes	Attributes with a strong linear relationship			
Temporal lobe	AR High-Gamma error	/	AR High-Gamma coefficient	
	IQR High-Gamma	/	MAD High-Gamma	
	Energy High-Gamma	/	VAR High-Gamma	
	complexity High-Gamma	/	complexity Low-Gamma	
	VAR Low-Gamma	/	Energy Low-Gamma	
Frontal lobe	Complexity Alpha	/	Mobility Beta	/ Energy Beta
	MAD Alpha	/	IQR Alpha	/ RMS Theta
	IQR Beta	/	MAD Beta	
	Energy Alpha	/	VAR Alpha	

The mRMR (Minimum Redundancy Maximum Relevance) method is one of the most well-known, powerful, and widespread approaches as filter-based feature selection that tends to rank features with the highest correlation with the class (output) and the least correlation among themselves [190]–[192]. Based on the results of KW, we decided to compare the winner approach as a filter with mRMR to compare the performance of the new approach to remove redundant attributes and select the largest relevance ones. We implemented mRMR for temporal lobe epilepsy data and we found out that this method of feature selection is indeed very fast (it took less than 3 minutes) but the results were not better than those obtained with KW. The reported MCC score is 0.33, which is too low compared to that of KW. The other issue with this ranking feature is about the ranked features; the parameters related to AR model are not the most important among the extracted features in this work. We checked the top 30 attributes and none of the attributes from AR model was selected by mRMR (figure 5-5) while the results of all feature ranking approaches utilized in this work proved the importance of the AR model parameters. Therefore, mRMR as a powerful feature ranking approach does not surpass the retained filter method, KW.

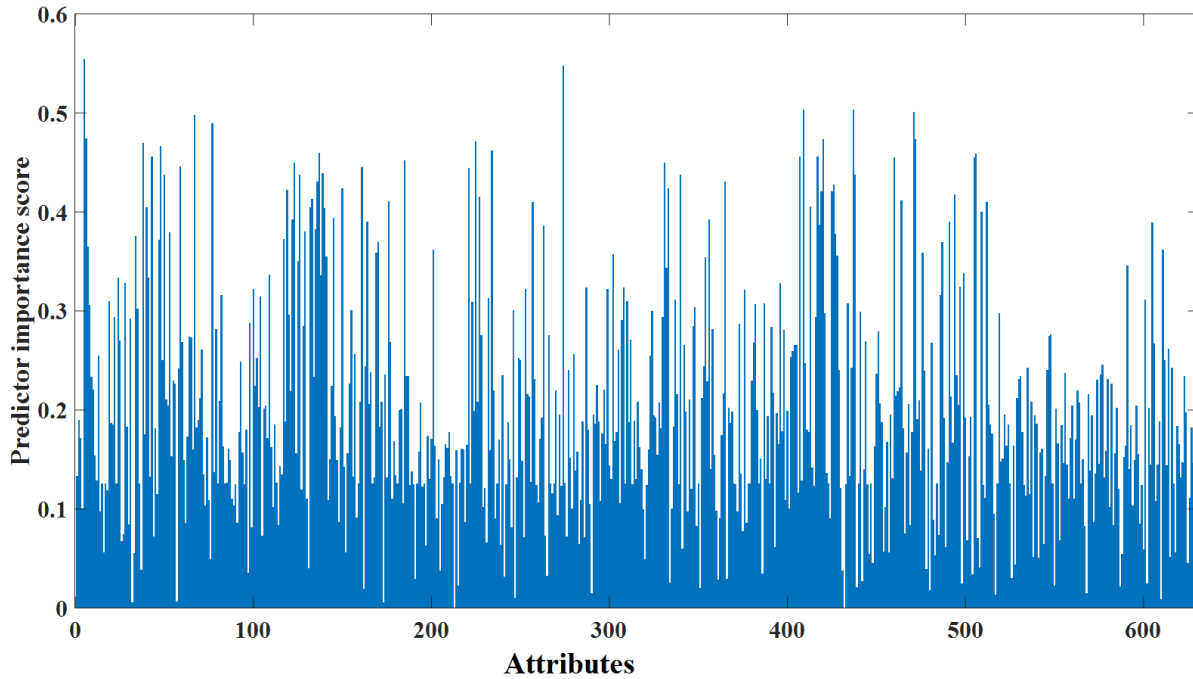


Figure 5-5. The result of mRMR feature ranking for temporal lobe.

5.7 Discussion

For different reasons, some researchers have divided the EEG sub-bands into various sub-bands [45], [46] while other did not [37], [47]. Our aim was, therefore, to consider both cases and evaluate the impact of dividing the EEG signal into various sub-bands. In fact, as shown in Figures 5-2-a and 5-2-b, dividing the EEG signal into 6 sub-bands will carry more predictive information than not splitting it.

In this study, we compared three various feature selection methods and the results showed that the filter method has the highest performance with the highest MCC and accuracy. Also, KW can rank the features a lot faster, with the shortest computational time.

A large panel of wrapper approaches have been proposed for feature selection but most of them are computationally expensive and complex in nature [177]–[179]. The results we obtained confirmed this fact: SVM-RFE has the lowest prediction performance and is the most intensive in terms of computation. Although, in most of existing works, it has been claimed that the embedded

methods that combine filters and wrappers take advantage of both, the obtained results did not really demonstrate that claim, showing that the non-parametric filter-based method, Kruskal Wallis, outperforms better than the above approaches.

The AR method estimates the power spectrum density (PSD) of a given signal. Then, this approach does not have the problem of spectral leakage and we can expect a better frequency resolution dissimilar the nonparametric method. PSD can be estimated by calculating the coefficients even when the order is low [193]–[195]. Furthermore, the prediction error term extracted from an AR model of the brain signals is claimed to reduce during preictal stage [146].

The maximum of cross-correlation, a bivariate feature, can be considered as a measure for lag synchronization due to estimating of the phase difference between two spatially separated sensors even with a low SNR [131], [196]. The key points for Kruskal Wallis as a winner, can be due to not just considering the parameters of AR model and MAX cross coloration. This feature selection tried to engage other important univariate features like complexity, which having an estimation of statistical moment of the power spectrum.

The temporal lobe is responsible to deal with the processing of the information and play a role in long term memory. Gamma rhythms are involved in higher processing tasks and cognitive functioning. These waves are the fundamental waves for learning, memory and information processing. The Frontal lobe is responsible for emotion control center, planning, judgment, and short-term memory. Theta rhythms are produced to help in creativity, relaxation, and emotional connection. Alpha waves help in feeling of deep relaxation and Beta waves are related to someone's conscious and problem solving [120].

In this study, we found that the low and high-gamma sub-bands are the most discriminating ones between preictal and interictal for TLE patients while the frequency ranges from Theta to low-Gamma were found to be the most discriminating features in six patients with FLE. Our results confirmed that gamma sub-bands are a promising biomarker in predicting of seizure for TLE [197]–[200]. However, for FLE, we should consider a wider range of frequencies (lower frequency compared to TLE) in the preictal stage. Note that some existing works in detection of FLEs

proposed that frequencies less than 50 Hz can play dominant roles among different brain waves [32], [33], [201].

Our results also demonstrated the complexity of seizure prediction and detection due to the fact that the frontal lobes of the brain control a wide variety of complex structural and functional roles [25]–[27], [202]. These findings can help establish specific relationships regarding the impact of each lobe in a specific function and the generation of waveforms based on that function.

Furthermore, by comparing the performance results of Kruskal Wallis for both lobes in Table 5-4, we see that MCC is not close to 1, the perfect prediction case by looking at the accuracy values. Not having very high accuracy can imply that the issue of the imbalanced was resolved and the reason for not having a high MCC is related to the low capacity of this version of Freiburg database due to having data up to 90 minutes of preictal, or to the fact that some seizures take few minutes. Based on Ramachandran *et al.* [15], we might need at least 3 to 16 hours before the onset of the seizure to efficiently predict seizures. It may be considered as a limitation in this study and weakness of this database. Another possible solution is employing non-linear measures, such as phase synchronization, to improve the model performance in forecasting seizures.

Moreover, according to [6], [119], a non-stationary signal like EEG can be anticipated as a stationary signal in a short duration epoch, like a two-second window. Also, based on the results in figures 5-3-a and 5-3-b, the variation of the mean feature for both lobes is nearly a constant value. This result partially confirms the previous claim, but this would require further investigations.

The effectiveness of the Kruskal Wallis as a nonparametric method is based on the fact that it does not need to assume a data distribution model, making our results promising in feature selection of EEG data of TLE and FLE. Therefore, dealing with a higher number of epilepsy patients will not be an issue and this approach would be applicable to a larger set of data. Furthermore, we divided the data into train and test sets and the model built by random forest was trained and validated with 10 fold cross validation. Also, we prevented overfitting, which can be considered as a generalized model for an unseen data.

Note that by applying mRMR, we spent not only less time in ranking the features but also we did not need to remove the redundant attributes, which we had to do for KW. However, by comparing the results of MCC for KW and mRMR, we found out that KW is really more efficient than its counterpart.

Furthermore, applying MCC in measuring the performance of the model, is a significant improvement over other existing works that deemed accuracy as the performance of the classifier while dealing with imbalanced data [15], [38], [108]. Moreover, according to [19], [20] a non-stationary signal like EEG can be anticipated as a stationary signal in a short duration epoch, like a two-second window. Also, based on the results in figures 5-3a and 5-3b, the variation of the mean feature for both lobes is nearly a constant value. This result partially confirms the previous claim, but this would require further investigations.

We do emphasized on the computation time in this chapter. Nevertheless, one may argue that since the training process can be performed offline, the computation time should not be an issue. However, as already mentioned in the objectives, our long-term target in implementing our algorithm in portable devices is to enhance the daily life of the patients and increase their autonomy. Thus, we opted for a dual-mode operation:

- The training can be done offline and the model can be updated afterwards via any available communication application. So, the model can be updated after an offline training.
- Knowing that (i) the application should need to frequently integrate new data in its early utilizations and (ii) that some patients could not have regular access to wireless connections, the other possible scenario is to update the tool online. Then, the training/update should be performed on the portable device itself while the application is still working on prediction mode.

To ensure this dual-mode operation, it is crucial to shorten the training CPU time while making the tool operation as simple as possible.

We believe that the findings of this work can be implemented on a low power hardware by efficiently considering less complex features for a specific sub band with the information from

only one patient, instead of building and deploying a model for the entire patients in the database. Note that, as already mentioned, the obtained results are based on a limited number of patients.

5.8 Conclusion

The epileptic seizures are the temporary occurrence of symptoms due to synchronization of abnormally excessive activities of the brain nerve cells. However, reviewing of the EEG signals will be a time consuming task for neurologists to analysis and monitor continuous electroencephalograms. Therefore, even it is quite challenging, implementing a high performance automated analysis of EEG signals is in high demand.

The Kruskal-Wallis feature selection strategy is simple and less time consuming as compared to other approaches. Among the time-domain features investigated, the parameters of AR model are ranked as the top features for both lobes. The second most important features are the maximum of cross-correlation and IQR for temporal and frontal lobes, respectively. Moreover, a high range of frequency like low and high-Gamma have been introduced as an interesting sub-band for the temporal lobe epilepsy, while the middle range of frequencies from Theta to Beta can be seen as important ranges of frequency for frontal lobe epilepsy.

Future efforts should be focused to reliably improve the performance of the prediction on test set for a patient-specific by considering a combination of various features that provide an estimation of phase, frequency and amplitude of the EEG signal.

Chapter 6

Probabilistic approach for efficient prediction of epileptic seizure with iEEG signal

6.1 Introduction

In this chapter, we will employ the time domain features thoroughly explained in the previous chapter. We extracted those features within a 2-second window while in this chapter we will vary the window size from 1 to 40 seconds and investigate the changes in the performance of various classifiers. After finding the optimum window size, we will introduce a new probabilistic approach in seizure prediction in order to improve the performance of the existing studies in this field.

6.2 SOP and SPH definitions in seizure forecasting

We introduce three additional important terms in the field of seizure forecasting (as illustrated in figure 6-1):

- Seizure occurrence period (SOP): The time interval in which a seizure is expected to occur [15], [203], [204].
- Preictal period (PP) or preictal zone: the period before a seizure, which is clinically obscure but implicitly determined in the dataset [41].
- Seizure prediction horizon (SPH): The interval between the alarm and a leading seizure expected to occur, is named SPH. This period is also vague though it should be in the preictal period [41], [203], [204].

During the SPH interval, a seizure-warning tool can effectively inform the patient to behave carefully or treatment plans can be employed [203]. Based on the clinical aspect, SPH should be high enough to allow sufficient time for the patient to behave cautiously or take medicines. By contrast, SOP should be low enough to soothe the patient's anxiety and stress [203], [204].

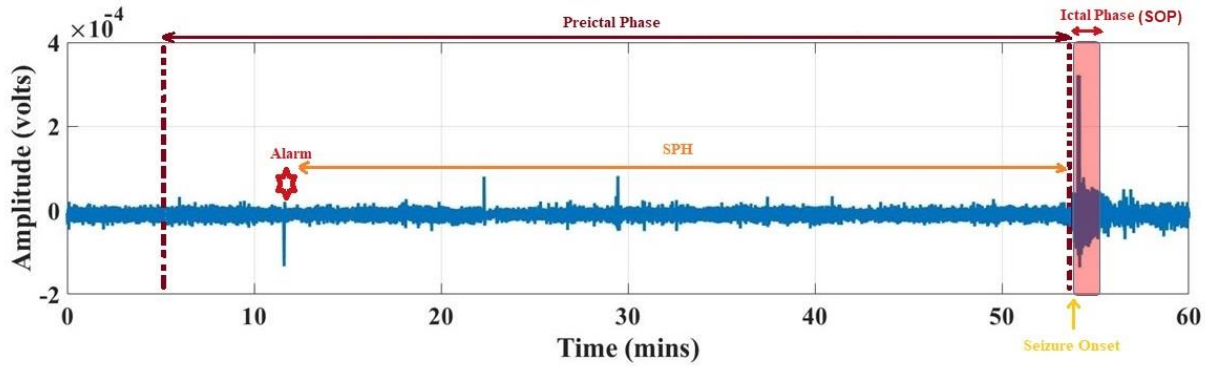


Figure 6-1: Durations of seizure occurrence period (SOP) and prediction horizon (SPH) are demonstrated.

6.3 Accuracy, an inadequate metric to evaluate an imbalanced dataset

The most common and natural technique to evaluate a classification model is to compute the accuracy, the ratio between the number of correct predictions and the total number of predictions. We consider the case with the task to forecast whether a given signal is preictal or abnormal (1) or interictal or normal (0).

In the bi-class scenario like our case, the class label being either 1, representing preictal, or 0 representing interictal. When the actual label is 1 and the predicted label is 1, then this is a true positive, TP, where the label is legitimately forecasted as positive. When both the actual and the predicted labels are 0, thus this is called TN (true negative), then the label is properly projected as negative. If the actual label is 0 and the predicted label is 1 we have the false positive case, FP. The case the label is fraudulently forecasted as positive while it should be negative is pondered as type 1 error. Reversely, the case when the actual label is 1 and the predicted label is 0, is considered as false negative, FN. We ponder this case as type 2 error [205]–[207].

A data with unequal distribution between its majority and minority groups is titled imbalanced dataset. Interestingly, a medical device to forecast a rare case as seizure cannot solely depend on plain accuracy in evaluation. The cost associated with missing a seizure is worse than fraudulently labeled a healthy signal as an abnormal signal. To investigate the issue of the imbalanced data, we considered the EEG signal of a patient from the database, #4, in this study. The imbalanced ratio was about 22 and 20 for train and test set respectively (the size of majority class to minority one). The result of the classification for this data has been inserted in Table 6-2.

Table. 6-1 The confusion matrix for a classifier.

Actual label	Predicted label	The result of comparison	Error type
1	1	True Positive (TP)	--
0	0	True Negative (TN)	--
0	1	False Positive (FP)	Type 1 error
1	0	False Negative (FN)	Type 2 error

Table. 6-2 Various measures for multiple classifiers in an imbalanced dataset.

Performance analysis of classification	SVM	XGB	RF
True Negatives	13731	13730	13731
False Positives	0	1	0
False Negatives	691	241	691
True Positives	0	450	0
Specificity	1.0	1.0	1.0
Sensitivity	0.0	0.65	0.0
Accuracy	0.95	0.98	0.95
F1 score	--	0.65	--
Average precision-recall score	0.05	0.67	0.05
AUC	0.50	0.83	0.50
MCC	0.0	0.80	0.0

This table demonstrates that SVM and RF classifiers could not deal with the imbalanced issue compared to XGB. Having a high value of accuracy while the rest of the measures like MCC are low can be consider as failure in dealing with the imbalanced classification problems. Thus, we found out that the plain accuracy may be misleading.

6.4 The optimum window

In this section, we describe the process of determining the optimum window length (figure 6-2).

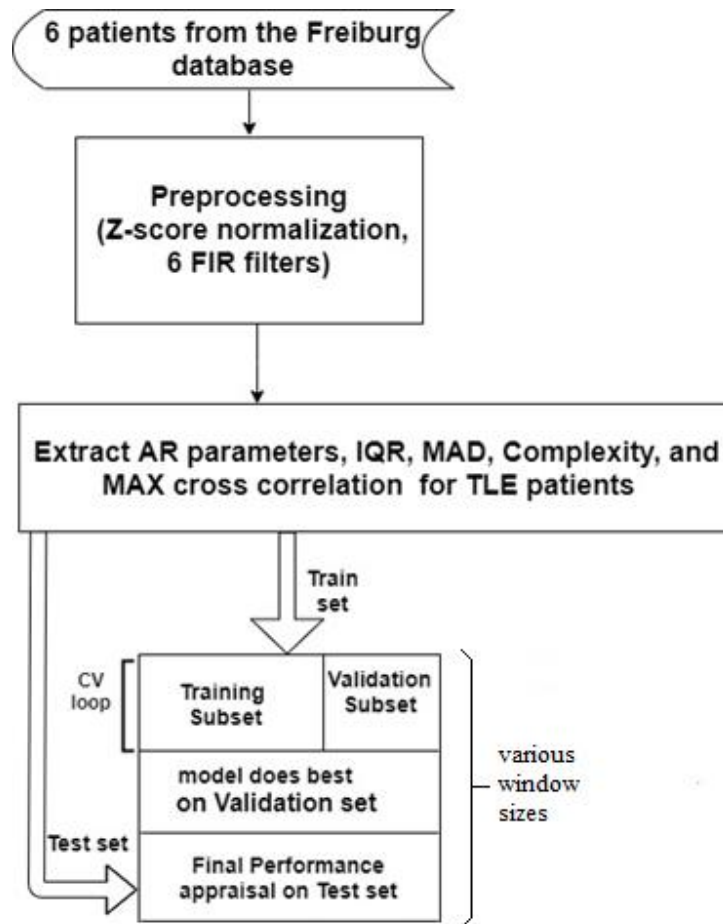


Figure 6-2. Flowchart of the work to select the best length of the window (CV stands for “cross-validation” loop).

As mentioned before, an iEEG signal was chunked into various sections and during each sliding window a specific feature was extracted. This is repeated until the whole signal is treated. The time window size varies from study to study from 1 to 20 seconds [15], [41], [203], [204], [208] and some works have been done about finding an optimum length of seizure prediction [38], [53],

[209], but due to lack of information related to the performance of the classifier, this issue still needs further investigation.

The length of the window is a crucial parameter that needs to be chosen carefully. This size cannot be deemed too small since the signal requires to be long enough to provide reliable values for the features. Furthermore, our aim in this work is to implement an algorithm on a low power device, which usually means limited computing power. Then with a very small window size, the features will require to be computed over and over in a short period of time. This can pose a threat once enough computational power is not available on a small implantable medical device [175], [209]. On the other hand, the window size cannot be taken too long because the characteristics extracted from the signal will not have a very smooth transition over time. In fact, because the targeted activity might occur at the beginning or in the middle of the window, the values extracted for the whole part of the sliding long-window may not exactly represent the type of the activity [175].

Six patients from Table 3-1 with temporal epilepsies originating from the hippocampal area were retained for this chapter namely, participants #2, #4, #7, #10, #12, and #16. 134 hours iEEG data available in the Freiburg database, i.e. 52 and 82 hours from preictal and interictal, respectively. Then based on the assumption in the previous chapter, we considered to employ besides accuracy other measures like MCC as the performance analysis in this part.

An hour EEG signal was divided into various non-overlapping window sizes (40 s, 20 s, 10 s, 4 s, 2 s, and 1 s) for interictal section, while the preictal section was split into chunks of same window sizes (i.e. from 40 s to 1 s) with 50% overlapping. The performances of various classifiers for different window sizes were reported in Table 6-3. We intended to employ various classifiers and measurements to increase the consistency in our results: SVM as linear approach, MLP as nonlinear method, the ancestor deep neural networks and two various ensemble tactics. Ensemble learning has been considered as a powerful machine learning algorithm that has demonstrated a great deal of advantages in various applications. This method of machine learning algorithm combines numerous base models in order to build one optimal predictive model [160], [210], [211]. We employed two various ensemble approaches in this study XGBoost (boosting) and Random Forest (bagging). The performance, MCC score, of the classifiers for different sets of window lengths were depicted in figure 6-3 based on the results from Table 6-3. XGBOOST had

the highest MCC score while MLP demonstrated the least performance. The result of MCC for XGBOOST are plotted on a graph, as shown in figure 6-4.

Among the available classifier candidates in the previous experiment, we found that XGBoost (XGB) gave the best performance in binary classification with the highest MCC score while the grid search has been performed to optimize the maximum depth parameter. XGB has been broadly utilized in numerous fields to demonstrate state-of-the-art results on some data challenges and it has shown a strong potential to solve the resulting difficulties in data analysis and is one of the most favourable classifiers in machine learning regarding classifiers [158], [212], [213].

Table 6-3. The performance of various classifiers for different window sizes.

Window /classifier	1s			2s			4s			10s			20s			40s		
	XGB	RF	SVM	XGB	RF	SVM	XGB	RF	SVM	XGB	RF	SVM	XGB	RF	SVM	XGB	RF	SVM
Accuracy	0.76	0.66	0.74	0.79	0.72	0.56	0.79	0.71	0.53	0.79	0.73	0.76	0.74	0.71	0.68	0.74	0.72	0.75
F1 score	0.75	0.54	0.75	0.79	0.66	0.61	0.78	0.63	0.73	0.78	0.67	0.76	0.70	0.64	0.70	0.71	0.65	0.76
F2 score	0.71	0.44	0.74	0.76	0.56	0.63	0.75	0.53	0.74	0.75	0.57	0.75	0.62	0.54	0.70	0.63	0.55	0.76
Average PR score	0.72	0.68	0.70	0.76	0.73	0.56	0.76	0.73	0.66	0.76	0.74	0.72	0.74	0.72	0.64	0.74	0.73	0.71
AUC	0.76	0.67	0.74	0.79	0.73	0.55	0.79	0.72	0.69	0.79	0.74	0.76	0.75	0.72	0.67	0.75	0.73	0.75
MCC	0.512	0.41	0.48	0.582	0.51	0.11	0.58	0.51	0.39	0.58	0.52	0.51	0.53	0.50	0.35	0.53	0.45	0.50

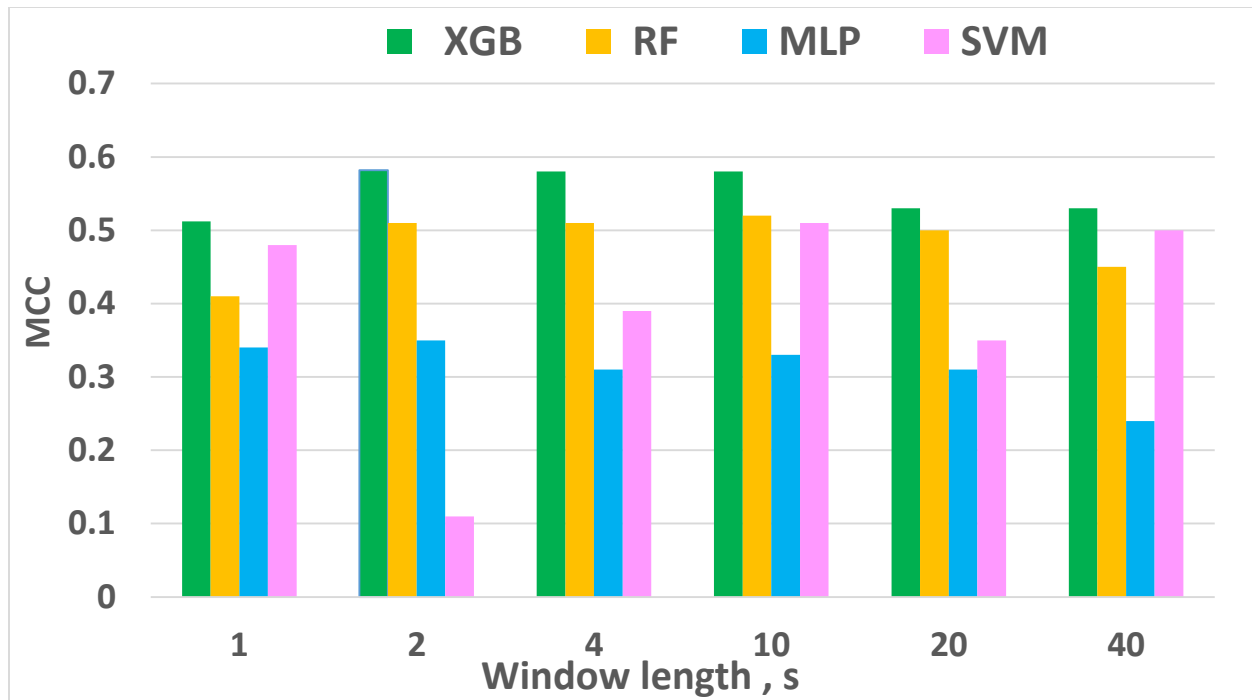


Figure 6-3: MCC for all the classifiers in a graph.

Finally, MCC score of XGB for various window sizes was displayed in figure 6-4, showing that the best MCC was located in the range of 2-10s window. We also drew the PR and ROC curves (Figures 6-5 and 6-6, respectively). In these plots, the line in the middle (the orange dashed line) shows that the classifier is a random guess while a perfect classifier assumes to cover the largest area under the curve [214], [215].

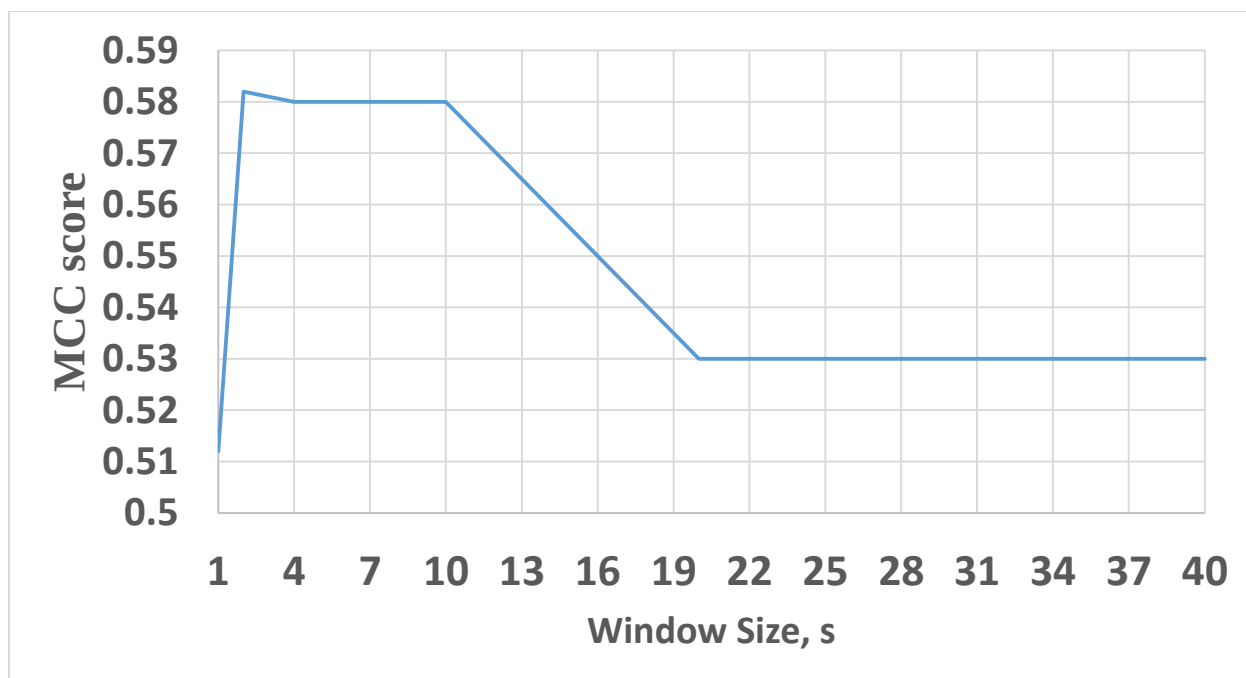


Figure 6-4: MCC of XGB for various window size.

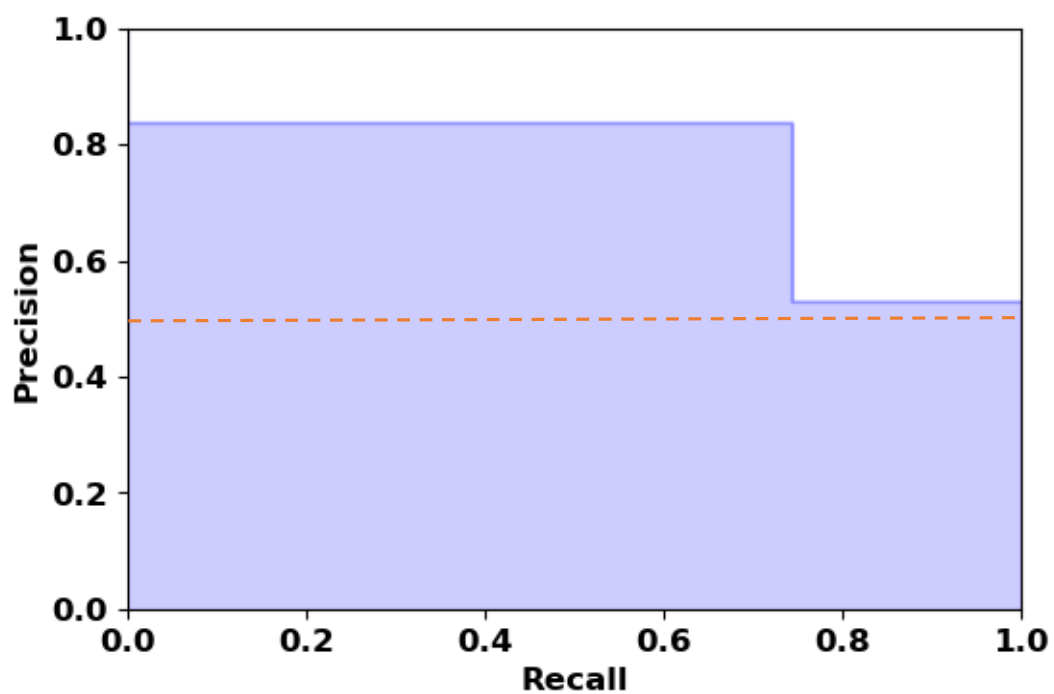


Figure 6-5: Precision and recall (PR) curve of XGB for 2 s window.

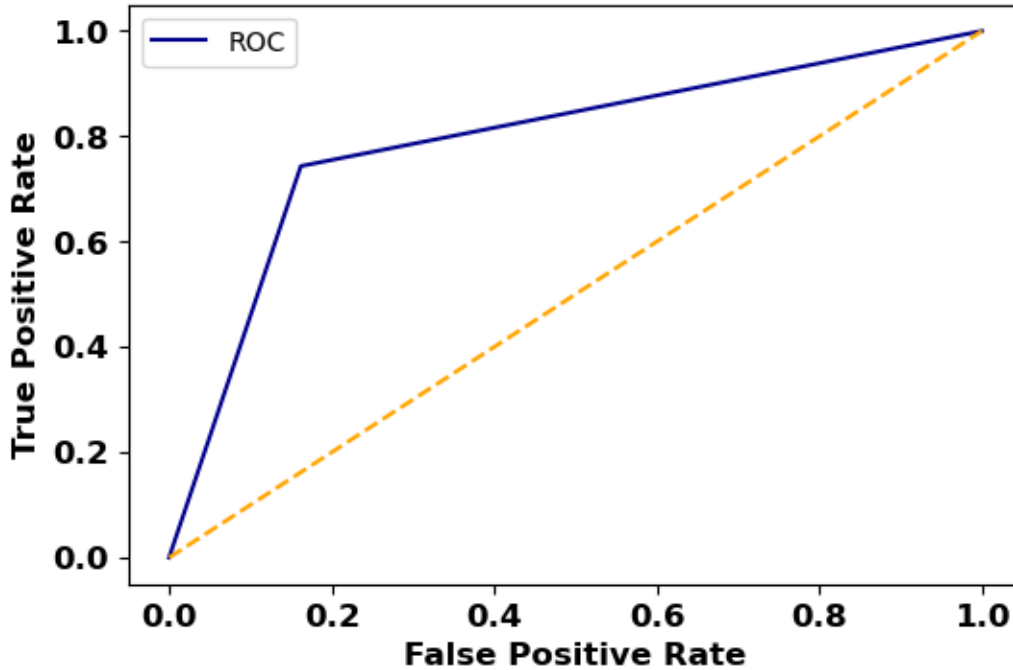


Figure 6-6: The Receiver Operating Characteristic (ROC) curve of XGB for 2 s window.

6.5 Discussion about the choice of the window size

Based on Table 6-4, Parvez [38] claimed that a 10 s window outperformed due to being more consistent with respect to the AECR values of preictal and interictal segments although the length of the data in this experiment was about 10 minutes. Then the above statement needs to be reinvestigated for a larger set of data. In another work [15], the authors employed accuracy on a highly imbalanced dataset to find the best window size, i.e. 20 s, from a set of 1 to 60 s window lengths. However, we investigated that accuracy alone is not the best metric to use. Therefore, the result of this work needs to be re-evaluated by applying other measures besides accuracy. In fact, in the above work, very few pre-ictal samples were studied while most of the samples belonged to the inter-ictal segments. Then, the model may gradually lose the capability of predicting the preictal patterns succeeding a long inter-ictal patterns. Same issue can be applied to [52].

In [53], Singh employed the posterior probability of the classifier and compared it for various window sizes and then, the 90 s window was chosen due to having the highest probability while

the performance of the model was not reported whether the under-fitting or over-fitting resolved during the training or not. In [54], the authors calculated the correlation coefficient during a certain time window (from 0.5 to 300 seconds) and investigated the performance of the classifier just by employing AUC. They found out that the time window of 60 s and 30 s demonstrated the highest AUC for seizure prediction in humans and dogs, respectively. Again, the imbalanced ratio was not reported while AUC demonstrated various flaws and drawbacks owing to the fact that it is sensitive to class imbalance [214]–[216].

Note that our work does not suffer from the imbalanced issue which this ratio is about 2:1. We applied various performance measures and classifiers to make sure the prediction system was investigated thoroughly. For a window length of 1 s, the classifiers performance has not been improved, which can confirm Islam's works [59], [119] that, by choosing a very short window, e.g., less than one second, the seizure waveform may not be recognized properly in such a short duration.

We intended to keep the stationarity of the EEG signal and based on the work of Islam [59] by choosing a very short window, e.g., less than one second, the seizure waveform may not be recognized properly in such a short duration. On the other hand, by defining a longer window length, e.g., more than 3 seconds, the assumption of stationarity of an EEG signal is not valid anymore. Therefore, artifacts and seizure cannot be distinguished from each other. In conclusion, a non-stationary signal like EEG can be assumed as a stationary signal in a short duration epoch like a two-second window.

Another point to note that, in order to optimize the device performance for a real time processing we need to take into consideration the factors of computing resources, power consumption and CPU time In a real time processing. In a real time scenario, a device will be always sequentially extracting and calculating some features then by assuming for a longer window size, the device needs to calculate the features based on a longer portion of the signal. The longer the window size then higher will be the power consumption and CPU time.

Table 6-4. The comparison of our study with existing works.

Name of the work, database	Feature used	Window size used in the paper	# of Subjects and seizures	Performance of the classifier	imbalance ratio (data length)
[15], 2018	Area, Normalized decay, Line length, Mean energy, Peak amplitude, Valley amplitude, Normalized peak number, Peak variation	1-100 s	7 subjects (5 dogs and 2 humans)	accuracy	20:1 for dogs 2:1 for patients
[53], 2016	Power Spectral Entropy, Fast furrier Transform, Higuchi Fractal Dimension, and Hurst Exponent	10-600 s			optimal window size can be argued to be around 90 seconds
[54], 2015	calculate the correlation coefficient in a certain time window between all possible pairs of EEG signals	0.5-300 s	Kaggle dataset, 5 epileptic dogs and 2 epileptic patients	AUC, no other information about the classifier performance	in humans, best classification is showed by SVM classifier for a time window $T_w = 60$ s (AUC = 0.9349); for seizure prediction in dogs, highest obtained AUC is 0.9432 for SVM classifier and $T_w = 30$ s
[52], 2008	Empirical Mode Decomposition (EMD) and AR model coefficients	12, 24, 35, and 47 s	19 patients the Freiburg database	Accuracy and variance	From 5 to 20 min of preictal and same length from interictal
[38], 2016	Average of the energy concentration ratio	5, 10, 15 seconds	21 patients	No information	10 second window outperformed due to being more consistent with respect to the AECR values
This work	AR parameters, IQR, MAD, complexity, and MAX cross coloration	1-40 seconds	6 patients from the Freiburg database	MCC: 0.582	1.6 (52 and 82 hours preictal and interictal respectively)

6.6 Proposed method for probabilistic approach

In order to undertake this research, various phases were recruited: data collection, data preprocessing and feature extraction, classification and optimization, evaluation and decision making. The overview of the proposed pipeline is illustrated in Fig. 6-7.

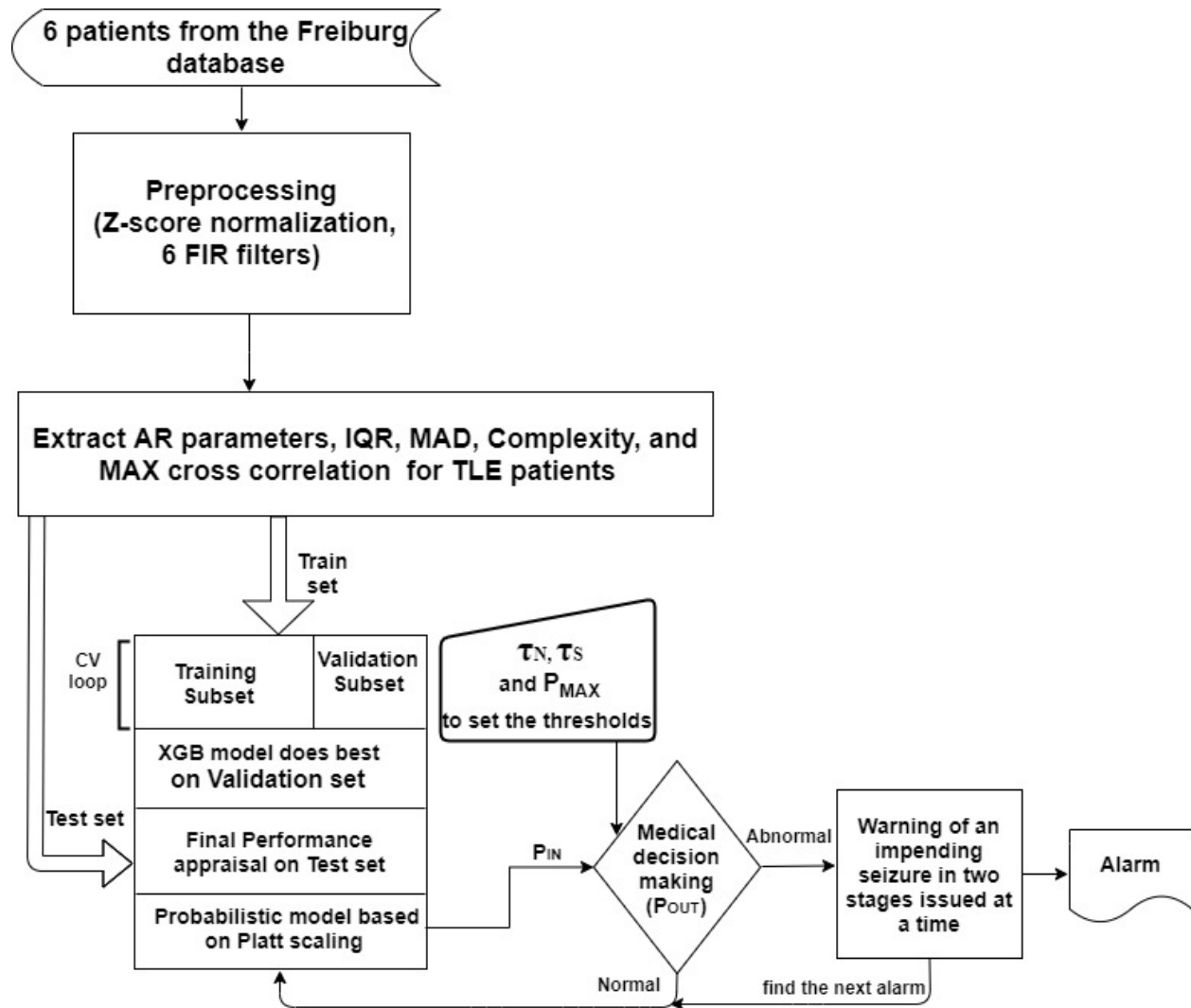


Fig. 6-7 Flowchart of the proposed work.

6.6.1 Preprocessing and feature extraction

We utilized the same preprocessing as in chapter 5 and the results of feature selection in figure 5-3 (a). The signal was chunked to 2 seconds window and 6 features were extracted namely, AR parameters, IQR, MAD, complexity, and MAX cross correlation. Afterwards, the data were divided into two groups: 75% of them allocated for training and validating, and the remaining 25% (never-seen-before) for testing the model. Both training and testing data segments have been numbered sequentially. “Numbered sequentially” means they are not randomly fetch to the classifier. The preictal sections are applied sequentially to the classifiers and from that, the classifiers try to find a pattern. Note that the validation set was used for estimating how precisely the classifier, XGB (with maximum depth of 1), performs after being trained with the learning set, including validating to prevent over-fitting of the algorithm and reducing of the bias estimation [217], [218] so that the model can generalize well to predict new samples of data. Once the classifier was fully optimized with the learning and validation sets, via 10-fold cross-validation, it was applied to the training set to assess final performance. We used the scikit-learn (machine learning) library through the python package for this work [219]. The experiments were executed on a desktop computer with Intel® Core™ i7 CPU @ 3.3 GHz processor, 16 GB RAM, Windows 7 Professional 64-bit operating system, and HDD storage.

6.6.2 Probabilistic framework

Platt scaling is used to extract the probabilities of each class, i.e. seizure vs. non-seizure. This calibration method passes the output of a classifier from a single model to a probability distribution through a sigmoid, as a result an estimation of posterior probability on (0,1) [220], [221].

In the next step, two important parameters are calculated namely, the mean and maximum probability of the non-seizure (P_{NS}) and the seizure occurrence period (P_{SOP}) segments for each participant along with the maximum probability of the P_{IN} . The histogram of Platt scaling for the whole test set (P_{IN}) is illustrated in figure 6-8. It shows that the model predicts most of the time

the lowest value and the chance to have a seizure is not too high. However, in the case the system creates a very high probability value then, possibly, we should have been dealing with a high rate of false positive rate since we had a larger set of interictal section.

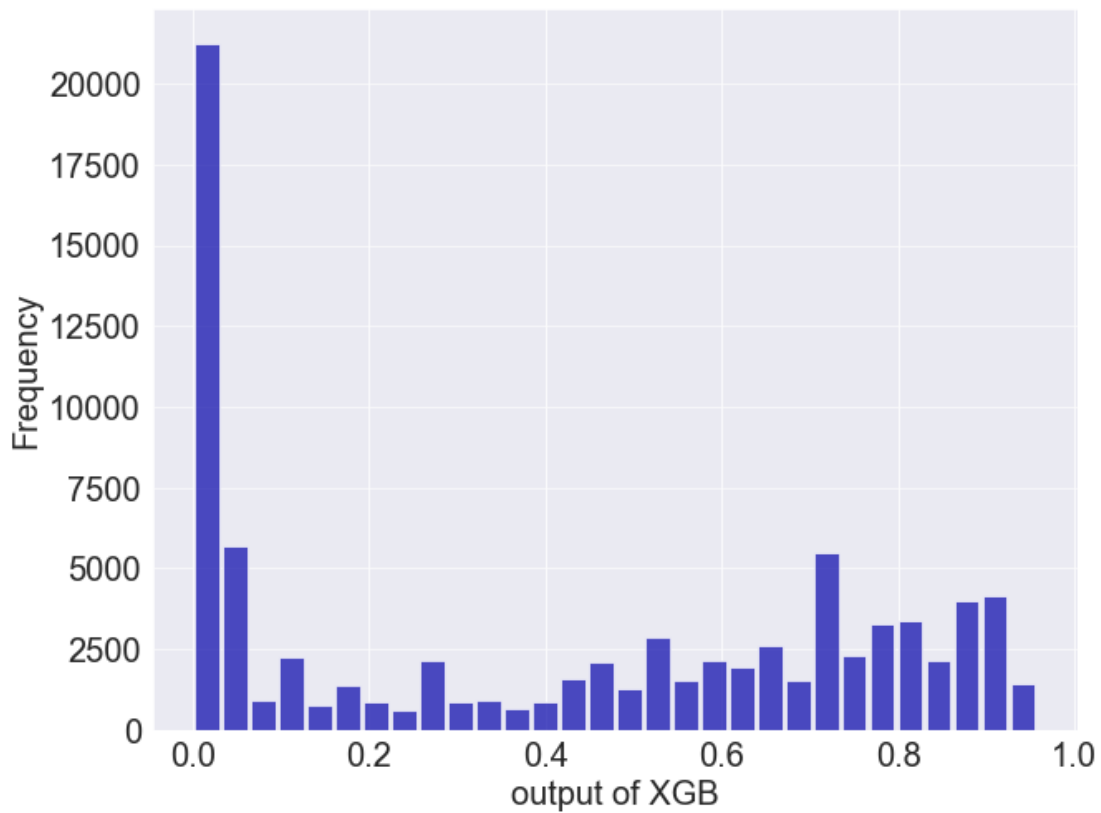


Fig. 6-8 The histogram of Platt scaling for the whole test set (P_{IN}).

Successive to learning the data using the optimized XGBoost framework and extracting the probabilities as described earlier, few adaptable thresholds were chosen to forecast the seizure efficiently based on the following procedure.

In order to set the thresholds (τ_N , τ_S) for 6 patients, the average and the maximum of the probability of non-seizure (NS), the seizure occurrence period (SOP), and the maximum probability of P_{IN} (P_{MAX}) were computed. Then τ_N and τ_S were defined as (Table 6-5):

$$\tau_{Ni} = \frac{P_{MEANi}}{P_{MAXi}} [for NS] \quad i=1, 2 \dots 6 \quad (6-1)$$

$$\tau_{Si} = \frac{P_{MEANi}}{P_{MAXi}} [for SOP] \quad i=1, 2 \dots 6 \quad (6-2)$$

These empirical constants were considered as thresholds based on numerous experiments for the 6 patients and, in order to predict an impending incident, a simple model was then retained.

The pseudo-code of the supervised prediction framework for an impending incident is shown in Algorithm 1.

Algorithm 1 Pseudo-code of seizure prediction framework with an adaptable threshold

```

1: procedure PREDICTION ( $P_{IN}$ ,  $\tau_N$ ,  $\tau_S$ ,  $P_{MAX}$ ,  $P_{OUT}$ )
2:   inputs:  $P_{IN}$ ,  $\tau_{Ni}$ ,  $\tau_{Si}$ ,  $P_{MAXi}$  ( $i=1, 2, \dots, 6$ )
3:   output:  $P_{OUT}$ 
4:   for each time point of  $P_{IN}$ 
5:     if  $P_{IN} \geq P_{MAXi}$  then
6:       go to line 11
7:     else if  $P_{MAXi} > P_{IN} \geq \text{MEAN}(\tau_{Ni} \text{ and } \tau_{Si})$  then
8:       return activate PA1 (Prediction Alarm level 1)
9:        $P_1 \leftarrow$  compute average of  $P_{IN}$  over a 1-minute window
10:      if  $P_1 > \text{MAX}(\tau_{Ni}/\tau_{Si} \text{ and } (\tau_{Ni} + \tau_{Si})/2)$  then
11:        return trigger PA2 (Prediction Alarm level 2)
12:      else go to line 7
13:    end if

```

```

14:         else go to line 5
15:     end if
16: end for
17: end procedure

```

As shown in this algorithm, the probabilistic prediction framework sequentially employs 270 dimensional feature-set (i.e., 3 univariate features, one bivariate feature, and 2 from AR: $3 \times 6 \times 6 + 15 \times 6 + 2 \times 6 \times 6 = 270$) to generate P_{IN} as one of the inputs to the probabilistic prediction framework (along with τ_{Ni} , τ_{Si} , and P_{MAX}). Based on the P_{IN} value and the first set of thresholds (τ_{Ni}, τ_{Si}), the system triggers the first prediction alarm (PA1) to warn the patient for an upcoming seizure with about 60% probability (the average of τ_N and τ_S) within the period T_1 (in Table 6-6). If P_{IN} is higher than P_{MAX} , then the second alarm (PA2 as T_2 in Table 6-6) will instantly be activated. Afterwards, the function accumulates the input for one minute (the future window length can be called the forecast horizon) to generate the second level of warning. This is done with the succeeding set of thresholds (MAX of $[\tau_{Ni} / \tau_{Si}$ and $(\tau_{Ni} + \tau_{Si})/2]$) which entrust 80% for an impending seizure (PA2) at a certain time T . Finally, it sends an urgent alert to the patient and caretaker. Interestingly, in all the published work the anticipation time of the first alarm has been reported while in this work we provided two sets of alarms and finally we will compare our work with others based on the first triggered alarm.

If the second condition is not met, at least the primary flag is raised and the system will continue to accumulate the next one minute P_{IN} to satisfy the PA2 condition. This procedure will continue as $P_{IN} \geq (\tau_{Ni} + \tau_{Si})/2$ and up to T_2 times (i.e., this procedure will continue, T_2 will be updated, and the system will remain active to make sure the seizure will happen as predicted earlier. The whole procedure is repeated until the end of the iEEG signal. It is worth mentioning that there are two values of T for patient #10 and #16 since we tested the model on two seizures for those patients.

Table 6-5. The implemented thresholds after numerous experiments for various patients.

The parameters of prediction for each patient		Pa#2	Pa#4	Pa#7	Pa#10	Pa#12	Pa#16	The average with 95% confidence interval
P_{NS}	P_{MEAN}	0.2068	0.002508	0.06607	0.02169	0.00249	0.0194	0.0532
	P_{MAX}	0.53349	0.01957	0.3059	0.05004	0.0233	0.03959	0.16198
	$\tau_N = P_{MEAN}/P_{MAX}$	0.387717	0.12815	0.21597	0.43337	0.1069	0.49087	
P_{SOP}	P_{MEAN}	0.34057	0.0644	0.16109	0.7457	0.76478	0.623	0.44997
	P_{MAX}	0.4499	0.10319	0.3059	0.9039	0.8478	0.863	0.57899
	$\tau_S = P_{MEAN}/P_{MAX}$	0.75699	0.62446	0.5266	0.8251	0.902	0.7219	
$(\tau_N + \tau_S)/2$		0.572	0.3763	0.3713	0.6292	0.5045	0.6065	0.51
τ_N / τ_S		0.51218	0.2052	0.41013	0.52527	0.11849	0.6799	0.41
$MAX([\tau_N + \tau_S]/2 \ \& \ \tau_N / \tau_S)$		0.572	0.3763	0.41013	0.6292	0.5045	0.6799	0.53
Maximum probability of P_{IN} , (P_{MAX})		0.6435	0.59142	0.7376	0.94712	0.9173	0.9577	0.7991±0.09

6.7 Results and discussions

After optimizing a classifier, the performance of the overall prediction was measured using common metrics utilized in the field to verify and estimate how successfully the proposed algorithm performs. Sensitivity of the triggered alarms and False Prediction Rate (FPR) were employed to demonstrate the results of this research and compare them with related works. Sensitivity measures the ratio of correctly predicted seizures divided by the total number of

seizures while the false prediction rate (FPR) is the number of false predictions over a period of time [222].

$$FPR = \frac{FP}{FP+TN} = 1 - \text{specificity} \quad (6-3)$$

Since in this work it is critical not to miss the seizure events (preictal), our aim was to maximize the sensitivity or True Positive Rate (TPR) [223] and, at the same time, to have a low FPR. Conventionally, in order to set a threshold, some researches plotted the TPR vs. FPR (the ROC curve). So there is a trade-off to find the best threshold, between avoiding a great number of false positives (FP) and benefiting from true positive (TP), in order to maximize the performance of the prediction. In fact, it not easy to simultaneously lower the number of false alerts and increase the sensitivity [212], [223]–[225].

In the last decade, various approaches have been employed by scientists to attain an increase of the sensitivity and a decrease of low false-positive rate while targeting a high anticipation time [53], [107], [168], [203], [225].

If the outcome is above threshold, a seizure alarm is triggered (in this work we inserted the values of the alarm in the first stage). As explained earlier, the threshold values of our seizure prediction framework are set so that they do not miss seizure episodes.

The results of this work for both sets of alarms are depicted in Table 6-6 and 6-7. We also illustrated the output of the probabilistic model based on Platt scaling (P_{IN} in figure 6-7) for patient #10 and #12 in figure 6-9 and 6-10, respectively. In figure 6-9, we illustrated P_{IN} for both preictal and interictal segments. The seizure is happening between 5980 and 6133 seconds (highlighted in red) which precede the interictal sections. The seizure period is highlighted in red for patient #12 in figure 6-10.

The proposed prediction system not only is capable of forecasting the seizures (a high sensitivity) but also will not generate a high false alarm (Table 6-6). As already noted, these results should be relativized because of the limited number of patients available in the used database.

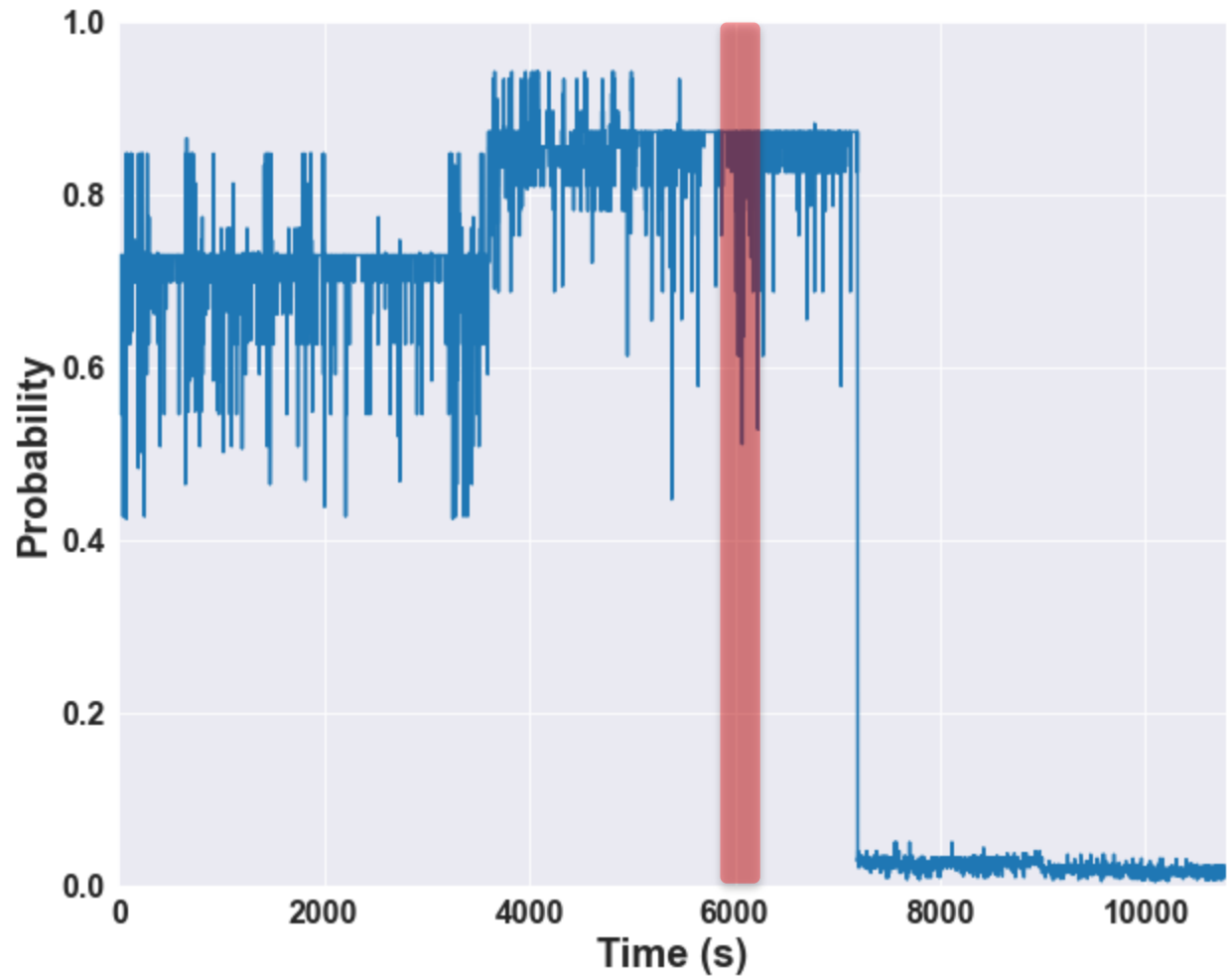


Fig. 6-9 P_{IN} related to file #166 included preictal and proceeding of the interictal portions for Patient #10. The red bar is the period of the seizure.

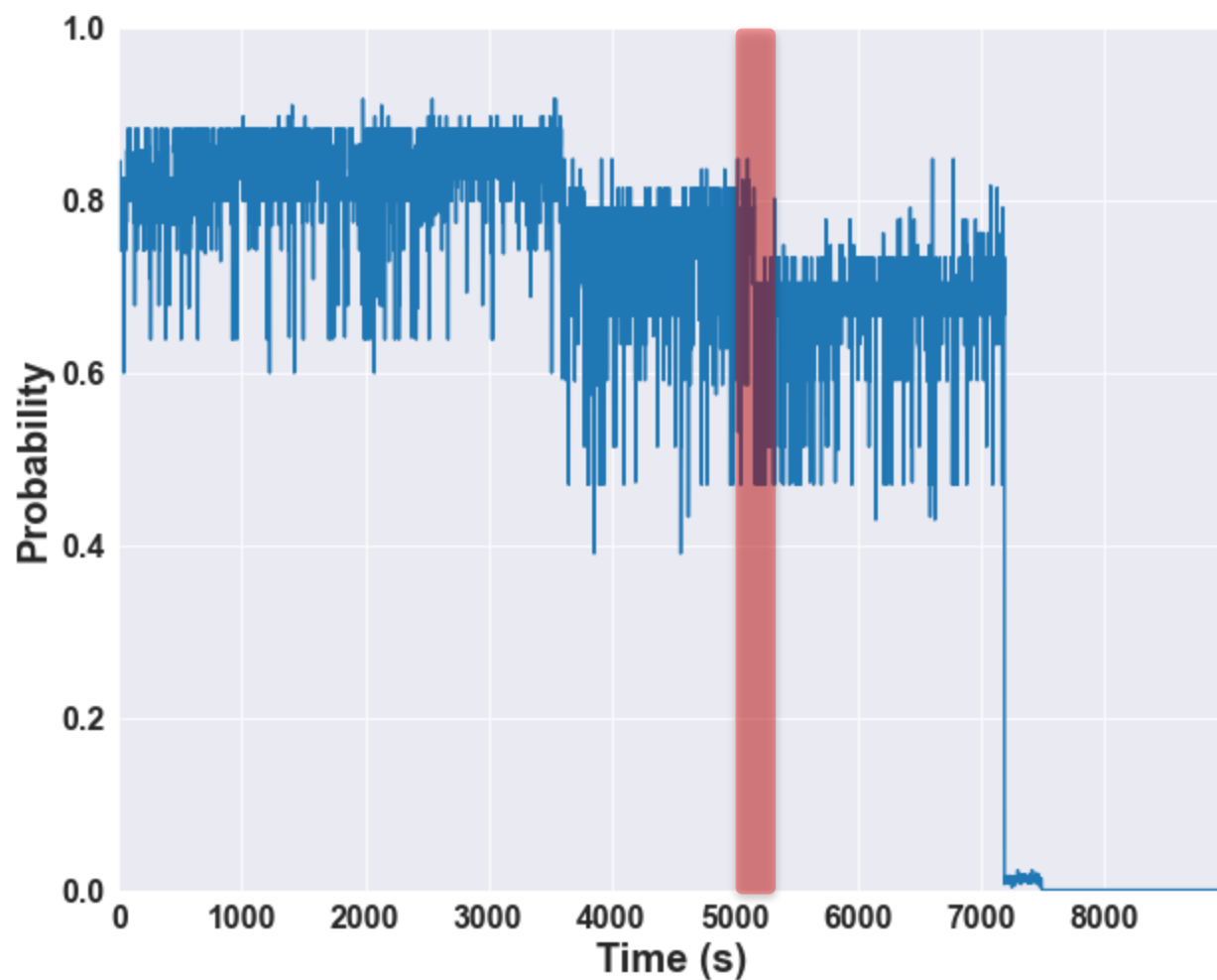


Fig. 6-10 P_{IN} related to file #19 included preictal and proceeding the interictal portions for Patient #12. The red bar is the period of the seizure.

Table 6-6. The result of two kinds of prediction time for 6 patients.

Time of prediction	Pa#2	Pa#4	Pa#7	Pa#10	Pa#12	Pa#16	The average with 95% confidence interval
for PA1, T_1 (min.)	59.6	70.2	68.4	88.7 & 80.3	85.1	67.1 & 84.9	75.5±7
for PA2, T_2 (min.)	55.9	70.1	43	82.7 & 38.8	55.6	15.6 & 33.1	49.4±14

Table 6-7. The overall Forecasting results using for 6 patients.

Results	Average for all the patients based on PA1, T_1	Average for all the patients based on PA2, T_2
Sensitivity	100%	100%
FPR	0.07%	Almost 0%
Anticipation Time	75.5 ± 7	49.4 ± 14

The obtained results were successfully compared to those already published (Table 6-8). As demonstrated, the proposed approach performs better with the highest sensitivity and the lowest false positive rate (figure 6-11). As mentioned earlier, our comparison will be based on the first set of alarm, PA1, and the anticipation time is considered T_1 .

With regard to Table 6-8, in [2], [15], [107] the classifiers were optimised based on one measure, i.e. accuracy, which is not, as discussed, a good measure for an imbalanced dataset. Furthermore, in [41] the performance of the classifier were not reported and if the performance of the classifier (other than accuracy) is low, then the result of the work really need to be reinvestigated.

Table 6-8. Our forecasting results using for 6 patients compared to the published works.

Reference, Year	Avg. prediction time	Feature used	Performance analyzing & classifier's performance	imbalance ratio (data length)	Database (patients, human/dog)	# seizures and # hours data
[107], 2019	0.25 minutes	Nonlinear features (which are computationally expensive)	sensitivity of 91% FP:36% (FP/TP was calculated and then compared & Accuracy	--	16 patients / Boston Hospital non-invasive EEG	The number of seizures is not mentioned Less than 60 hours data investigated

[15], 2018	Not reported	time and frequency domain	FP/h=0.03-0.6 TP/h=40-97% For human: TP/h:0.4-0.74 FP/h: 0.26-0.6 & Accuracy	From 2:1 to 20:1 677 hours (42 hours for human) 111 seizures (10 for humans)	MSEL-LAB 7 subjects (5 dogs and 2 humans)	MSEL and IEEG.ORG
[41], 2017	No information provided	time and frequency domain	Prediction can achieve a sensitivity of about 90–100%, and the false-positive rate of about 0–0.3 times per day. No performance classification is reported.	8:1 to 10:1 (607:77)	Kaggle competition (6 Dogs) canine epilepsy is an excellent analog for human epilepsy	20 seizures
[2], 2017	33	time and frequency domain	Sensitivity=92.2%, and specificity 93.38%. or 0.06 1/h & accuracy	Not mentioned	CHB-MIT, scalp	84 seizures, 22 subjects
[108], 2017	25.66	26 univariate and 3 bivariate features	Sensitivity 79% Specificity 82%	Not mentioned	10 patients 26 electrode iEEG	154 seizures 86.20 days data
Our work	Average time of 75 min	Time domain features	100% sensitivity and 0.07 false positive rate MCC : 0.582	1.6	the Freiburg Seizure Prediction EEG (FSPEEG) database	36 seizures 134 hours data

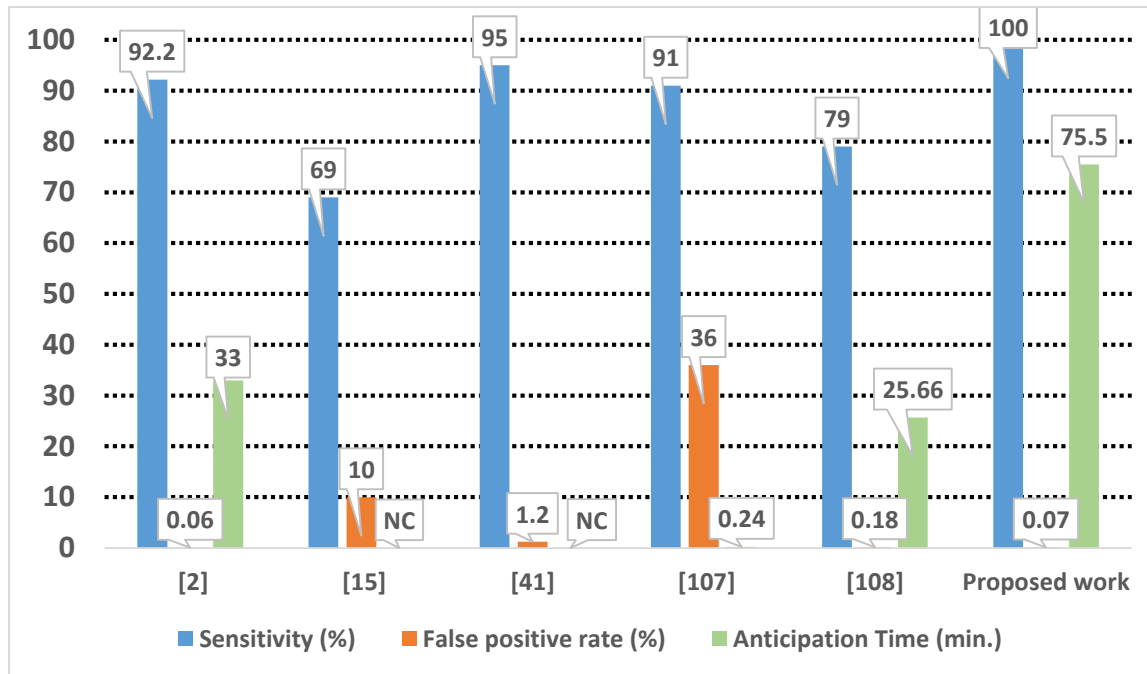


Fig. 6-11 Comparison of sensitivity, false positive rate and the prediction time (NC: not clear).

6.8 Conclusion

While finding a unique model for all patients is really challenging, our modelling results demonstrated that the proposed algorithm can be an efficient tool for reliable and clinically relevant seizure prediction. The presented XGBOOST-based system for seizure prediction uses iEEG signals and exhibits two original features. First, we found out that a 2s window showed the highest performance (MCC score) while assuring that the data did not suffer from imbalanced issue. Second, the previous published work had lower sensitivity, higher false positive rate while the proposed showed a sensitivity rate of 100% with no false alarm. As already mentioned, these results should be relativized because of the limited number of patients available in the used database.

Chapter 7

Conclusions and Future Work

7.1 Conclusions

Predicting of epileptic seizures remains an active research field with several unanswered questions that must be determined before deploying a commercial medical warning device to predict the seizure successfully. The provision for a warning medical device is widely demanding to enhance the life quality of more than 70 million individuals afflicted by this debilitating disease. The proposed probabilistic seizure prediction framework is one small part of this complex medical and social problem and could help doctors, scientists, and finally epileptic patients to tackle treating in a new era.

In this thesis, we introduced numerous features of EEG signals that can be considered in seizure prediction. We transformed the EEG samples to a proper feature space by extracting some useful features and measurements from such signals, and then to feed them to a classifier. Our aim was to implement an algorithm to reduce CPU time with simpler algorithm and less complexity for a low-cost computing device. As preliminary results, we introduced a probabilistic approach based on SVM classifier in which the data of the detection were used in the prediction. This approach later was completed with one of the most powerful classifiers, XBGboost.

Before introducing the novel probabilistic approach, we applied three principle feature selection methods on time domain features and the Kruskal-Wallis, a simple and less time consuming method compared to other approaches, showed the highest performance metric. Also, we found out that among the time-domain features investigated, the parameters of AR model are ranked as the top features for both lobes. The second most important set of features includes the maximum of cross-correlation and IQR for temporal and frontal lobes, respectively. Moreover, a high range of frequency like low and high-Gamma have been introduced as interesting sub-bands for the

temporal lobe epilepsy, while the middle range of frequencies from Theta to Beta can be seen as important ranges of frequency for frontal lobe epilepsy.

Implementing a unique model for all patients is really challenging, but our promising results demonstrated that the novel probabilistic XGBOOST-based system can be an efficient tool for reliable and clinically relevant seizure prediction. The result of the prediction showed a sensitivity and specificity of almost 100% and the average of the anticipation time was 75 minutes.

7.2 Future work

Based on Figure 1-1, the number of papers in the prediction of seizure is increasing while the unpredictable nature of seizures makes it an ever-growing research field. Many algorithms have been developed to better detect, analyze, and study the seizure and its mechanisms, like the Phase-Amplitude Coupling Measure [226], Phase and Amplitude Lock Values [227], Cross-Frequency Coupling [228] while there is still room to enhance or develop new techniques. To distinguish between the normal and abnormal synchronization of a neural activity, a coherence-based analysis can be proposed. Coherence is a measure that provides synchrony between pairs of brain regions while global coherence offers coordinated neural activity across multiple brain areas. Unlike pairwise coherence, global coherence gives a better understanding of neural synchronization across several brain regions due to rendering a higher coordinate spatial activity [229], [230]. Global coherence can be described based on the eigenvalues of a cross spectral matrix for a range of frequencies and, therefore, spectral analysis should be applied [231]. Global coherence can be defined as the ratio of the largest eigenvalue of the cross spectral matrix to the sum of its eigenvalues at a given frequency [231]–[233].

To the best of the author's knowledge, global coherence has not been used to study seizures, evaluate the potential of this measure in prediction of the seizure, as well as to explore the synchronous activities during and before the seizure for both lobes. The following preliminary results proved that this direction can be promising.

In fact, we illustrated the global coherence for both preictal and interictal over various frequencies in Figures 7-1 and 7-2. Figure 7-1 shows the global coherence of preictal (two non-consecutive hours) and interictal (one hour) portions for patient #2 with temporal epilepsy. Figures 7-1a and 7-1b demonstrate that there is a wide range of frequencies can be considered prior to the seizure whereas figure 7-1c helps us to limit this broad range.

Figure 7-2 shows the global coherence of preictal (3 consecutive hours) and interictal (one hour) portions for patient #1 with frontal lobe epilepsy. From figure 7-2 (a), (b), and (c) we can see there is a meaningful change in global coherence for the range of frequencies of 20-65 Hz before and during seizure. However, the results of the global coherence for non-seizure file, figure 7-2 (d), do not show the pattern in preictal files. We evaluated the potential of the global coherence in prediction of the seizure, as well as to discover the synchronous activities during and prior to the seizure. Finally, our investigation demonstrated that the global coherence can be a promising measure prior to the seizure for both lobes.

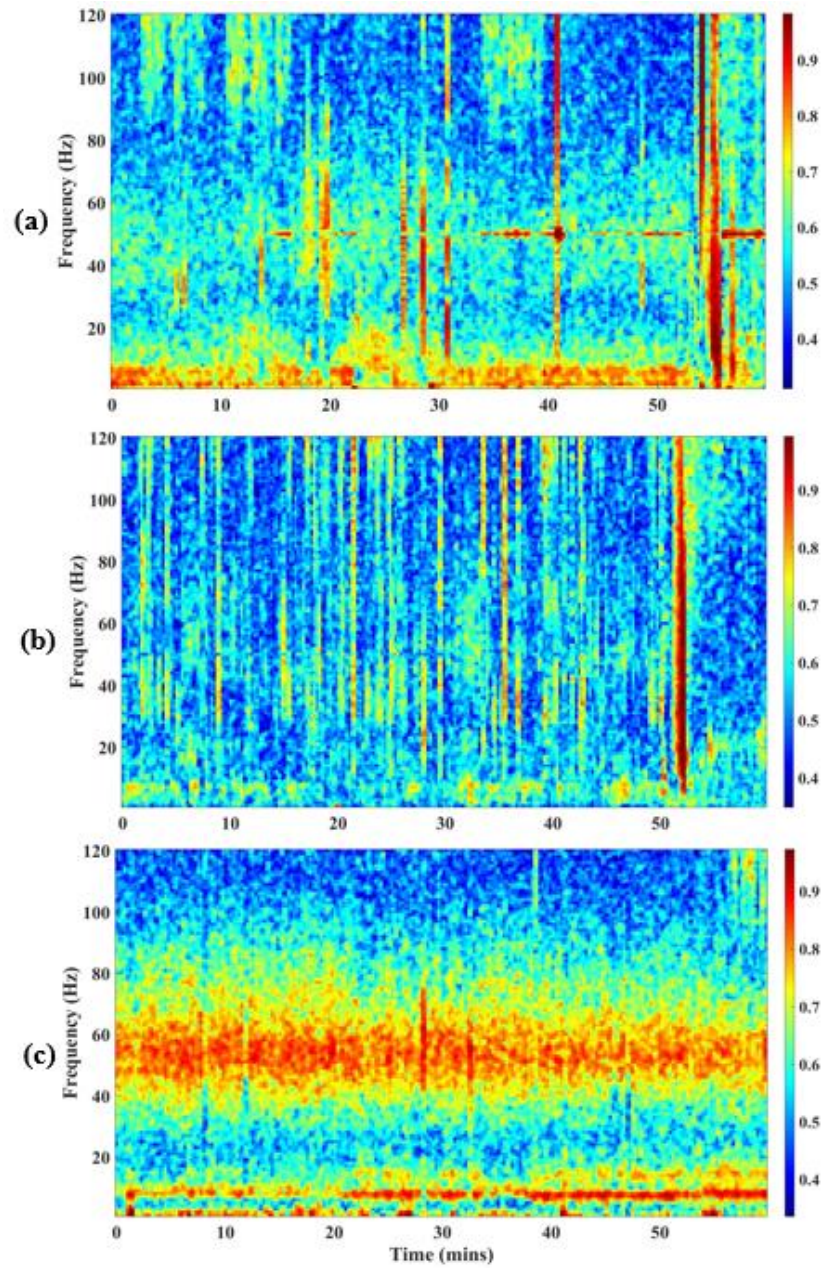


Figure 7-1 Global coherence for various frequencies of a temporal lobe epilepsy patient (patient # 2) a) one hour EEG signal contains the seizure from 53 to 56 minutes b) a seizure occurs between 50 to 52 minutes c) one hour interictal file.

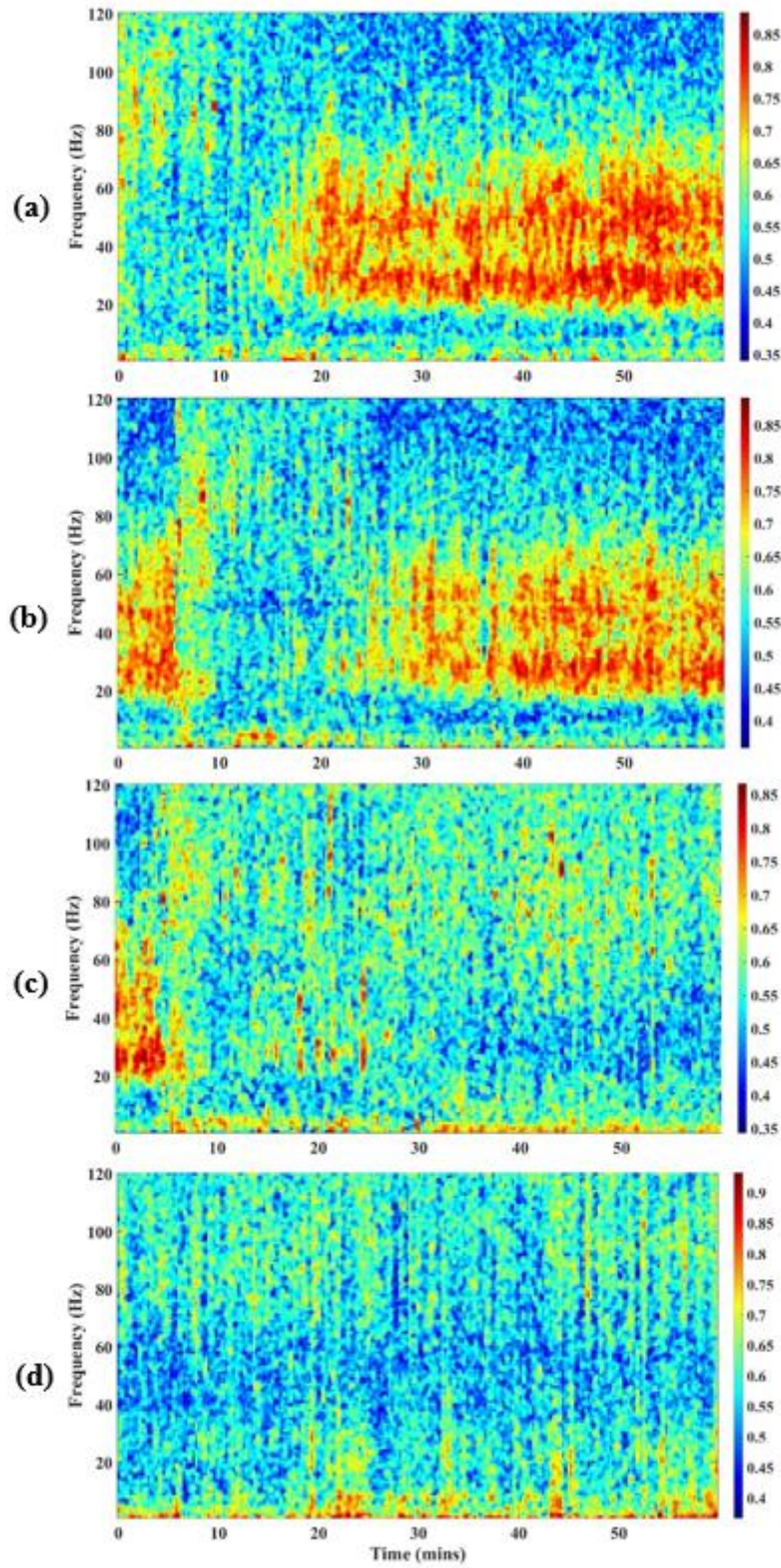


Fig 7-2. Global coherence for various frequencies of a frontal lobe epilepsy patient (patient # 1) a) one hour EEG signal prior to the seizure b) one hour data contains seizure from 356 to 375 seconds and precede a seizure c) A seizure occurs between 287- 305 seconds d) one hour interictal file.

From that, the following directions can be undertaken as future work to further investigate the outcomes of the obtained results:

- By comparing the performance results of Kruskal Wallis for both lobes, we found out that MCC is not close to 1 (perfect prediction case). Not having a high MCC value can imply the low capacity of the version of Freiburg database we used, due to having data up to 90 minutes of preictal. Based on Ramachandran [15], one might need at least 3 to 16 hours before the onset of the seizure to efficiently predict seizures. It may be considered as a limitation in this study and weakness of this database. In future, a newer version of the Freiburg dataset should be used to alleviate this issue.
- As future work, the window length (called forecast horizon in this work and of 1 minute long), can be further investigated to give the patient various kinds of options and warning time zones with a certain confidence for an impending seizure. The proposed approach can be implemented on a low power device due to considering few simple time-domain features, particularly when compared to other works that employed computationally expensive features. Also, this framework can be extended to an online semi-supervised seizure prediction in which the values of the future thresholds can be adjusted, based on the current and past parameters.
- Knowing that seizure patterns differ from patient to patient, and sometimes may need different approaches for different seizures with various origins, future efforts should be deployed to reliably improve the performance of the prediction on train and test set for a patient-specific.
- The MCC score for FLE patients was low compared to TLEs. Therefore, a new set of features that can characterize the dynamic of the brain network for this kind of epilepsy should be considered. Since there are not many publications investigating frontal lobe epilepsy then the findings in the work indicate promising results.
- We found that the Global Coherence is a promising measure prior to the seizure for TLEs and FLEs. Thus, this measure needs to be investigated further and compared to other measures such as cross-frequency coupling (CFC) and phase and amplitude lock values (PLV/ALV) to characterize the patterns of coordinated neural activity within and across the brain regions.

- Various experiments have been done with respect to mice subjects, to quantitatively tracking the brain-heart interactions (such as the frequency and the strength of the signals from brain affecting the heart) during and prior to the epilepsy [117]. In addition, in [111], the generalized partial directed coherence (GPDC) was applied to find the most useful range of the frequency spectrum that the brain and heart interact with each other. However, we still do not have a similar analysis on human database subjects. In future, further investigation by the GPDC analysis should be done on the heart-brain interactions for prediction of the seizure with the newest available database of epilepsy. Figure 7-3 could illustrate the kind of future work that should take in prediction of the seizure, by showing the relations between ECG and EEG signals.

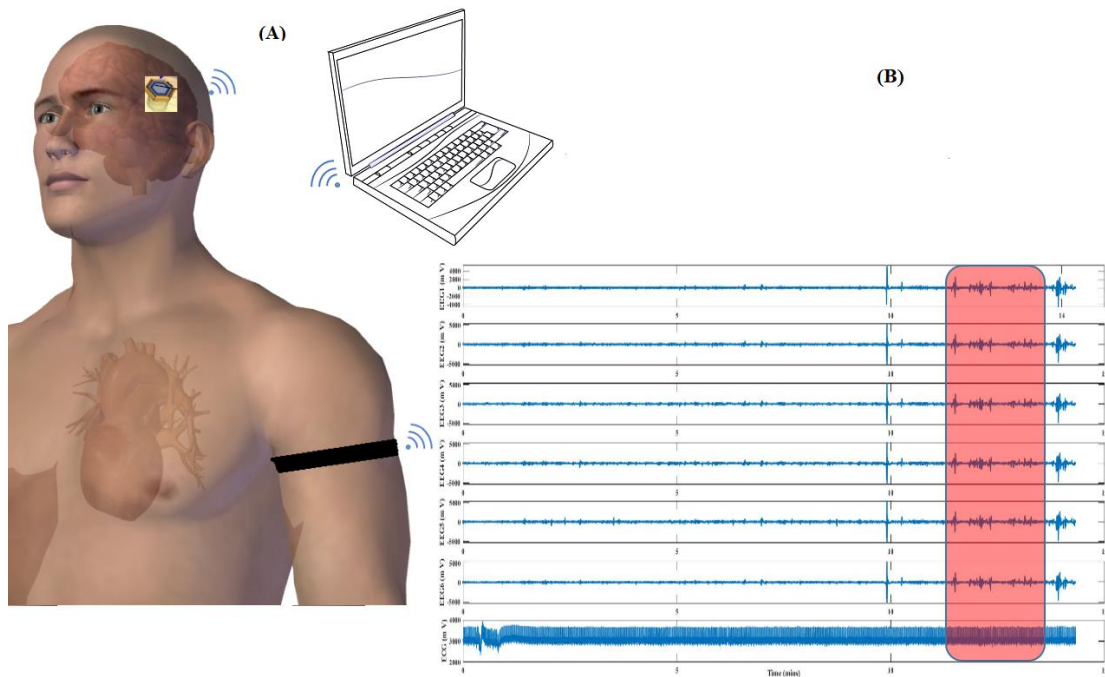


Figure 7-3: The interaction of EEG and ECG: (A) a device predict seizure based on ECG and EEG signals (B) the first top signals are related to EEG and the last one is the ECG signal. The heightened area shoes the seizure period.

- The first stage in deploying a machine learning model on a medical device is having access to an online wireless recording system with an implantable wireless ECoG or iEEG. Wireless ECoG electrodes utilized on a monkey [234] , a sheep [235], and recently, fully

implantable wireless ECoG electrodes for humans [236], [237] have already been designed. Figure 7-4 shows a wireless ECoG recorder, WIMAGINE (Wireless Implantable Multi-channel Acquisition system for Generic Interface with Neurons) implanted on the brain cortex of a sheep. In the next stage, a Brain Computer Interface (BCI) is required to receive and analyze the ECoG signal, and then to produce a particular output based on the proposed algorithm [236], [237]. This data and power management unit can be considered as an abdominal device and finally can be connected as external devices like Wi-Fi access point to communicate with an external base station [236].

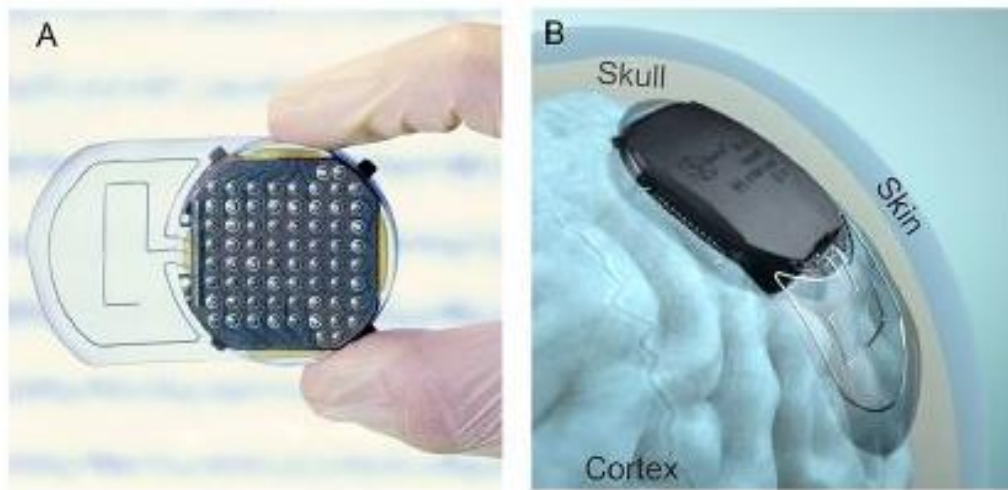


Figure 7-4: (A) View of the cortical electrode array. (B) Lateral implantation of a WIMAGINE implant on an anatomical model [235].

- Other than sending an alarm, the implanted device can perform as a trigger therapy. This system can deliver a small dose of fast acting antiepileptic drugs to cool down a certain part of the brain [237] or for the patients with drug-resistant epilepsy, deep brain stimulation can be a solution to cease the seizure instantaneously [225], [230]. Furthermore, with an electric stimulation we may try to reset the brain dynamics to a state which no seizure may be developed [136]. This is certainly another direction worth to explore. Finally, figure 7-5 shows an autonomous closed loop seizure prevention system with three essential elements: prediction, stimulation, and advisory sections. Therefore,

based on this work, the next step would be to create this system is studying the stimulation of electrical impulses of the defected brain region.

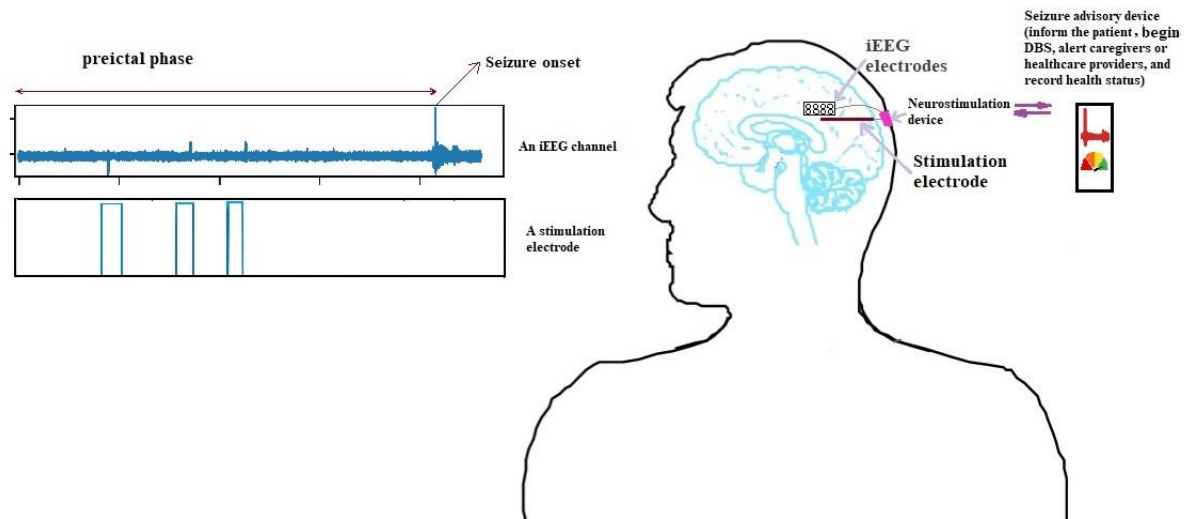


Figure 7-5: An illustration of an autonomous medical device with a closed loop seizure suppression system.

References

- [1] J. Ramakrishnan and B. R. Kanagaraj, "Analysis of non-seizure and seizure activity using intracranial EEG signals and empirical mode decomposition based approximate entropy," *Biomed. Res.*, vol. 2018, no. Special Issue ComputationalLifeSciencesandSmarterTechnologicalAdvancement, pp. S47–S51, 2018, doi: 10.4066/biomedicalresearch.29-16-2323.
- [2] S. M. Usman, M. Usman, and S. Fong, "Epileptic Seizures Prediction Using Machine Learning Methods," *Comput. Math. Methods Med.*, vol. 2017, 2017, doi: 10.1155/2017/9074759.
- [3] S. Tewolde, K. Oommen, D. Y. C. Lie, Y. Zhang, and M. C. Chyu, "Epileptic seizure detection and prediction based on continuous cerebral blood flow monitoring - A review," *J. Healthc. Eng.*, vol. 6, no. 2, pp. 159–178, Jun. 2015, doi: 10.1260/2040-2295.6.2.159.
- [4] L. Dalic and M. J. Cook, "Managing drug-resistant epilepsy: Challenges and solutions," *Neuropsychiatric Disease and Treatment*, vol. 12. Dove Medical Press Ltd., pp. 2605–2616, Oct. 12, 2016, doi: 10.2147/NDT.S84852.
- [5] M. Larmuseau, "Epileptic Seizure Prediction using Deep Learning," Master thesis, Ghent University, 2016.
- [6] A. Gupta, P. Singh, and M. Karlekar, "A novel signal modeling approach for classification of seizure and seizure-free EEG signals," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 26, no. 5, pp. 925–935, May 2018, doi: 10.1109/TNSRE.2018.2818123.
- [7] H. Khammari and A. Anwar, "A Spectral Based Forecasting Tool of Epileptic Seizures," *International Journal of Computer Science Issues* 9(3), 2012.
- [8] L. Kuhlmann, K. Lehnertz, M. P. Richardson, B. Schelter, and H. P. Zaveri, "Seizure prediction — ready for a new era," *Nature Reviews Neurology*, vol. 14, no. 10. Nature Publishing Group, pp. 618–630, Oct. 01, 2018, doi: 10.1038/s41582-018-0055-2.
- [9] B. Abbaszadeh and M. C. E. Yagoub, "Optimum Window Size and Overlap for Robust Probabilistic Prediction of Seizures with iEEG," in *2019 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology, CIBCB 2019*, Jul. 2019, pp. 1–5, doi: 10.1109/CIBCB.2019.8791450.
- [10] K. Tsiouris, V. C. Pezoulas, M. Zervakis, S. Konitsiotis, D. D. Koutsouris, and D. I. Fotiadis, "A Long Short-Term Memory deep learning network for the prediction of epileptic seizures using EEG signals," *Comput. Biol. Med.*, vol. 99, pp. 24–37, Aug. 2018, doi: 10.1016/j.combiomed.2018.05.019.
- [11] O. College, "Anatomy and Physiology - OpenStax," *Rice University*, 2013. <https://openstax.org/details/books/anatomy-and-physiology> (accessed Mar. 25, 2021).
- [12] J. Webster, *Medical instrumentation : application and design*, Fourth editon, Wiley, 2010.
- [13] M. Bandarabadi, C. A. Teixeira, J. Rasekhi, and A. Dourado, "Epileptic seizure prediction using relative spectral power features," *Clin. Neurophysiol.*, vol. 126, no. 2, pp. 237–248, 2015, doi:

10.1016/j.clinph.2014.05.022.

- [14] E. B. Assi, D. K. Nguyen, S. Rihana, and M. Sawan, "A functional-genetic scheme for seizure forecasting in canine epilepsy," *IEEE Trans. Biomed. Eng.*, vol. 65, no. 6, pp. 1339–1348, Jun. 2018, doi: 10.1109/TBME.2017.2752081.
- [15] V. R. K. Ramachandran, H. J. Alblas, D. V. Le, and N. Meratnia, "Towards an online seizure advisory system—An adaptive seizure prediction framework using active learning heuristics," *Sensors (Switzerland)*, vol. 18, no. 6, Jun. 2018, doi: 10.3390/s18061698.
- [16] R. S. Fisher *et al.*, "Epileptic seizures and epilepsy: Definitions proposed by the International League Against Epilepsy (ILAE) and the International Bureau for Epilepsy (IBE)," *Epilepsia*, vol. 46, no. 4, pp. 470–472, Apr. 2005, doi: 10.1111/j.0013-9580.2005.66104.x.
- [17] A. Korotkov, J. D. Mills, J. A. Gorter, E. A. Van Vliet, and E. Aronica, "Systematic review and meta-analysis of differentially expressed miRNAs in experimental and human temporal lobe epilepsy," *Scientific Reports*, vol. 7, no. 1, Nature Publishing Group, pp. 1–13, Dec. 01, 2017, doi: 10.1038/s41598-017-11510-8.
- [18] Y. Fu *et al.*, "Systems-level analysis identifies key regulators driving epileptogenesis in temporal lobe epilepsy," *Genomics*, vol. 112, no. 2, pp. 1768–1780, Mar. 2020, doi: 10.1016/j.ygeno.2019.09.020.
- [19] M. T. Venø *et al.*, "A systems approach delivers a functional microRNA catalog and expanded targets for seizure suppression in temporal lobe epilepsy," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 117, no. 27, pp. 15977–15988, Jul. 2020, doi: 10.1073/pnas.1919313117.
- [20] M. Alizadeh *et al.*, "Hemispheric Regional Based Analysis of Diffusion Tensor Imaging and Diffusion Tensor Tractography in Patients with Temporal Lobe Epilepsy and Correlation with Patient outcomes," *Sci. Rep.*, vol. 9, no. 1, pp. 1–8, Dec. 2019, doi: 10.1038/s41598-018-36818-x.
- [21] L. M. Gibson, M. F. Hanby, S. M. Al-Bachari, L. M. Parkes, S. M. Allan, and H. C. A. Emsley, "Late-onset epilepsy and occult cerebrovascular disease," *Journal of Cerebral Blood Flow and Metabolism*, vol. 34, no. 4, Nature Publishing Group, pp. 564–570, 2014, doi: 10.1038/jcbfm.2014.25.
- [22] M. N. LEVY, M. NG, P. MARTIN, H. ZIESKE, and T. Rogoff, "Sympathetic and Parasympathetic Interactions upon the Left Ventricle of the Dog," *Circ. Res.*, vol. 19, no. 1, pp. 5–10, Jul. 1966, doi: 10.1161/01.res.19.1.5.
- [23] T. Netoff, Y. Park, and K. Parhi, "Seizure prediction using cost-sensitive support vector machine," in *Proceedings of the 31st Annual International Conference of the IEEE Engineering in Medicine and Biology Society: Engineering the Future of Biomedicine, EMBC 2009*, 2009, pp. 3322–3325, doi: 10.1109/IEMBS.2009.5333711.
- [24] S.-H. Oh, Y.-R. Lee, and H.-N. Kim, "A Novel EEG Feature Extraction Method Using Hjorth Parameter," *Int. J. Electron. Electr. Eng.*, pp. 106–110, 2014, doi: 10.12720/ijeee.2.2.106-110.
- [25] A. De, A. Konar, A. Samanta, S. Biswas, and P. Basak, "Seizure prediction using low frequency EEG waves from WAG/Rij rats," in *2017 2nd International Conference for Convergence in Technology, I2CT 2017*, Dec. 2017, vol. 2017-January, pp. 244–249, doi: 10.1109/I2CT.2017.8226129.
- [26] E. Verche, C. San Luis, and S. Hernández, "Neuropsychology of frontal lobe epilepsy in children

- and adults: Systematic review and meta-analysis," *Epilepsy and Behavior*, vol. 88. Academic Press Inc., pp. 15–20, Nov. 01, 2018, doi: 10.1016/j.yebeh.2018.08.008.
- [27] B. Klugah-Brown *et al.*, "Altered structural and causal connectivity in frontal lobe epilepsy," *BMC Neurol.*, vol. 19, no. 1, p. 70, Apr. 2019, doi: 10.1186/s12883-019-1300-z.
 - [28] M. O. Baud, S. Vulliemoz, and M. Seeck, "Recurrent secondary generalization in frontal lobe epilepsy: Predictors and a potential link to surgical outcome?," *Epilepsia*, vol. 56, no. 9, pp. 1454–1462, Sep. 2015, doi: 10.1111/epi.13086.
 - [29] M. M. Siddiqui, G. Srivastava, and H. Saeed, "Diagnosis of Nocturnal Frontal Lobe Epilepsy (NFLE) sleep disorder using short time frequency analysis of PSD approach applied on EEG signal," *Biomed. Pharmacol. J.*, vol. 9, no. 1, pp. 393–403, 2016, doi: 10.13005/bpj/951.
 - [30] S. Yin *et al.*, "Complete prefrontal lobe isolation surgery for recurrent epilepsy: A case report," *Exp. Ther. Med.*, vol. 12, no. 5, pp. 3029–3034, Nov. 2016, doi: 10.3892/etm.2016.3714.
 - [31] F. Pisano *et al.*, "Convolutional Neural Network for Seizure Detection of Nocturnal Frontal Lobe Epilepsy," *Complexity*, vol. 2020, 2020, doi: 10.1155/2020/4825767.
 - [32] G. Busonera *et al.*, "EEG Spectral Coherence Analysis in Nocturnal Epilepsy," *IEEE Trans. Biomed. Eng.*, vol. 65, no. 12, pp. 2713–2719, Dec. 2018, doi: 10.1109/TBME.2018.2814479.
 - [33] B. Pisano *et al.*, "Autosomal dominant nocturnal frontal lobe epilepsy seizure characterization through wavelet transform of EEG records and self organizing maps," in *IEEE International Workshop on Machine Learning for Signal Processing, MLSP*, Nov. 2016, vol. 2016-November, doi: 10.1109/MLSP.2016.7738861.
 - [34] B. Direito *et al.*, "Feature selection in high dimensional EEG features spaces for epileptic seizure prediction," in *IFAC Proceedings Volumes (IFAC-PapersOnline)*, Jan. 2011, vol. 44, no. 1 PART 1, pp. 6206–6211, doi: 10.3182/20110828-6-IT-1002.03331.
 - [35] S. Priyanka, D. Dema, and T. Jayanthi, "Feature selection and classification of Epilepsy from EEG signal," in *2017 International Conference on Energy, Communication, Data Analytics and Soft Computing, ICECDS 2017*, Jun. 2018, pp. 2404–2406, doi: 10.1109/ICECDS.2017.8389880.
 - [36] B. Karlik and Ş. Hayta, "Comparison Machine Learning Algorithms for Recognition of Epileptic Seizures in EEG," *International Work-Conference on Bioinformatics and Biomedical Engineering Proceedings (IWBBIO 2014)*, Granada 7-9 April, 2014.
 - [37] T. Das, A. Ghosh, S. Guha, and P. Basak, "Classification of EEG Signals for Prediction of Seizure using Multi-Feature Extraction," *1st International Conference on Electronics, Materials Engineering and Nano-Technology (IEMENTech)*, Kolkata, India, Oct. 2017, doi: 10.1109/IEMENTECH.2017.8076992.
 - [38] M. Z. Parvez and M. Paul, "Epileptic seizure prediction by exploiting spatiotemporal relationship of EEG signals using phase correlation," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 24, no. 1, pp. 158–168, Jan. 2016, doi: 10.1109/TNSRE.2015.2458982.
 - [39] A. Shaikh and M. Dhopeswarkar, "A Novel Approach to Optimal Feature Selection for Epileptic Seizure Prediction," *IOSR J. Comput. Eng.*, vol. 13, pp. 13–16, 2013, Accessed: Apr. 13, 2021. [Online]. Available: www.iosrjournals.org.
 - [40] T. Wen and Z. Zhang, "Effective and extensible feature extraction method using genetic

- algorithm-based frequency-domain feature search for epileptic EEG multiclassification," *Med. (United States)*, vol. 96, no. 19, May 2017, doi: 10.1097/MD.00000000000006879.
- [41] H. T. Shiao *et al.*, "SVM-Based System for Prediction of Epileptic Seizures from iEEG Signal," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 5, pp. 1011–1022, May 2017, doi: 10.1109/TBME.2016.2586475.
 - [42] S. Farahmand, T. Sobayo, and D. J. Mogul, "Noise-Assisted Multivariate EMD-Based Mean-Phase Coherence Analysis to Evaluate Phase-Synchrony Dynamics in Epilepsy Patients," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 26, no. 12, pp. 2270–2279, Dec. 2018, doi: 10.1109/TNSRE.2018.2881606.
 - [43] P. Patrikelis, E. Angelakis, and S. Gatzonis, "Neurocognitive and behavioral functioning in frontal lobe epilepsy: A review," *Epilepsy and Behavior*, vol. 14, no. 1, Epilepsy Behav, pp. 19–26, Jan. 2009, doi: 10.1016/j.yebeh.2008.09.013.
 - [44] J. Klatt *et al.*, "The EPILEPSIAE database: An extensive electroencephalography database of epilepsy patients," *Epilepsia*, vol. 53, no. 9, pp. 1669–1676, Sep. 2012, doi: 10.1111/j.1528-1167.2012.03564.x.
 - [45] P. Fergus, A. Hussain, D. Hignett, D. Al-Jumeily, K. Abdel-Aziz, and H. Hamdan, "A machine learning system for automated whole-brain seizure detection," *Appl. Comput. Informatics*, vol. 12, no. 1, pp. 70–89, Jan. 2016, doi: 10.1016/j.aci.2015.01.001.
 - [46] A. Sharma, J. K. Rai, and R. P. Tewari, "Epileptic seizure anticipation and localisation of epileptogenic region using EEG signals," *J. Med. Eng. Technol.*, vol. 42, no. 3, pp. 203–216, Apr. 2018, doi: 10.1080/03091902.2018.1464074.
 - [47] N. Mohan, P. P. Muhammed Shanir, N. Sulthan, K. A. Khan, and S. Sofiya, "Automatic epileptic seizure prediction in scalp EEG," in *Proceedings - 2nd International Conference on Intelligent Circuits and Systems, ICICS 2018*, Oct. 2018, pp. 281–285, doi: 10.1109/ICICS.2018.00063.
 - [48] L. K. McCorry, "Physiology of the autonomic nervous system," *Am. J. Pharm. Educ.*, vol. 71, no. 4, 2007, doi: 10.5688/aj710478.
 - [49] B. Scupham, "Boundless Anatomy and Physiology," *LUMEN LEARNING*. <https://courses.lumenlearning.com/boundless-ap/> (accessed Apr. 09, 2021).
 - [50] S. Selim, M. Tantawi, H. Shedeed, and A. Badr, "Reducing execution time for real-time motor imagery based BCI systems," in *Advances in Intelligent Systems and Computing*, Oct. 2017, vol. 533, pp. 555–565, doi: 10.1007/978-3-319-48308-5_53.
 - [51] T. Haddad, N. Ben-Hamida, L. Talbi, A. Lakhssassi, and S. Aouini, "Temporal epilepsy seizures monitoring and prediction using cross-correlation and chaos theory," *Healthc. Technol. Lett.*, vol. 1, no. 1, pp. 45–50, Jan. 2014, doi: 10.1049/htl.2013.0010.
 - [52] A. K. Tafreshi, A. M. Nasrabadi, and A. H. Omidvarnia, "Empirical mode decomposition in epileptic seizure prediction," in *Proceedings of the 8th IEEE International Symposium on Signal Processing and Information Technology, ISSPIT 2008*, 2008, pp. 275–280, doi: 10.1109/ISSPIT.2008.4775729.
 - [53] D. A. Singh and O. Aktas, "The Window Size for Classification of Epileptic Seizures based on Analysis of EEG Patterns," 2016. Accessed: Mar. 25, 2021. [Online]. Available: <http://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-186439>.

- [54] O. Panichev, A. Popov, and V. Kharytonov, "Patient-specific epileptic seizure prediction using correlation features," 2015 Signal Processing Symposium (SPSymposium), Debe, Poland, Jul. 2015, doi: 10.1109/SPS.2015.7168309.
- [55] X. Wei, L. Zhou, Z. Zhang, Z. Chen, and Y. Zhou, "Early prediction of epileptic seizures using a long-term recurrent convolutional network," *J. Neurosci. Methods*, vol. 327, p. 108395, Nov. 2019, doi: 10.1016/j.jneumeth.2019.108395.
- [56] Y. Yang *et al.*, "Epileptic seizure prediction based on permutation entropy," *Front. Comput. Neurosci.*, vol. 12, Jul. 2018, doi: 10.3389/fncom.2018.00055.
- [57] F. Moslehi and A. Haeri, "An evolutionary computation-based approach for feature selection," *J. Ambient Intell. Humaniz. Comput.*, vol. 11, no. 9, pp. 3757–3769, Sep. 2020, doi: 10.1007/s12652-019-01570-1.
- [58] "EEG Database — Seizure Prediction Project Freiburg," 2012. <http://epilepsy.uni-freiburg.de/freiburg-seizure-prediction-project/eeg-database> (accessed Mar. 25, 2021).
- [59] M. K. Islam, A. Rastegarnia, and Z. Yang, "A Wavelet-Based Artifact Reduction from Scalp EEG for Epileptic Seizure Detection," *IEEE J. Biomed. Heal. Informatics*, vol. 20, no. 5, pp. 1321–1332, Sep. 2016, doi: 10.1109/JBHI.2015.2457093.
- [60] Edward E Smith; Stephen Michael Kosslyn, *Cognitive psychology : mind and brain*. Upper Saddle River, N.J.: Pearson/Prentice Hall., 2007.
- [61] J. Rosenow, "Anatomy of the nervous system," Elsevier, Boston, MA, 2009. Accessed: Apr. 09, 2021. [Online]. Available: <https://www.scholars.northwestern.edu/en/publications/anatomy-of-the-nervous-system-2>.
- [62] R. Stuffelbeam, "Neurons, Synapses, Action Potentials- The Mind Project," 2006. https://mind.ilstu.edu/curriculum/neurons_intro/ (accessed Apr. 09, 2021).
- [63] M. Kutz, *Standard Handbook of Biomedical Engineering & Design*. McGraw-Hill Education, 2003.
- [64] H. Lodish, A. Berk, S. L. Zipursky, P. Matsudaira, D. Baltimore, and J. Darnell, "The Action Potential and Conduction of Electric Impulses," in *Molecular Cell Biology*, 4th editio., W. H. Freeman, 2000.
- [65] Josef Fontana *et al.*, "Introduction to the Central Nervous System," *University Development Fund (FRVŠ)*, 2018. <http://fbilt.cz/en/skripta/regulacni-mechanismy-2-nervova-regulace/4-synapticky-prenos/> (accessed Apr. 09, 2021).
- [66] D. S. Faber and A. E. Pereda, "Two forms of electrical transmission between neurons," *Frontiers in Molecular Neuroscience*, vol. 11. Frontiers Media S.A., p. 427, Nov. 21, 2018, doi: 10.3389/fnmol.2018.00427.
- [67] H. Jasper, "The Ten Twenty Electrode System of the International Federation.," *Electroencephalogr. Clin. Neurophysiol.*, vol. 10, pp. 371–375, 1958, Accessed: Apr. 09, 2021. [Online]. Available: [https://www.scirp.org/\(S\(i43dyn45teexjx455qlt3d2q\)\)/reference/ReferencesPapers.aspx?ReferenceID=1192064](https://www.scirp.org/(S(i43dyn45teexjx455qlt3d2q))/reference/ReferencesPapers.aspx?ReferenceID=1192064).
- [68] O. Väisänen, "Multichannel EEG methods to improve the spatial resolution of cortical potential distribution and the signal quality of deep brain sources," PhD thesis, Tampere University of

Technology, 2008.

- [69] G. Ramantani, L. Maillard, and L. Koessler, "Correlation of invasive EEG and scalp EEG," *Seizure*, vol. 41. W.B. Saunders Ltd, pp. 196–200, Oct. 01, 2016, doi: 10.1016/j.seizure.2016.05.018.
- [70] A. K. Shah and S. Mittal, "Invasive electroencephalography monitoring: Indications and presurgical planning," *Ann. Indian Acad. Neurol.*, vol. 17, no. SUPPL. 1, 2014, doi: 10.4103/0972-2327.128668.
- [71] W. Stacey, M. Le Van Quyen, F. Mormann, and A. Schulze-Bonhage, "What is the present-day EEG evidence for a preictal state?," *Epilepsy Res.*, vol. 97, no. 3, pp. 243–251, Dec. 2011, doi: 10.1016/j.epilepsyres.2011.07.012.
- [72] Y. Nagahama *et al.*, "Intracranial EEG for seizure focus localization: Evolving techniques, outcomes, complications, and utility of combining surface and depth electrodes," *J. Neurosurg.*, vol. 130, no. 4, pp. 1180–1192, Apr. 2019, doi: 10.3171/2018.1.JNS171808.
- [73] S. Makeig and J. Onton, "ERP Features and EEG Dynamics: An ICA Perspective," in *The Oxford Handbook of Event-Related Potential Components*, Oxford University Press, 2012.
- [74] A. H. Shoeb, "Patient-Specific Seizure Onset Detection," Master thesis, Massachusetts Institute of Technology, 2003.
- [75] L. F. Q. Jerrold S. Meyer, *Psychopharmacology: Drugs, the Brain, and Behavior*, 3rd ed. Oxford University Press, 2018.
- [76] C. E. Stafstrom and L. Carmant, "Seizures and epilepsy: An overview for neuroscientists," *Cold Spring Harb. Perspect. Biol.*, vol. 7, no. 5, pp. 1–19, 2015, doi: 10.1101/cshperspect.a022426.
- [77] G. Baier, M. Goodfellow, P. N. Taylor, Y. Wang, and D. J. Garry, "The importance of modeling epileptic seizure dynamics as spatio-temporal patterns," *Front. Physiol.*, vol. 3 JUL, 2012, doi: 10.3389/fphys.2012.00281.
- [78] "Hiroshima MEG." <http://meg.aalip.jp/vsEEG/vsEEGE.html> (accessed Apr. 13, 2021).
- [79] A. Tanner, "Automatic seizure detection using a two-dimensional EEG feature space," Master thesis, Aalto University, 2011.
- [80] J. P. Cunha, "Epilepsy (Seizures): Learn about Facts and Treatments." https://www.medicinenet.com/seizure/article.htm#epilepsy_facts (accessed Apr. 13, 2021).
- [81] N. Sinha *et al.*, "Predicting neurosurgical outcomes in focal epilepsy patients using computational modelling," *Brain*, vol. 140, no. 2, pp. 319–332, Feb. 2017, doi: 10.1093/brain/aww299.
- [82] S. J. M. Smith, "EEG in the diagnosis, classification, and management of patients with epilepsy," *Neurology in Practice*, vol. 76, no. 2. BMJ Publishing Group, pp. 2–7, Jun. 2005, doi: 10.1136/jnnp.2005.069245.
- [83] M. J. Shen and D. P. Zipes, "Role of the autonomic nervous system in modulating cardiac arrhythmias," *Circulation Research*, vol. 114, no. 6. Circ Res, pp. 1004–1021, Mar. 14, 2014, doi: 10.1161/CIRCRESAHA.113.302549.
- [84] Z. Chen, P. Venkat, D. Seyfried, M. Chopp, T. Yan, and J. Chen, "Brain-Heart Interaction: Cardiac Complications after Stroke," *Circulation Research*, vol. 121, no. 4. Lippincott Williams and Wilkins,

- pp. 451–468, Aug. 04, 2017, doi: 10.1161/CIRCRESAHA.117.311170.
- [85] R. E. Klabunde, *Cardiovascular Physiology Concepts*, Second edition, Lippincott Williams & Wilkins, 2012.
 - [86] M. N. Levy, “Sympathetic-parasympathetic interactions in the heart,” *Circulation research*, vol. 29, no. 5. Circ Res, pp. 437–445, 1971, doi: 10.1161/01.RES.29.5.437.
 - [87] J. B. Martins and D. P. Zipes, “Effects of sympathetic and vagal nerves on recovery properties of the endocardium and epicardium of the canine left ventricle,” *Circ. Res.*, vol. 46, no. 1, pp. 100–110, 1980, doi: 10.1161/01.RES.46.1.100.
 - [88] B. -X Yuan, J. L. Ardell, D. A. Hopkins, A. M. Losier, and J. A. Armour, “Gross and microscopic anatomy of the canine intrinsic cardiac nervous system,” *Anat. Rec.*, vol. 239, no. 1, pp. 75–87, 1994, doi: 10.1002/ar.1092390109.
 - [89] J. A. Armour, D. A. Murphy, B. X. Yuan, S. Macdonald, and D. A. Hopkins, “Gross and microscopic anatomy of the human intrinsic cardiac nervous system,” *Anat. Rec.*, vol. 247, no. 2, pp. 289–298, Feb. 1997, doi: 10.1002/(SICI)1097-0185(199702)247:2<289::AID-AR15>3.0.CO;2-L.
 - [90] N. Assefa and Y. Tsige, “Human Anatomy and Physiology,” 2003.
 - [91] J. L. Dennis Kasper, Anthony Fauci, Stephen Hauser, Dan Longo, J. Larry Jameson, *Harrison’s Manual of Medicine*, 19th ed. McGraw-Hill Medical, 2016.
 - [92] R. D. Dawn Collins, John Goodfellow, Dulanka Silva and S. Nagaraja, “Neurology & Neurosurgery eureka,” 2018.
 - [93] J. F. Thayer and R. D. Lane, “Claude Bernard and the heart-brain connection: Further elaboration of a model of neurovisceral integration,” *Neuroscience and Biobehavioral Reviews*, vol. 33, no. 2. Neurosci Biobehav Rev, pp. 81–88, Feb. 2009, doi: 10.1016/j.neubiorev.2008.08.004.
 - [94] G. Ernst, “Heart-Rate Variability—More than Heart Beats?,” *Front. Public Heal.*, vol. 5, p. 1, Sep. 2017, doi: 10.3389/fpubh.2017.00240.
 - [95] K. Fujiwara *et al.*, “Epileptic Seizure Prediction Based on Multivariate Statistical Process Control of Heart Rate Variability Features,” *IEEE Trans. Biomed. Eng.*, vol. 63, no. 6, pp. 1321–1332, Jun. 2016, doi: 10.1109/TBME.2015.2512276.
 - [96] van der Kruijs SJM *et al.*, “Autonomic Nervous System Functioning Associated with Epileptic Seizures: Analysis of Heart Rate Variability,” *J. Neurol. Neurophysiol.*, vol. 5, no. 4, 2014, doi: 10.4172/2155-9562-5-1000215.
 - [97] E. N. Marieb, *Essentials of Human Anatomy & Physiology*, 11th ed. Pearson, 2015.
 - [98] D. J. Dalziel, B. M. Uthman, S. P. McGorray, and R. L. Reep, “Seizure-alert dogs: A review and preliminary study,” *Seizure*, vol. 12, no. 2, pp. 115–120, 2003, doi: 10.1016/S105913110200225X.
 - [99] V. Strong, S. Brown, M. Huyton, and H. Coyle, “Effect of trained Seizure Alert Dogs® on frequency of tonic-clonic seizures,” *Seizure*, vol. 11, no. 6, pp. 402–405, 2002, doi: 10.1053/seiz.2001.0656.
 - [100] V. Strong, S. W. Brown, and R. Walker, “Seizure-alert dogs: Fact or fiction?,” *Seizure*, vol. 8, no. 1, pp. 62–65, 1999, doi: 10.1053/seiz.1998.0250.
 - [101] M. Zijlmans, D. Flanagan, and J. Gotman, “Heart Rate Changes and ECG Abnormalities During

- Epileptic Seizures: Prevalence and Definition of an Objective Clinical Sign," *Epilepsia*, vol. 43, no. 8, pp. 847–854, Aug. 2002, doi: 10.1046/j.1528-1157.2002.37801.x.
- [102] K. B. Nilsen, M. Haram, S. Tangedal, T. Sand, and E. Brodtkorb, "Is elevated pre-ictal heart rate associated with secondary generalization in partial epilepsy?," *Seizure*, vol. 19, no. 5, pp. 291–295, 2010, doi: 10.1016/j.seizure.2010.03.003.
 - [103] T. H. Schwartz, S. B. Hong, A. P. Bagshaw, P. Chauvel, and C. G. Bénar, "Preictal changes in cerebral haemodynamics: Review of findings and insights from intracerebral EEG," *Epilepsy Res.*, vol. 97, no. 3, pp. 252–266, Dec. 2011, doi: 10.1016/j.eplepsyres.2011.07.013.
 - [104] K. S. Patel, M. Zhao, H. Ma, and T. H. Schwartz, "Imaging preictal hemodynamic changes in neocortical epilepsy," *Neurosurg. Focus*, vol. 34, no. 4, Apr. 2013, doi: 10.3171/2013.1.FOCUS12408.
 - [105] H. Firpi, E. Goodman, and J. Echauz, "Genetic programming artificial features with applications to epileptic seizure prediction," in *Annual International Conference of the IEEE Engineering in Medicine and Biology - Proceedings*, 2005, vol. 7 VOLS, pp. 4510–4513, doi: 10.1109/iembs.2005.1615471.
 - [106] M. K. Hasan, M. A. Ahamed, M. Ahmad, and M. A. Rashid, "Prediction of Epileptic Seizure by Analysing Time Series EEG Signal Using k -NN Classifier," *Appl. Bionics Biomech.*, vol. 2017, 2017, doi: 10.1155/2017/6848014.
 - [107] Z. S. Zaghloul and M. Bayoumi, "Early Prediction of Epilepsy Seizures VLSI BCI System," *arXiv*, Jun. 2019, Accessed: Mar. 25, 2021. [Online]. Available: <http://arxiv.org/abs/1906.02894>.
 - [108] C. Xiao, S. Wang, L. Iasemidis, S. Wong, and W. A. Chaovalitwongse, "An Adaptive Pattern Learning Framework to Personalize Online Seizure Prediction," *IEEE Trans. Big Data*, pp. 1–1, Mar. 2017, doi: 10.1109/tbdata.2017.2675982.
 - [109] M. K. Moridani and H. Farhadi, "Heart rate variability as a biomarker for epilepsy seizure prediction," *Bratislava Med. J.*, vol. 118, no. 01, pp. 3–8, Jan. 2017, doi: 10.4149/BLL_2017_001.
 - [110] K. Hoyos-Osorio, J. Castaneda-Gonzalez, and G. Daza-Santacoloma, "Automatic epileptic seizure prediction based on scalp EEG and ECG signals," 2016 XXI Symposium on Signal Processing, Images and Artificial Vision (STSIVA), Bucaramanga, Colombia, Nov. 2016, doi: 10.1109/STSIVA.2016.7743357.
 - [111] T. N. Hutson, F. Rezaei, N. M. Gautier, J. Indumathy, E. Glasscock, and L. Iasemidis, "Directed Connectivity Analysis of the Neuro-Cardio- and Respiratory Systems Reveals Novel Biomarkers of Susceptibility to SUDEP," *IEEE Open J. Eng. Med. Biol.*, vol. 1, pp. 301–311, Nov. 2020, doi: 10.1109/ojemb.2020.3036544.
 - [112] O. Van Der Stelt and A. Belger, "Application of electroencephalography to the study of cognitive and brain functions in schizophrenia," in *Schizophrenia Bulletin*, Jun. 2007, vol. 33, no. 4, pp. 955–970, doi: 10.1093/schbul/sbm016.
 - [113] D. Pandis and N. Scarmeas, "Seizures in alzheimer disease: Clinical and epidemiological data," *Epilepsy Curr.*, vol. 12, no. 5, pp. 184–187, Sep. 2012, doi: 10.5698/1535-7511-12.5.184.
 - [114] P. M. R. Reis, F. Hebenstreit, F. Gabsteiger, V. von Tschärner, and M. Lochmann, "Methodological aspects of EEG and body dynamics measurements during motion," *Front. Hum. Neurosci.*, vol. 8,

- no. MAR, Mar. 2014, doi: 10.3389/fnhum.2014.00156.
- [115] G. Alba, J. Vila, B. Rey, P. Montoya, and M. Á. Muñoz, "The Relationship Between Heart Rate Variability and Electroencephalography Functional Connectivity Variability Is Associated With Cognitive Flexibility," *Front. Hum. Neurosci.*, vol. 13, p. 64, Feb. 2019, doi: 10.3389/fnhum.2019.00064.
 - [116] S. Ramgopal *et al.*, "Seizure detection, seizure prediction, and closed-loop warning systems in epilepsy," *Epilepsy and Behavior*, vol. 37. Academic Press Inc., pp. 291–307, 2014, doi: 10.1016/j.yebeh.2014.06.023.
 - [117] F. Bahari, P. Ssentongo, S. J. Schiff, and B. J. Gluckman, "A brain–heart biomarker for epileptogenesis," *J. Neurosci.*, vol. 38, no. 39, pp. 8473–8483, Sep. 2018, doi: 10.1523/JNEUROSCI.1130-18.2018.
 - [118] "European database on epilepsy Epilepsiae." http://www.epilepsiae.eu/project_outputs/european_database_on_epilepsy (accessed Apr. 14, 2021).
 - [119] M. K. Islam, "Artifact Characterization, Detection and Removal from Neural Signals," National University of Singapore, 2015.
 - [120] P. A. Abhang, B. W. Gawali, and S. C. Mehrotra, *Introduction to EEG- and Speech-Based Emotion Recognition*. Elsevier Inc., 2016.
 - [121] I. S. Isa, B. S. Zainuddin, Z. Hussain, and S. N. Sulaiman, "Preliminary study on analyzing EEG alpha brainwave signal activities based on visual stimulation," in *Procedia Computer Science*, Jan. 2014, vol. 42, no. C, pp. 85–92, doi: 10.1016/j.procs.2014.11.037.
 - [122] J. Malmivuo and R. Plonsey, *Bioelectromagnetism: Principles and Applications of Bioelectric and Biomagnetic Fields*. Oxford University Press, 2012.
 - [123] A. H. Mawia A Hassan, Elwathiq A Mahmoud, Abdalla and and A. M. Wedaa, "A Comparison between Windowing FIR Filters for Extracting the EEG Components," *J. Biosens. Bioelectron.*, 2015, doi: 10.4172/2155-6210.1000191.
 - [124] S. J. Orfanidis, *Introduction to Signal Processing - Download link*. Prentice Hall, 2009.
 - [125] Y. Park, L. Luo, K. K. Parhi, and T. Netoff, "Seizure prediction with spectral power of EEG using cost-sensitive support vector machines," *Epilepsia*, vol. 52, no. 10, pp. 1761–1770, Oct. 2011, doi: 10.1111/j.1528-1167.2011.03138.x.
 - [126] S. Roy, M. Yadav, S. Roy, and P. R. Gautam, "QRS COMPLEX DETECTION AND FILTERING OF ECG SIGNAL USING WAVELET TRANSFORM," *Int. J. Comput. Eng. Appl.*, Accessed: Apr. 14, 2021. [Online]. Available: www.ijcea.com.
 - [127] B. Direito, C. Teixeira, M. Bandarabadi, F. Sales, and A. Dourad, "Automatic warning of epileptic seizures by SVM: The long road ahead to success," in *IFAC Proceedings Volumes (IFAC-PapersOnline)*, Jan. 2014, vol. 19, no. 3, pp. 1158–1163, doi: 10.3182/20140824-6-za-1003.00658.
 - [128] N. Mahmoodian, A. Boese, M. Friebe, and J. Haddadnia, "Epileptic seizure detection using cross-bispectrum of electroencephalogram signal," *Seizure*, vol. 66, pp. 4–11, Mar. 2019, doi: 10.1016/j.seizure.2019.02.001.

- [129] Y. Paul, "Various epileptic seizure detection techniques using biomedical signals: a review," *Brain Informatics*, vol. 5, no. 2. Springer Berlin Heidelberg, p. 6, Dec. 01, 2018, doi: 10.1186/s40708-018-0084-z.
- [130] G. Wu, Z. Li, Y. Zhang, X. Dong, and L. Ye, "Study of feature extraction algorithms for epileptic seizure prediction based on SVM," in *Lecture Notes in Electrical Engineering*, Jul. 2019, vol. 463, pp. 2370–2377, doi: 10.1007/978-981-10-6571-2_289.
- [131] F. Mormann *et al.*, "On the predictability of epileptic seizures," *Clin. Neurophysiol.*, vol. 116, no. 3, pp. 569–587, Mar. 2005, doi: 10.1016/j.clinph.2004.08.025.
- [132] W. Dargie, "Analysis of time and frequency domain features of accelerometer measurements," 2009, doi: 10.1109/ICCCN.2009.5235366.
- [133] D. Tkach, H. Huang, and T. A. Kuiken, "Study of stability of time-domain features for electromyographic pattern recognition," *J. Neuroeng. Rehabil.*, vol. 7, no. 1, p. 21, May 2010, doi: 10.1186/1743-0003-7-21.
- [134] B. V K, "Multiparameter index for the estimation of Depth of Anaesthesia based on Signal processing and Soft computing techniques," *University*, 2012, Accessed: Apr. 15, 2021. [Online]. Available: <http://hdl.handle.net/10603/221775>.
- [135] N. Moghim and D. W. Corne, "Predicting epileptic seizures in advance," *PLoS One*, vol. 9, no. 6, Jun. 2014, doi: 10.1371/journal.pone.0099334.
- [136] F. Mormann, R. G. Andrzejak, C. E. Elger, and K. Lehnertz, "Seizure prediction: The long and winding road," *Brain*, vol. 130, no. 2. Oxford University Press, pp. 314–333, Feb. 01, 2007, doi: 10.1093/brain/awl241.
- [137] P. Balakrishnan, S. Hemalatha, and D. N. S. Keshav, "Detection of Startle-Type Epileptic Seizures using Machine Learning Technique," *Int. J. Epilepsy*, vol. 5, no. 2, pp. 92–98, Oct. 2018, doi: 10.1055/s-0039-1693072.
- [138] M. Bedeuzzaman, O. Farooq, and Y. U Khan, "Automatic Seizure Detection using Inter Quartile Range," *Int. J. Comput. Appl.*, vol. 44, no. 11, pp. 1–5, Apr. 2012, doi: 10.5120/6304-8614.
- [139] W. van Drongelen, *Signal Processing for Neuroscientists*, Second. ScienceDirect, 2018.
- [140] U. R. Acharya, S. Vinitha Sree, G. Swapna, R. J. Martis, and J. S. Suri, "Automated EEG analysis of epilepsy: A review," *Knowledge-Based Syst.*, vol. 45, pp. 147–165, Jun. 2013, doi: 10.1016/j.knosys.2013.02.014.
- [141] S. Aydin, "Determination of autoregressive model orders for seizure detection," *Comp Sci*, vol. 18, no. 1, 2010, doi: 10.3906/elk-0906-83.
- [142] R. M. Kunst, "kunst2." <https://homepage.univie.ac.at/robert.kunst/> (accessed Mar. 31, 2021).
- [143] P. S. Nagpaul, "Time Series Analysis in WinIDAMS," 2005.
- [144] E. E. Holmes, M. D. Scheuerell, and E. J. Ward, "Applied Time Series Analysis for Fisheries and Environmental Sciences," *NOAA Fisheries, Northwest Fisheries Science Center*, 2020. <https://nwfsc-timeseries.github.io/atsa-labs/sec-dlm-forecasting-with-a-univariate-dlm.html> (accessed Mar. 31, 2021).

- [145] R. Sharma, P. Sircar, and R. B. Pachori, "Computer-aided diagnosis of epilepsy using bispectrum of EEG signals," in *Application of Biomedical Engineering in Neuroscience*, Springer Singapore, 2019, pp. 197–220.
- [146] C. Teixeira *et al.*, "Brainatic: A system for real-time epileptic seizure prediction," in *Biosystems and Biorobotics*, vol. 6, Springer International Publishing, 2014, pp. 7–17.
- [147] N. Dhulekar, S. Nambirajan, B. Oztan, and B. Ü. L. Yener, "Seizure prediction by graph mining, transfer learning, and transformation learning," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2015, vol. 9166, pp. 32–52, doi: 10.1007/978-3-319-21024-7_3.
- [148] I. Halperin, "Fundamentals of Machine Learning in Finance," *Coursera*.
<https://www.coursera.org/learn/fundamentals-machine-learning-in-finance> (accessed Apr. 15, 2021).
- [149] K. Collins-Thompson, "Applied Machine Learning in Python | Coursera," *University of Michigan*.
<https://www.coursera.org/learn/python-machine-learning> (accessed Mar. 25, 2021).
- [150] A. Ben-Hur, "Machine Learning," 2015.
<https://www.cs.colostate.edu/~cs545/fall15/doku.php?id=start> (accessed Apr. 15, 2021).
- [151] A. Y. Ng, "Feature selection, L1 vs. L2 regularization, and rotational invariance," in *Proceedings, Twenty-First International Conference on Machine Learning, ICML 2004*, 2004, pp. 615–622, doi: 10.1145/1015330.1015435.
- [152] EliteDataScience, "Overfitting in Machine Learning: What It Is and How to Prevent It," 2017.
<https://elitedatascience.com/overfitting-in-machine-learning> (accessed Apr. 15, 2021).
- [153] J. D. Rodríguez, A. Pérez, and J. A. Lozano, "Sensitivity Analysis of k-Fold Cross Validation in Prediction Error Estimation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 3, pp. 569–575, 2010, doi: 10.1109/TPAMI.2009.187.
- [154] H. Ide and T. Kurita, "Improvement of learning for CNN with ReLU activation by sparse regularization," in *Proceedings of the International Joint Conference on Neural Networks*, Jun. 2017, vol. 2017-May, pp. 2684–2691, doi: 10.1109/IJCNN.2017.7966185.
- [155] H. Ahn, B. Chung, and C. Yim, "Super-resolution convolutional neural networks using modified and bilateral ReLU," May 2019, doi: 10.23919/ELINFOCOM.2019.8706394.
- [156] K. Collins-Thompson, "Applied Machine Learning in Python | Coursera," *coursera*.
<https://www.coursera.org/learn/python-machine-learning> (accessed Apr. 15, 2021).
- [157] S. Bhattacharya *et al.*, "A novel PCA-firefly based XGBoost classification model for intrusion detection in networks using GPU," *Electron.*, vol. 9, no. 2, Feb. 2020, doi: 10.3390/electronics9020219.
- [158] M. Chen, Q. Liu, S. Chen, Y. Liu, C. H. Zhang, and R. Liu, "XGBoost-Based Algorithm Interpretation and Application on Post-Fault Transient Stability Status Prediction of Power System," *IEEE Access*, vol. 7, pp. 13149–13158, 2019, doi: 10.1109/ACCESS.2019.2893448.
- [159] C. P. Hsieh, Y. T. Chen, W. K. Beh, and A. Y. A. Wu, "Feature Selection Framework for XGBoost Based on Electrodermal Activity in Stress Detection," in *IEEE Workshop on Signal Processing*

- Systems, SiPS: Design and Implementation*, Oct. 2019, vol. 2019-October, pp. 330–335, doi: 10.1109/SiPS47522.2019.9020321.
- [160] P. Mehta *et al.*, “A high-bias, low-variance introduction to Machine Learning for physicists,” *Physics Reports*, vol. 810. Elsevier B.V., pp. 1–124, May 30, 2019, doi: 10.1016/j.physrep.2019.03.001.
 - [161] D. Zhang, L. Qian, B. Mao, C. Huang, B. Huang, and Y. Si, “A Data-Driven Design for Fault Detection of Wind Turbines Using Random Forests and XGboost,” *IEEE Access*, vol. 6, pp. 21020–21031, Mar. 2018, doi: 10.1109/ACCESS.2018.2818678.
 - [162] T. T. Nguyen, J. Z. Huang, and T. T. Nguyen, “Unbiased feature selection in learning random forests for high-dimensional data,” *Sci. World J.*, vol. 2015, 2015, doi: 10.1155/2015/471371.
 - [163] M. A. Elnaggar, M. A. EL Azeem, and F. A. Maghraby, “Machine Learning Model for Predicting Non-performing Agricultural Loans,” in *Advances in Intelligent Systems and Computing*, Apr. 2020, vol. 1153 AISC, pp. 395–404, doi: 10.1007/978-3-030-44289-7_37.
 - [164] R. J. Urbanowicz, M. Meeker, W. La Cava, R. S. Olson, and J. H. Moore, “Relief-based feature selection: Introduction and review,” *Journal of Biomedical Informatics*, vol. 85. Academic Press Inc., pp. 189–203, Sep. 01, 2018, doi: 10.1016/j.jbi.2018.07.014.
 - [165] M. Cherrington, F. Thabtah, J. Lu, and Q. Xu, “Feature selection: Filter methods performance challenges,” May 2019, doi: 10.1109/ICCIsci.2019.8716478.
 - [166] G. Niu, *Data-driven technology for engineering systems health management*. Springer Singapore, 2017.
 - [167] G. Chandrashekar and F. Sahin, “A survey on feature selection methods,” *Comput. Electr. Eng.*, vol. 40, no. 1, pp. 16–28, Jan. 2014, doi: 10.1016/j.compeleceng.2013.11.024.
 - [168] C. Sammut and G. Webb, *Encyclopedia of Machine Learning*. Springer, 2011.
 - [169] “Subspace, Latent Structure and Feature Selection - Statistical and Optimization Perspectives Workshop, SLSFS 2005 Bohinj, Slovenia, February 23-25, 2005, Revised Selected Papers | Craig Saunders | Springer.” <https://www.springer.com/gp/book/9783540341376> (accessed Mar. 31, 2021).
 - [170] I. Guyon, S. Gunn, M. Nikravesh, and L. A. Zadeh, Eds., *Feature Extraction - Foundations and Applications*, Springer-Verlag Berlin Heidelberg, 2006.
 - [171] S. Vora and H. Yang, “A comprehensive study of eleven feature selection algorithms and their impact on text classification,” in *Proceedings of Computing Conference 2017*, Jan. 2018, vol. 2018-January, pp. 440–449, doi: 10.1109/SAI.2017.8252136.
 - [172] B. Mwangi, T. S. Tian, and J. C. Soares, “A review of feature reduction techniques in Neuroimaging,” *Neuroinformatics*, vol. 12, no. 2. Humana Press Inc., pp. 229–244, 2014, doi: 10.1007/s12021-013-9204-3.
 - [173] S. Liang *et al.*, “Relationship between dynamical characteristics of sit-to-walk motion and physical functions of elderly humans,” in *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, Oct. 2016, vol. 2016-October, pp. 871–875, doi: 10.1109/EMBC.2016.7590839.

- [174] Z. Wang and Z. Xie, "Infrared face recognition based on local binary patterns and Kruskal-Wallis test," in *2014 IEEE/ACIS 13th International Conference on Computer and Information Science, ICIS 2014 - Proceedings*, Sep. 2014, pp. 185–188, doi: 10.1109/ICIS.2014.6912131.
- [175] H. Sanz, C. Valim, E. Vegas, J. M. Oller, and F. Reverter, "SVM-RFE: Selection and visualization of the most relevant features through non-linear kernels," *BMC Bioinformatics*, vol. 19, no. 1, p. 432, Nov. 2018, doi: 10.1186/s12859-018-2451-4.
- [176] W. Du, Z. Cao, T. Song, Y. Li, and Y. Liang, "A feature selection method based on multiple kernel learning with expression profiles of different types," *BioData Min.*, vol. 10, no. 1, pp. 1–16, Feb. 2017, doi: 10.1186/s13040-017-0124-x.
- [177] N. Li, M. Shepperd, and Y. Guo, "A systematic review of unsupervised learning techniques for software defect prediction," *Information and Software Technology*, vol. 122, Elsevier B.V., Jun. 01, 2020, doi: 10.1016/j.infsof.2020.106287.
- [178] H. Zhang, J. Zhou, D. Jahed Armaghani, M. M. Tahir, B. T. Pham, and V. Van Huynh, "A Combination of Feature Selection and Random Forest Techniques to Solve a Problem Related to Blast-Induced Ground Vibration," *Appl. Sci.*, vol. 10, no. 3, p. 869, Jan. 2020, doi: 10.3390/app10030869.
- [179] A. Barnwal, H. Cho NVIDIA, and T. Hocking, "Survival regression with accelerated failure time model in XGBoost," Lab report, Northern Arizona University, 2020.
- [180] A. B. Bastiaan Sjardin, Luca Massaron, "Large Scale Machine Learning with Python | Packt," *Packt Publishing*, 2016. <https://www.packtpub.com/product/large-scale-machine-learning-with-python/9781785887215> (accessed Mar. 25, 2021).
- [181] X. Yang and W. Wen, "Ridge and Lasso Regression Models for Cross-Version Defect Prediction," *IEEE Trans. Reliab.*, vol. 67, no. 3, pp. 885–896, Sep. 2018, doi: 10.1109/TR.2018.2847353.
- [182] I. Martin-Diaz, D. Morinigo-Sotelo, O. Duque-Perez, and R. D. J. Romero-Troncoso, "Advances in Classifier Evaluation: Novel Insights for an Electric Data-Driven Motor Diagnosis," *IEEE Access*, vol. 4, pp. 7028–7038, 2016, doi: 10.1109/ACCESS.2016.2622679.
- [183] R. Delgado and X. A. Tibau, "Why Cohen's Kappa should be avoided as performance measure in classification," *PLoS One*, vol. 14, no. 9, p. e0222916, Sep. 2019, doi: 10.1371/journal.pone.0222916.
- [184] D. Chicco and G. Jurman, "The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation," *BMC Genomics*, vol. 21, no. 1, p. 6, Jan. 2020, doi: 10.1186/s12864-019-6413-7.
- [185] M. A. U. H. Tahir, S. Asghar, A. Manzoor, and M. A. Noor, "A Classification Model for Class Imbalance Dataset Using Genetic Programming," *IEEE Access*, vol. 7, pp. 71013–71037, 2019, doi: 10.1109/ACCESS.2019.2915611.
- [186] S. Boughorbel, F. Jarray, and M. El-Anbari, "Optimal classifier for imbalanced data using Matthews Correlation Coefficient metric," *PLoS One*, vol. 12, no. 6, p. e0177678, Jun. 2017, doi: 10.1371/journal.pone.0177678.
- [187] S. X. McLachlan, G. J., Rathnayake, S., and Lee, "Comprehensive Chemometrics-Chemical and Biochemical Data Analysis," 2nd ed. Elsevier, 2020.

- [188] Syed Naeem Ahmed , “*Physics and Engineering of Radiation Detection*,” 2nd ed. Elsevier, 2015.
- [189] V. A. Profillidis and G. N. Botzoris, “Statistical Methods for Transport Demand Modeling,” in *Modeling of Transport Demand*, Elsevier, 2019, pp. 163–224.
- [190] M. Radovic, M. Ghalwash, N. Filipovic, and Z. Obradovic, “Minimum redundancy maximum relevance feature selection approach for temporal gene expression data,” *BMC Bioinformatics*, vol. 18, no. 1, p. 9, Jan. 2017, doi: 10.1186/s12859-016-1423-9.
- [191] P. Bugata and P. Drotar, “On some aspects of minimum redundancy maximum relevance feature selection,” *Sci. China Inf. Sci.*, vol. 63, no. 1, pp. 1–15, Jan. 2020, doi: 10.1007/s11432-019-2633-y.
- [192] M. Billah and S. Waheed, “Minimum redundancy maximum relevance (mRMR) based feature selection from endoscopic images for automatic gastrointestinal polyp detection,” *Multimed. Tools Appl.*, vol. 79, no. 33–34, pp. 23633–23643, Sep. 2020, doi: 10.1007/s11042-020-09151-7.
- [193] A. Mulye, “Power Spectrum Density Estimation Methods for Michelson Interferometer Wavemeters,” Master thesis, University of Ottawa, 2016.
- [194] A. S. Al-Fahoum and A. A. Al-Fraihat, “Methods of EEG Signal Features Extraction Using Linear Analysis in Frequency and Time-Frequency Domains,” *ISRN Neurosci.*, vol. 2014, pp. 1–7, Feb. 2014, doi: 10.1155/2014/730218.
- [195] J. M. Girault, F. Ossant, A. Ouahabi, D. Kouamé, and F. Patat, “Time-varying autoregressive spectral estimation for ultrasound attenuation in tissue characterization,” *IEEE Trans. Ultrason. Ferroelectr. Freq. Control*, vol. 45, no. 3, pp. 650–659, 1998, doi: 10.1109/58.677609.
- [196] X. Sun, ao Shen, and X. Mu, “New Phase Difference Measurement Method for Non-Integer Number of Signal Periods Based on Multiple Cross-Correlations,” 2015, 7, pp. 1537-1541.
- [197] G. Buzsáki and F. L. da Silva, “High frequency oscillations in the intact brain,” *Progress in Neurobiology*, vol. 98, no. 3. Prog Neurobiol, pp. 241–249, Sep. 2012, doi: 10.1016/j.pneurobio.2012.02.004.
- [198] B. Frauscher *et al.*, “High-Frequency Oscillations in the Normal Human Brain,” *Ann. Neurol.*, vol. 84, no. 3, pp. 374–385, Sep. 2018, doi: 10.1002/ana.25304.
- [199] A. Thomschewski, A. S. Hincapié, and B. Frauscher, “Localization of the epileptogenic zone using high frequency oscillations,” *Frontiers in Neurology*, vol. 10, no. FEB. Frontiers Media S.A., p. 94, 2019, doi: 10.3389/fneur.2019.00094.
- [200] E. H. Smith *et al.*, “Dual mechanisms of ictal high frequency oscillations in rhythmic onset seizures,” *Scientific report* 10, Jan. 13, 2020, doi: 10.1101/2020.01.09.20017053.
- [201] T. Uchida, K. Fujiwara, T. Inoue, Y. Maruta, M. Kano, and M. Suzuki, “Analysis of VNS Effect on EEG Connectivity with Granger Causality and Graph Theory,” in *2018 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference, APSIPA ASC 2018 - Proceedings*, Mar. 2019, pp. 861–864, doi: 10.23919/APSIPA.2018.8659723.
- [202] X. Wu *et al.*, “Altered intrinsic brain activity associated with outcome in frontal lobe epilepsy,” *Sci. Rep.*, vol. 9, no. 1, pp. 1–8, Dec. 2019, doi: 10.1038/s41598-019-45413-7.
- [203] H. Shokouh Alaei, M. A. Khalilzadeh, and A. Gorji, “Optimal selection of SOP and SPH using fuzzy inference system for on-line epileptic seizure prediction based on EEG phase synchronization,”

- Australas. Phys. Eng. Sci. Med.*, vol. 42, no. 4, pp. 1049–1068, Dec. 2019, doi: 10.1007/s13246-019-00806-w.
- [204] N. D. Truong, L. Kuhlmann, M. R. Bonyadi, D. Querlioz, L. Zhou, and O. Kavehei, “Epileptic Seizure Forecasting with Generative Adversarial Networks,” *IEEE Access*, vol. 7, pp. 143999–144009, 2019, doi: 10.1109/ACCESS.2019.2944691.
 - [205] A. Koop, “Machine Learning Algorithms: Supervised Learning Tip to Tail,” *Coursera*. <https://www.coursera.org/learn/machine-learning-classification-algorithms> (accessed Apr. 15, 2021).
 - [206] A. Koop, “Machine Learning: Algorithms in the Real World,” *Coursera*. <https://www.coursera.org/specializations/machine-learning-algorithms-real-world> (accessed Apr. 15, 2021).
 - [207] M. N. and I. Altintas, “Machine Learning With Big Data,” *Coursera*. <https://www.coursera.org/learn/big-data-machine-learning#instructors> (accessed Apr. 15, 2021).
 - [208] P. Salami *et al.*, “Seizure onset location shapes dynamics of initiation,” *Clin. Neurophysiol.*, vol. 131, no. 8, pp. 1782–1797, Aug. 2020, doi: 10.1016/j.clinph.2020.04.168.
 - [209] A. Ben Amar, A. B. Kouki, and H. Cao, “Power approaches for implantable medical devices,” *Sensors (Switzerland)*, vol. 15, no. 11. MDPI AG, pp. 28889–28914, Nov. 13, 2015, doi: 10.3390/s151128889.
 - [210] A. Goyal and N. Sardana, “Empirical Analysis of Ensemble Machine Learning Techniques for Bug Triaging,” Aug. 2019, doi: 10.1109/IC3.2019.8844876.
 - [211] F. Huang, G. Xie, and R. Xiao, “Research on ensemble learning,” in *2009 International Conference on Artificial Intelligence and Computational Intelligence, AICI 2009*, 2009, vol. 3, pp. 249–252, doi: 10.1109/AICI.2009.235.
 - [212] A. B. B. Ashok Bhowmick, Tamer Abdou, “Predictive analytics in healthcare epileptic seizure recognition,” in *Proceedings of the 28th Annual International Conference on Computer Science and Software Engineering*, pp. 323–330, Accessed: Mar. 26, 2021. [Online]. Available: <https://dl.acm.org/doi/10.5555/3291291.3291327>.
 - [213] P. Vanabelle, P. De Handschutter, R. El Tahry, M. Benjelloun, and M. Boukhebouze, “Epileptic seizure detection using EEG signals and extreme gradient boosting,” *J. Biomed. Res.*, vol. 34, no. 3, pp. 228–239, 2020, doi: 10.7555/JBR.33.20190016.
 - [214] T. Saito and M. Rehmsmeier, “The Precision-Recall Plot Is More Informative than the ROC Plot When Evaluating Binary Classifiers on Imbalanced Datasets,” *PLoS One*, vol. 10, no. 3, p. e0118432, Mar. 2015, doi: 10.1371/journal.pone.0118432.
 - [215] Q. Zou, S. Xie, Z. Lin, M. Wu, and Y. Ju, “Finding the Best Classification Threshold in Imbalanced Classification,” *Big Data Res.*, vol. 5, pp. 2–8, Sep. 2016, doi: 10.1016/j.bdr.2015.12.001.
 - [216] G. Rivera, R. Florencia, V. García, A. Ruiz, and J. P. Sánchez-Solís, “News classification for identifying traffic incident points in a Spanish-speaking country: A real-world case study of class imbalance learning,” *Appl. Sci.*, vol. 10, no. 18, Sep. 2020, doi: 10.3390/AP10186253.
 - [217] M. Tonini, M. D’Andrea, G. Biondi, S. Degli Esposti, A. Trucchia, and P. Fiorucci, “A Machine

- Learning-Based Approach for Wildfire Susceptibility Mapping. The Case Study of the Liguria Region in Italy,” *Geosciences*, vol. 10, no. 3, p. 105, Mar. 2020, doi: 10.3390/geosciences10030105.
- [218] M. Luckner, B. Topolski, and M. Mazurek, “Application of XGboost algorithm in fingerprinting localisation task,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2017, vol. 10244 LNCS, pp. 661–671, doi: 10.1007/978-3-319-59105-6_57.
 - [219] “1. Supervised learning — scikit-learn 0.24.1 documentation.” https://scikit-learn.org/stable/supervised_learning.html#supervised-learning (accessed Mar. 25, 2021).
 - [220] R. S. Smith and T. Windeatt, “Facial action unit recognition using multi-class classification,” *Neurocomputing*, vol. 150, no. PB, pp. 440–448, Feb. 2015, doi: 10.1016/j.neucom.2014.07.066.
 - [221] B. Böken, “On the appropriateness of Platt scaling in classifier calibration,” *Inf. Syst.*, vol. 95, p. 101641, Jan. 2021, doi: 10.1016/j.is.2020.101641.
 - [222] T. Haslwanter, “An Introduction to Statistics with Python - With Applications in the Life Sciences | Thomas Haslwanter | Springer,” *Springer*, 2016. <https://www.springer.com/gp/book/9783319283159> (accessed Mar. 25, 2021).
 - [223] J. Rasekhi, M. R. K. Mollaei, M. Bandarabadi, C. A. Teixeira, and A. Dourado, “Preprocessing effects of 22 linear univariate features on the performance of seizure prediction methods,” *J. Neurosci. Methods*, vol. 217, no. 1–2, pp. 9–16, Jul. 2013, doi: 10.1016/j.jneumeth.2013.03.019.
 - [224] H. Wang and H. Zheng, “True Positive Rate,” in *Encyclopedia of Systems Biology*, Springer New York, 2013, pp. 2302–2303.
 - [225] Duun-Henriksen and Jonas, “Detection and Prediction of Epileptic Seizures,” PhD thesis, Technical university of Denmark, 2013.
 - [226] T. T. K. Munia and S. Aviyente, “Time-Frequency Based Phase-Amplitude Coupling Measure For Neuronal Oscillations,” *Sci. Rep.*, vol. 9, no. 1, pp. 1–15, Dec. 2019, doi: 10.1038/s41598-019-48870-2.
 - [227] M. H. Myers, A. Padmanabha, G. Hossain, A. L. De Jongh Curry, and C. D. Blaha, “Seizure prediction and detection via phase and amplitude lock values,” *Front. Hum. Neurosci.*, vol. 10, no. MAR2016, Mar. 2016, doi: 10.3389/fnhum.2016.00080.
 - [228] Y. Salimpour and W. S. Anderson, “Cross-frequency coupling based neuromodulation for treating neurological disorders,” *Front. Neurosci.*, vol. 13, no. FEB, 2019, doi: 10.3389/fnins.2019.00125.
 - [229] A. Cimenser *et al.*, “Tracking brain states under general anesthesia by using global coherence analysis,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 108, no. 21, pp. 8832–8837, May 2011, doi: 10.1073/pnas.1017041108.
 - [230] N. Zangiabadi, L. Di. Ladino, F. Sina, J. P. Orozco-Hernández, A. Carter, and J. F. Téllez-Zenteno, “Deep brain stimulation and drug-resistant epilepsy: A review of the literature,” *Frontiers in Neurology*, vol. 10, no. JUN. Frontiers Media S.A., p. 601, 2019, doi: 10.3389/fneur.2019.00601.
 - [231] A. Yousefi, R. S. Fard, U. T. Eden, and E. N. Brown, “State-Space Global Coherence to Estimate the Spatio-Temporal Dynamics of the Coordinated Brain Activity,” in *Proceedings of the Annual*

- International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, Jul. 2019, vol. 2019, pp. 5794–5798, doi: 10.1109/EMBC.2019.8856634.
- [232] G. V. Kumar, T. Halder, A. K. Jaiswal, A. Mukherjee, D. Roy, and A. Banerjee, “Large Scale Functional Brain Networks Underlying Temporal Integration of Audio-Visual Speech Perception: An EEG Study,” *Front. Psychol.*, vol. 7, no. OCT, p. 1558, Oct. 2016, doi: 10.3389/fpsyg.2016.01558.
 - [233] L. Cornelissen, S. E. Kim, P. L. Purdon, E. N. Brown, and C. B. Berde, “Age-dependent electroencephalogram (Eeg) patterns during sevoflurane general anesthesia in infants,” *Elife*, vol. 4, no. JUNE 2015, pp. 1–25, Jun. 2015, doi: 10.7554/eLife.06513.
 - [234] F. Sauter-Starace *et al.*, “Long-Term Sheep Implantation of WIMAGINE®, a Wireless 64-Channel Electrocorticogram Recorder,” *Front. Neurosci.*, vol. 13, p. 847, Aug. 2019, doi: 10.3389/fnins.2019.00847.
 - [235] K. Matsushita *et al.*, “A Fully Implantable Wireless ECoG 128-Channel Recording Device for Human Brain–Machine Interfaces: W-HERBS,” *Front. Neurosci.*, vol. 12, p. 511, Jul. 2018, doi: 10.3389/fnins.2018.00511.
 - [236] S. Il Chang, S. Y. Park, and E. Yoon, “Minimally-invasive neural interface for distributed wireless electrocorticogram recording systems,” *Sensors (Switzerland)*, vol. 18, no. 1, Jan. 2018, doi: 10.3390/s18010263.
 - [237] B. Csernyus *et al.*, “Recent antiepileptic and neuroprotective applications of brain cooling,” *Seizure*, vol. 82. W.B. Saunders Ltd, pp. 80–90, Nov. 01, 2020, doi: 10.1016/j.seizure.2020.09.018.