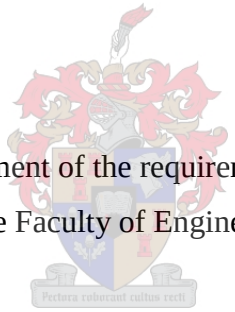
Application of Machine Learning with Electroencephalography in Seizure Detection

André Daniël Volschenk



Thesis presented in partial fulfilment of the requirements for the degree of Master of Engineering (Mechatronic) in the Faculty of Engineering at Stellenbosch University

Supervisor: J. van der Merwe

Co-supervisor: Prof. P.R. Fourie

December 2017

Declaration

By submitting this thesis/dissertation electronically, I declare that the entirety of the work contained therein is my own, original work, that I am the sole author thereof (save to the extent explicitly otherwise stated), that reproduction and publication thereof by Stellenbosch University will not infringe any third party rights and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

Signature .....
A.D. Volschenk

Date December 2017
.....

Abstract

Application of Machine Learning with Electroencephalography in Seizure Detection

Volschenk, A.D.

Department of Mechanical and Mechatronic Engineering,
Stellenbosch University,
Private Bag X1, 7602 Matieland, RSA.

Thesis: M.Eng (Mechatronic)
December
2017

INTRODUCTION: Seizures are periods of abnormal electrical activity in the brain, which induce brain injury to the sufferer. A patient that suffer seizures may need to be monitored for several hours, days, or even weeks. Seizure identification using electroencephalography (EEG) can be achieved through the use of seizure detection algorithms. Continuous EEG monitoring with early-detection algorithms to warn of the onset of seizures has many benefits as it allows for early intervention. In this study, the desired seizure monitoring software is designed for immediate application in the clinical environment to any patient. The aim of this research is to develop a robust, completely automatic software solution intended for real-time whole-brain seizure detection that uses EEG data, and no patient- or seizure-specific tuning. The training and testing is performed using a large, publicly available data corpus. The current state-of-the-art algorithm is improved upon. Detection should be possible as soon as a patient is rushed into the intensive care unit (ICU) and the EEG electrodes are connected properly.

METHODS: The CHB-MIT data corpus is used. Included for analysis are 24 patients, 185 seizures, 979.9 hours of data, and 18 channels. Independent training and testing sets are used, with a train:test ratio of 80:20. Preprocessing: If a frame is corrupted by abnormal channel amplitude, mains noise, or phase reversal, then it is rejected without being passed to the next processes. Otherwise, the frame is bandpass filtered between 0.5 and 70 Hz, and a 5-level db2 wavelet filterbank is used for sub-band coding. Frequency bands γ(high), γ(low), β, α, θ, and δ are thereby approximated. The Relative Average Amplitude (RAA), Relative Scale Energy (RSE), and Coefficient of Variation of Amplitude (CVA) features of bands β, α, and θ are taken. Classification: A probabilistic Bayes classifier is trained and used for classification. Ictal/inter-ictal and high-/low-α classifiers are used. A novel automatic procedure for α training-data selection is implemented. Postprocessing: A sequential hypothesis test and persistence

is used for false positive reduction. The objective function in the train-validate phase is the F_1 score, which is the harmonic mean of Positive Predictive Value (PPV) and True Positive Rate (TPR). Leave-one-out-cross-validation (LOOCV) is used in the train-validate phase. The TPR, PPV, and False Positive Rate (FPR) are reported for convenience.

RESULTS: The offline train-validate phase yielded $TPR = 58.73$ %, $PPV = 59.89$ %, $FPR = 0.2045$ /h. The online test phase yielded $TPR = 58.5$ %, $PPV = 40.61$ %, $FPR = 0.3536$ /h.

CONCLUSIONS: The algorithm presented here is an improvement to the current state-of-the-art. For clinical applicability, the issues of overall algorithm performance and inter-patient variability should be further improved.

Uittreksel

Toepassing van Masjienleer met Elektroënsefalografie in Stuipe Deteksie

Volschenk, A.D.

Departement Meganiese en Megatroniese Inginieurswese,
Universiteit Stellenbosch,
Privaatsak X1, 7602 Matieland, RSA.

Tesis: M.Ing (Megatronies)
Desember
2017

INLEIDING: Stuipe is periodes van abnormale elektriese aktiwiteit in die brein, wat breinbesering aan die lyer veroorsaak. 'n Pasiënt wat aan stuipe ly moet vir 'n paar uur, dae, of selfs weke gemonitor word. Die identifikasie van stuipe word gedoen met behulp van elektroënsefalografie (EEG) deur die gebruik van stuipe-opsporingsalgoritmes. Die gebruik van deurlopende EEG monitering met vroeë opsporingsalgoritmes waarsku teen die aanvang van stuipe en het baie voordele aangesien dit voorsiening maak vir vroeë ingryping. In hierdie studie is die gewenste stuipe-monitering sagteware ontwerp vir onmiddellike toepassing op enige pasiënt in die kliniese omgewing. Die doel van hierdie navorsing is om 'n robuuste, heeltemal outomatiese sagteware-oplossing te ontwikkel wat gebruik kan word vir intydse hele-brein stuipe opsporing wat EEG data gebruik, en geen pasiënt- of stuip-spesifieke verfyning benodig nie. Die opleiding en toetsing is uitgevoer deur gebruik te maak van 'n groot, openlik-beskikbare data corpus. Daar word verbeteringe aangebring op die huidige beste-van-die-beste algoritme. Opsporing moet moontlik wees sodra 'n pasiënt in die intensiewe sorgeenheid ingebring word en die EEG-elektrodes behoorlik aangeheg is.

METODES: Die KHB-MIT data corpus word gebruik. Vir analise is 24 pasiënte, 185 stuipe, 979.9 ure se data, en 18 kanale ingesluit. Onafhanklike opleiding- en toetsstelle word gebruik, met 'n oplei:toets verhouding van 80:20. Voorverwerking: Indien 'n raam besmet is deur abnormale kanaalamplitude, kraglyn-geraas of fase-omkering, dan word dit afgekeur sonder om aan die volgende prosesse oorgedra te word. Andersins word die raam deur 'n banddeurlaat filter tussen 0.5 en 70 Hz gefiltreer, en 'n 5-vlak db2 golfie filterbank word gebruik vir subband kodering. Frekwensiebande γ(hoog), γ(laag), β, α, θ, en δ word sodoende beraam. Die relatiewe gemiddelde amplitude (RGA), relatiewe skaalenergie (RSE) en koeffisiënt van variasie van amplitude (KVA) eienskappe van bande β, α, en θ word geneem. Klassifikasie: 'n Waarskynlikheids

Bayes-klassifiseerder word opgelei en gebruik vir klassifikasie. Ictale/inter-ictale en hoë/lae α klassifiseerders word gebruik. 'n Nuwe outomatiese prosedure vir α opleiding-data seleksie word geïmplementeer. Na-verwerking: 'n Opeenvolgende hipotese toets en blywendheid word gebruik vir vals-positiewe vermindering. Die teiken funksie in die opleidingsvalidereringsfase is die F_1 telling, wat die harmoniese gemiddeld van Positiewe Voorspellende Waarde (PVW) en Ware Positiewe Koers (WPK) is. Laat-een-uit-kruis-validering (LEUKV) word gebruik in die opleidingsvalidereringsfase. Die WPK, PVW en vals-positiewe koers (VPK) word gemeld vir gerief.

RESULTATE: Die aflyn opleidingsvalidereringsfase het $WPK = 58.73$ %, $PVW = 59.89$ %, $VPK = 0.2045$ /h opgelewer. Die aanlyn toetsfase het $WPK = 58.5$ %, $PVW = 40.61$ %, $VPK = 0.3536$ /h opgelewer.

GEVOLGTREKKINGS: Die algoritme wat hier aangebied word is 'n verbetering van die huidige beste-van-die-beste. Vir kliniese toepaslikheid moet die kwessies van algehele algoritme prestasie en interpasiënt veranderlikheid verder verbeter word.

Toewyding (Dedication)

Opgedra aan Dr. Daniël Johannes Volschenk, Almine Karin Volschenk, en Hendrik Smith Volschenk
wat my belangstelling in die ingenieurswese en mediese gebiede aanmoedig, en wat die toewyding, volharding, en werksetiek in my bevorder wat nodig is om te presteer.

Acknowledgements

I would like to express my sincere gratitude to a few individuals who have assisted me with the technical development and finalization of this work.
First and foremost, Johan van der Merwe, my supervisor in this study. He has provided excellent guidance throughout this work to keep me focussed and on track at all times. Johan has furthermore been invaluable in the editing process. I truly owe a great deal of gratitude to him.
To Pieter Fourie, co-supervisor in this study, I would like to express sincere appreciation. My meetings with Pieter have given me so much inspiration and enthusiasm. I have thoroughly enjoyed discussing and deliberating this research with him.
To Quinton Hendrikse of the computer laboratories, and Charl Möller of the HPC (high performance computer) cluster. I thank each of them for their technical assistance with setting up the computation facilities. Their patience and helpfulness have taught me much and enabled my rather extensive use of the computation facilities.
Finally, to my friend and colleague, Brandon Wakefield. He has introduced me to some fundamental Machine Learning concepts, propelling my study into the field. I sincerely enjoy our comprehensive discussions concerning Machine Learning, and computer programming, as well as a host of other technical fields.

More personally, I would like to extend sincere gratitude to Roedolf David Hendrik Vorster and Mariona Prat Plana, along with all my other friends and family who are not listed here. The times we share together have kept me human, I think. And finally to the Biomedical Engineering Research Group (BERG), with whom I feel kinship.

Contents

List of Tables

List of Figures

Nomenclature

Abbreviations

		Defined on page:
A5	Approx. db2 coefficient with frequency 0.5 → 4 Hz in this study	54
AED	Anti-Epileptic Drug	1
CHB-MIT	Children's Hospital of Boston - Massachusetts Institute of Technology	27
CVA	Coefficient of Variation of Amplitude	56
D1	Detail db2 coefficient with frequency 64 → 70 Hz in this study	54
D2	Detail db2 coefficient with frequency 32 → 64 Hz in this study	54
D3	Detail db2 coefficient with frequency 16 → 32 Hz in this study	54
D4	Detail db2 coefficient with frequency 8 → 16 Hz in this study	54
D5	Detail db2 coefficient with frequency 4 → 8 Hz in this study	54
db2	Daubechies wavelet with 2 vanishing-moments	54
ECoG	Electrocorticography	13
EEG	Electroencephalography	12
EMG	Electromyography	14
FFT	Fast Fourier Transform	35
FSPP	Freiburg Seizure Prediction Project	27
ICU	Intensive Care Unit	1
iEEG	Intra-cranial Electroencephalography	13
IEEG	Intra-cranial Electroencephalography	13
LOOCV	Leave-One-Out Cross-Validation	70
NBC	Naïve Bayes Classifier	44
RAA	Relative Average Amplitude	55
RSE	Relative Scale Energy	56

List of symbols

		Defined on page:
A_{TH}	Alpha detection threshold	65
amp_{EMG}	EMG amplitude ratio	61
F_1	The F_1 score (also called F-score or F-measure)	25
FN	False Negative	21
FP	False Positive	21
FPR	False Positive Rate (also called Fall-out or Probability of false alarm)	23
I_{TH}	Ictal detection threshold	65
N	Number of frames used for probability estimation	66
$\hat{p}$	Probability estimate for an ictal event within the last N frames	66
p	Detection threshold probability	66
PPV	Positive Predictive Value (also called *Selectivity* or Precision)	23
t_{dl}	Detection latency (also called detection delay)	23
t_{ph}	Prediction horizon (also called prediction time)	23
T	Persistence refractory parameter	67
TN	True Negative	21
TP	True Positive	21
TPR	True Positive Rate (also called *Sensitivity* or Recall or Probability of detection)	23
x_{ALP}	Percentile of α data from the top used for training	61
x_{chn}	Number of highest value channels used	63
x_{EMG}	Percentile of EMG data below which training data is taken	61
x_{es}	Temporal context in probability calculation	65
x_{high}	The maximum allowable channel-amplitude threshold	53
x_{mains}	The maximum allowable mains-noise-amplitude threshold	53
x_{NAL}	Percentile of α data below which training data is taken	61
x_{NTH}	Scaling factor on amp_{EMG}	65
x_{phase}	The phase-reversal threshold scaling parameter	54
α	EEG frequency band taken as 8 → 16 Hz in this study	15
β	EEG frequency band taken as 16 → 32 Hz in this study	15
γ	EEG frequency band taken as 32 → ∞ Hz in this study	15
δ	EEG frequency band taken as 0 → 4 Hz in this study	15
θ	EEG frequency band taken as 4 → 8 Hz in this study	15
$\Pi_{h(m)}$	Set of α-model hyper-parameters	69
$\Pi_{h(r)}$	Set of artefact rejection hyper-parameters	69
Π_p	Set of parameters (excl. hyper-parameters)	69

Chapter 1

Introduction

1.1 Background

Seizures are periods of abnormal electrical activity in the brain, which may or may not manifest clinically. A seizure will take place when a burst of electrical impulses in the brain exceed their normal limits. These impulses will spread to adjacent areas and cause an uncontrolled storm of electrical activity in the brain. If the electrical impulses are conducted to muscle fibres, twitches or convulsions may result. If this happens, the seizure is said to be a clinical (also called convulsive) seizure.

There is an extensively long list of possible causes for seizures. Just some of the pathological causes for seizures include diet, medical conditions (including brain tumours, brain abscesses, epilepsy, encephalitis, meningitis, among many others), some medications and drug or alcohol abuse, fevers (especially in young children), head injury, and hypoglycaemia. Seizures induce brain injury to the sufferer (Bergen, 2006; Bronen, 2000), and so a patient diagnosed with a serious case of any of the possible causes for seizures may need to be monitored for several hours, days, or even weeks.

The first line of treatment for seizures is anticonvulsant medication, also called anti-epileptic drugs (AEDs) or anti-seizure drugs. AEDs are a successful form of treatment for about 70 % of patients (Sander, 2004). Besides AEDs, electrical stimulation, and therapeutic hypothermia are treatments used for harm reduction. Other interventions may be more appropriate for non-epileptic seizures. Electroencephalography (EEG) is commonly used for diagnosis and accurate quantification of convulsive- as well as non-convulsive seizures (NCS) (also called sub-clinical seizures). There is an increasing amount of evidence that suggest that NCS occur in a significant portion of obtunded or unresponsive patients in Intensive Care Unit (ICU) settings. Retrospective analysis suggest that 11 % to 55 % of patients in neurologic ICUs may be experiencing NCS (Scheuer, 2002). NCS may prolong the need for intensive care and worsen the degree of brain injury. A diagnosis of NCS often results in intensification of AED therapy, however NCS is often identified late in the course of an illness, and as such the clinical impact of the treatment may be suboptimal (Scheuer, 2002).

EEG can be used to screen for clinical- as well as sub-clinical seizures and to

provide additional prognostic information related to neurologic outcome. Only continuous EEG monitoring can however reliably provide timely and therapeutically important information to guide AED therapy, since seizures may occur outside the routine EEG recording session, and because many seizures are NCS which makes it difficult to know when to use EEG (Abend *et al.*, 2011; Scheuer, 2002). Modern ICUs have equipment to monitor almost all vital functions of a patient, except for the brain, despite recommendations in literature (Ponten *et al.*, 2010; Shellhaas and Clancy, 2007).

Seizure identification can be very difficult. Intensivists and ICU staff are not adequately trained for interpreting EEG data (overall mean of 61 % for recognition of epileptiform discharges) (Rijsdijk *et al.*, 2008). Another issue is that long term continuous EEG would be incredibly labour intensive. Perhaps the greatest challenge to using continuous EEG in clinical practise is the lack of reliable method for online seizure detection to determine when ICU staff evaluation of the patient is required (Ponten *et al.*, 2010).

The problem of seizure identification using continuous EEG can be alleviated through the use of seizure detection algorithms. Continuous EEG monitoring with early detection and prediction algorithms to warn of onset of seizures has many benefits as it allows for early intervention (van Putten and Tavy, 2004), such as timely administering of fast-acting AEDs, electrical stimulation, or therapeutic hypothermia (Fisher *et al.*, 2010; Hill *et al.*, 2000; Morrell, 2006; Mormann *et al.*, 2007; Stein *et al.*, 2000). One offline post-monitoring benefit of automated seizure detection is the potentially massive reduction in the amount of EEG data that needs to be stored and reviewed (Fisch, 1999).

Two scenarios for how a seizure can occur are proposed: First is that the seizure is caused by a sudden and abrupt state transition, in which case it is not preceded by detectable dynamic change. In the second case, the transition is gradual. In the abrupt first case, seizure detection algorithms could be used to detect a seizure, but prediction algorithms would fail to predict the onset of seizures. In the second, gradual case, detection and prediction algorithms could potentially be successful (Niedermeyer and Lopes da Silva, 2012). This implies that an ideal monitor may need both a detection, and a prediction algorithm concurrently.

1.2 Problem identification

The ideal monitoring algorithm should be designed for application in the clinical environment. For the purposes of this study, the aspects in which publications in literature do not meet the requirements of a monitoring algorithm designed for immediate clinical implementation shall be referred to as ‘limitations’. It is

acknowledged that not all publications in literature have the aim of developing such a monitor, and as such these limitations should be regarded as impediments to immediate clinical use of the given algorithm, rather than a limitation of the research publication itself.

Reports of the performance of monitoring techniques suffer from one or more of the following limitations:

1. Seizure prediction accuracy is highly variable between patients.
2. Researchers report different performance metrics.
3. The technique requires careful patient- or seizure-specific tuning.
4. The technique is not evaluated on long-term continuous EEG and/or the technique is not evaluated on independent data.
5. The software is tested using intracranial EEG (iEEG), or with some other idealized data that do not accurately simulate clinical conditions.
6. Training and testing are done on a private data corpus and/or small data corpora are used for performance evaluation.
7. Only 1 channel (usually the focus channel) or 1 small cluster or channels and/or only 1 seizure-type is used to train and evaluate the system.
8. Data preprocessing includes manual removal of data intervals of ocular- or muscle artefacts, or some other data pre-selection.

Limitations 1 and 2 are further described in Section 2.3.2, and limitations 3 through 8 are further described in both Sections 2.3.3 and 2.3.5.

1.3 Research motivation

Ideally a continuous EEG monitoring algorithm should address all limitations listed in Section 1.2. The ideal monitor should be able to function immediately and independently as soon as the EEG device is placed correctly on the patient. Monitoring can then be started immediately. EEG hardware is available commercially in various grades and prices from a number of international suppliers. In fact it is the software component of the proposed device that requires most of the research and development.

In order to fulfil the need identified, software needs to be developed in order to detect seizures online using EEG as input data. Online testing in this context implies feeding data to the software consecutively from start to end, without prior manipulation, to simulate the clinical environment conditions. In this

research, a detection algorithm is trained and optimized, and then tested online. The development of a prediction algorithm is not within the scope of this thesis.

The Saab and Gotman (2005) method addresses the limitations listed in Section 1.2, with the exception that a private data corpus is used for training and testing. The method in this study is based on the Saab and Gotman (2005) method, however it will be applied to a publically available data corpus. In this way the results are reproducible for future researchers. The method presented in this study will not replicate the Saab and Gotman (2005) method exactly. The Saab and Gotman (2005) method makes use of many parameters which it does not attempt to optimize. Optimizing many inter-dependent parameters simultaneously is time- and computationally expensive. Previously optimized as well as previously unoptimized parameters are all optimized in this study. Furthermore, additional procedures are introduced in this work in an attempt to improve on the Saab and Gotman (2005) method.

1.4 Aims and objectives

Study aims

The aim of this research is to develop a robust, completely automatic software solution intended for real-time whole-brain seizure detection that uses EEG data, and no patient- or seizure-specific tuning. The training and testing is to be performed using a large, publicly available data corpus. The current state-of-the-art seizure detector is improved upon. The final deliverable of this research is the online performance data of the algorithm and a discussion that addresses each limitation in Section 1.2.

Study objectives

1. Develop software for EEG seizure detection based on the current state-of-the-art.
2. Introduce improvements to the method, by reducing FPR while attempting to maintain TPR.
3. Train and optimize the technique offline.
4. Test the technique online with independent data and report the online performance results.

1.5 Scope

In this study, the Saab and Gotman (2005) method is applied to the publicly available CHB-MIT corpus. Previously unoptimized parameters (x_{phase}, x_{NTH}, x_{es}) are optimized. Leave-one-out-cross-validation (LOOCV) is used in the train-validation phase iterations, and training data from each patient is given equal weight. A sequential hypothesis test is introduced. The α-data selection heuristic is completely automated. LOOCV, weighing patient data, and the sequential hypothesis test are additional procedures to improve on the method. The online performance is given and discussed with reference to the limitations given in Section 1.2. The study was started on 03 June 2016 and the deadline for final submission is 08 September 2017.

1.6 Contributions

- The current state-of-the-art method is applied to a publicly available data corpus, in order to make it reproducible.
- The method is improved upon by optimizing more of its parameters, introducing a sequential hypothesis test, weighing patient data equally in the training phase, and applying LOOCV to the train-validate phase iterations.
- The tedious manual procedure for α-data selection is automated. Automated procedures save time and bode well for possible future dynamic learning implementations.

Chapter 2

Literature study

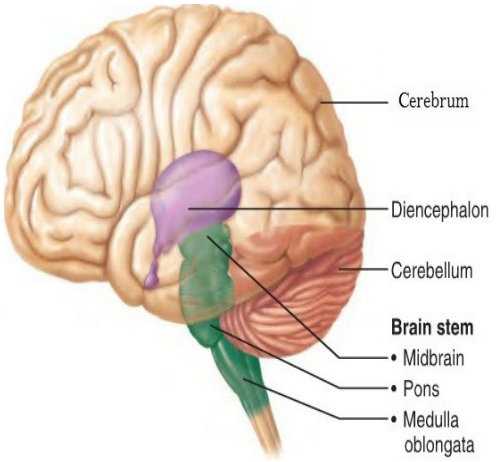
2.1 Physiology and pathology

The human brain is a remarkable organ of the most incredible complexity. Such an organ is understandably expensive to maintain. Despite the fact that the brain is only about 2 % of the weight of the body, it uses approximately 20 % of both its total energy, as well as approximately 20 % of its oxygen intake (Raichle and Gusnard, 2002). The brain, like any other organ, is composed of cells. The cells that enable the brain to perform its function as a biological computer are called neurons. The latest estimates for the number of neurons in the brain is set at 86 billion, with an average of 40 000 synapses on each neuron, and about 10 times as many neuroglia as neurons to maintain homoeostasis.

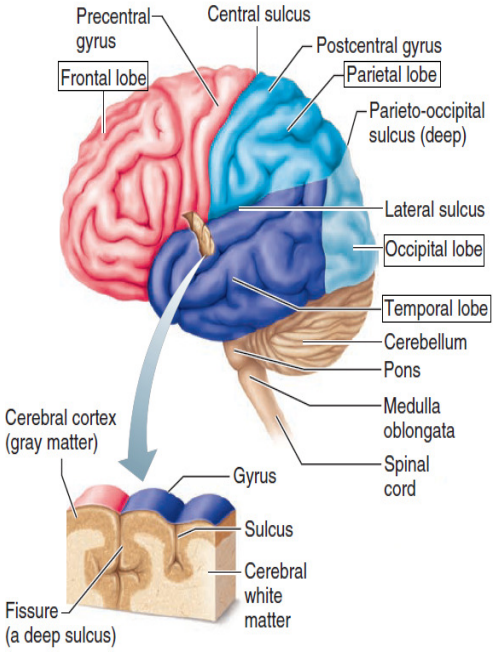
2.1.1 The human brain

The human brain is the major functional unit of the central nervous system (CNS). The brain is composed of four major regions as shown in Figure 2.1a, namely the cerebrum (also called cerebral cortex), diencephalon, cerebellum, and the brain stem. The cerebrum is the largest part of the human brain and is divided into four lobes which express its location, as shown in Figure 2.1b. The locations are: the frontal lobe, parietal lobe, occipital lobe, and temporal lobe.

The frontal lobe contains the majority of dopamine-sensitive neurons. It is associated with many functions, including: reward, attention, short-term memory, planning, decision-making and problem-solving, motivation, behaviour, consciousness, and emotion. The parietal lobe integrates sensory information and various modalities, including spatial sense and navigation (proprioception), in its somatosensory cortex in the postcentral gyrus. The *homunculus* is often used to show distribution of the somatosensory cortex according to which body part it renders. The parietal lobe is furthermore important for language processing, mathematical analysis, and writing tasks. The occipital lobe is the visual processing centre and contains the vast majority of the visual cortex. The temporal lobe is involved in processing sensory input into derived meanings for the retention of visual memory, language comprehension and speech, hearing, learning, and emotional association.



(a) The regions of the brain (Marieb, 2015)



(b) The lobes of the cerebrum (Marieb, 2015)

Figure 2.1: The human brain (Marieb, 2015)

The cerebrum, also called cerebral cortex, is further divided into the left- and right cerebral hemispheres by a deep longitudinal fissure. Figure 2.1b also shows the ridges (gyri), grooves (sulci), and deep grooves (fissures) on the cerebrum. The entire CNS is protected by three connective tissue membranes that are collectively referred to as the meninges. The CNS is further protected by cerebrospinal fluid (CSF), which is a fluid with similar composition as blood plasma.

2.1.2 Aetiology and epidemiology of seizures

In this section some of the causes (aetiology) of seizures, and the prevalence (epidemiology) of the cause, are discussed briefly. First a distinction must be drawn between epileptic seizures and non-epileptic seizures (NESs). The aetiology of epileptic seizures are abnormal electrical activity originating the brain only, whereas the aetiology of NES are factors external to the brain that in turn induce abnormal electrical activity in the brain. For example, high temperature (fever), low oxygen levels, low blood sugar, poisons, and high levels of alcohol are all factors with origin external to the brain that may induce NESs.

Epilepsy is the most common cause for seizures, but is the second most common neurological disorder in humans, after stroke. Epilepsy is a chronic disorder

and its mechanisms are still poorly understood. There are a number of possible causes for a patient to develop epilepsy, namely brain injury, stroke, genetic disorders, birth defects, cerebral infection, or brain tumors. According to Shorvon (2005) there are a significant amount of cases of epilepsy that have unknown aetiology. In childhood the most common causes may be genetic disorders or birth defects, in adulthood non-genetic external factors are more likely, and among the elderly vascular diseases are an increasingly common cause. The prevalence of epilepsy is relatively high. It is estimated that 1 % of humans (approximately 65 million) suffer epilepsy. The highest incidence of epilepsy is found in the toddlers and the elderly (Forsgren, 2008; Shorvon, 2005).

Brain injury is the second most common cause of seizures. There are a number of causes for brain injury: A concussion occurs when the brain collides with the inner skull wall. A skull fracture results due to severe trauma, usually severe enough to cause brain injury. A haematoma is the collection or clotting of blood just outside the intracranial blood vessels, and leads to intracranial pressure build up. A haemorrhage is an uncontrolled intracranial bleeding. Brain injury can result in a number of symptoms: A swelling of the brain leads to an oedema, where intracranial pressure builds up and may cause the brain to press against the skull. A diffuse axonal injury, or sheer injury, is damage to the white matter neurons over a large area. Although no bleeding may occur, the result is permanent brain damage and even death. Traumatic Brain Injury (TBI) is a common pathology globally. In the United States of America the incidence is approximately 538 per 100 000 (Rutland-Brown *et al.*, 2006), while the incidence is estimated at 235 per 100 000 in Europe and 322 per 100 000 in Australia (Tagliaferri *et al.*, 2006). The incidence of TBI is peak in young adults (15 to 24 years) and the elderly (> 65 years). The epidemiology in developing nations are difficult to estimate, however it is reported that the incidence rate is growing due to an increase in motorization combined with inadequate traffic education and implementation (Roozenbeek *et al.*, 2013). The prevalence of seizures in TBI patients are estimated at between 5 % to 7 % (Teasell *et al.*, 2007).

Malignant hypertension (arteriolar nephrosclerosis) is an abnormally high blood pressure (above 180/120 mm Hg). This condition may lead to heart attack, stroke, and kidney failure, among others. Possible causes for hypertension include autonomic hyperactivity, head trauma, pre-eclampsia, and eclampsia, among others. Eclampsia is the condition whereby seizures are caused in women only during pregnancy. Fortunately eclampsia is a rare condition that follows pre-eclampsia, characterized by high blood pressure after the 20th week of pregnancy. Although rare, eclampsia is a highly serious condition with high risk of contraction in women that have: hypertension, diabetes, a history of poor diet or malnutrition, their first time pregnancy, pregnancy with twins, or are over 35 or under 20 years of age. Approximately 10 % of pregnancies

are complicated by hypertensive disorders, with eclampsia and pre-eclampsia accounting for about half of such cases worldwide (Hutcheon *et al.*, 2011). More generally, hypertension is suffered by between 1 % and 5 % of children, 15 % of young adults, and more than 60 % of adults above the age of 65 years (Sharifian, 2012). One study found that long-term severe hypertension increase the risk of seizures by 11 fold compared to a control group (Hesdorffer *et al.*, 1996).

Alcohol withdrawal delirium (AWD) is experienced by an approximately 51 % of individuals with alcohol addiction when they are denied alcohol. Of the individuals that do experience AWD, between 3 to 5 % experience *grand mal* seizures (see Section 2.1.3) and severe confusion (McKeon *et al.*, 2008).

Cerebral palsy is caused by abnormal brain development or injury to the developing brain before-, during-, or shortly after birth, and often causes seizures in the sufferer. Cerebral palsy cause motor-, coordination-, and posture disabilities in children. The prevalence of cerebral palsy is well over 2.0 in every 1000 births, with an incidence of epilepsy at 20 % to 40 % (Odding *et al.*, 2006). In another study, the prevalence of seizures in patients with cerebral palsy is estimated at as high as 62 % (Bruck *et al.*, 2001).

Cerebral hypoxia refers to the condition in which the brain is deprived of adequate oxygen supply at the tissue level. Cerebral ischaemia is the restriction of blood supply to neural tissue, causing cerebral hypoxia and neuroglycopaenia. Any disease or disorder of the brain is referred to as encephalopathy. Hypoxic ischaemic encephalopathy (HIE) is the condition that occurs when the entire brain endured a period of below normal oxygen level (but not total oxygen deprivation) due to inadequate blood supply. HIE (also called hypoxic brain damage) occurs most often as a result of cardiac arrest (CA), or neonatal asphyxia in the case of a neonate (Busl and Greer, 2010). According to the American Heart Association (AHA), prolonged untreated seizures are detrimental to the brain, and are common after return of spontaneous circulation (ROSC), occurring in 5 % to 20 % of comatose CA survivors (Peberdy *et al.*, 2010). According to the World Health Organization (WHO), in developed nations neonatal asphyxia affects 3-5 neonates per 1000 live births, with 0.5-1 neonates per 1000 developing HIE. The estimate for HIE in developing nations are difficult to estimate, but the danger posed by HIE is even greater than in developed nations (WHO, 2016). It has been estimated that only 30 % of neonatal encephalopathy cases are in developed nations (Kurinczuk *et al.*, 2010). HIE is incredibly harmful, as it causes approximately two thirds of neonatal seizures (Tekgul *et al.*, 2006).

Hyponatraemia is the condition whereby the sodium electrolyte levels in the blood is abnormally low. Hyponatraemia is the most common electrolyte abnor-

mality encountered clinically. Homoeostasis of intra-cellular water balance and blood pressure is complicated by hyponatraemia. In one retrospective study, hyponatraemia was the cause of seizures in 70 % of infants under 6 months old who lacked other aetiology (Farrar *et al.*, 1995).

Brain cancer is the uncontrolled overgrowth of brain cells that form masses called tumours. Whether cancerous or malignant, the brain tumour displays a characteristic of fast expansion, thereby disrupting normal bodily control and presenting a life-threatening condition. Fortunately brain cancer is an uncommon condition with far lower than 1 % chance of development in humans (American Cancer Society, 2017). The epidemiology of seizures due to brain tumours are dependent on tumour type, grade, and location. The prevalence is estimated at between 20 % to 45 % in patients with brain tumour (Maschio, 2012).

Hypoglycaemia is an abnormally low level of blood sugar (below 50 mg/dL), common in diabetics. The epidemiology of hypoglycaemia is difficult to ascertain. In one retrospective study, the frequency of hypoglycaemia ($\leq$ 55 mg/dL) in non-critical hospital admissions was 36 per 10 000 admissions (Nirantharakumar *et al.*, 2012). Another study found that a severe event (seizures or coma) due to hypoglycaemia had an incidence of 4.8/100 patient-years (Davis *et al.*, 1997).

A brain aneurysm (also called intracranial- or cerebral aneurysm) occurs when a weak point in the wall of an artery or vein that supplies blood to the brain dilates locally, thereby ballooning the vessel when the dilation is filled with blood. In the United States of America, the prevalence of brain aneurysms are 1 % to 5 % (10-12 million), with incidence 1 per 10 000 per year. The highest incidence is encountered in persons aged 30 to 60, with higher occurrence in women (Brisman *et al.*, 2006). A brain aneurysm, whether ruptured or not, may lead to seizures in the sufferer (American Heart Association and American Stroke Association, 2012). A ruptured aneurysm has incidence of seizure or epilepsy of about 11 %, whereas an unruptured aneurysm has an incidence of seizure or epilepsy of between 6 % and 9 % (Hoh *et al.*, 2011).

Encephalitis is an inflammation of the brain. Seizures and convulsions are known symptoms. Infection is the most common cause for encephalitis, with viruses being the most common aetiological agents. The incidence of encephalitis is estimated to be between 3.5 and 7.4 per 100 000 patient-years (Granerod and Crowcroft, 2007). One study found that seizures occurred in 42.6 % of patients with encephalitis (Misra and Kalita, 2009).

Other causes of seizures exist, however it is not within the scope of this thesis to review all possible causes. Evidently, from the discussion provided,

there are a host of possible causes for seizures. Patients suffering from any of the above conditions should ideally be monitored continuously for seizures to allow for early diagnosis and intervention.

2.1.3 Taxonomy of seizures

Seizure types are grouped and defined according to the classifications by the International League Against Epilepsy (ILAE, 1981). Seizure classification is based on clinical and EEG observation, instead of the underlying pathophysiological or anatomical distinctions. With regards to the origin within the brain, seizures are classified as either *focal* or *generalized*. A simple classification scheme is illustrated in Figure 2.2.

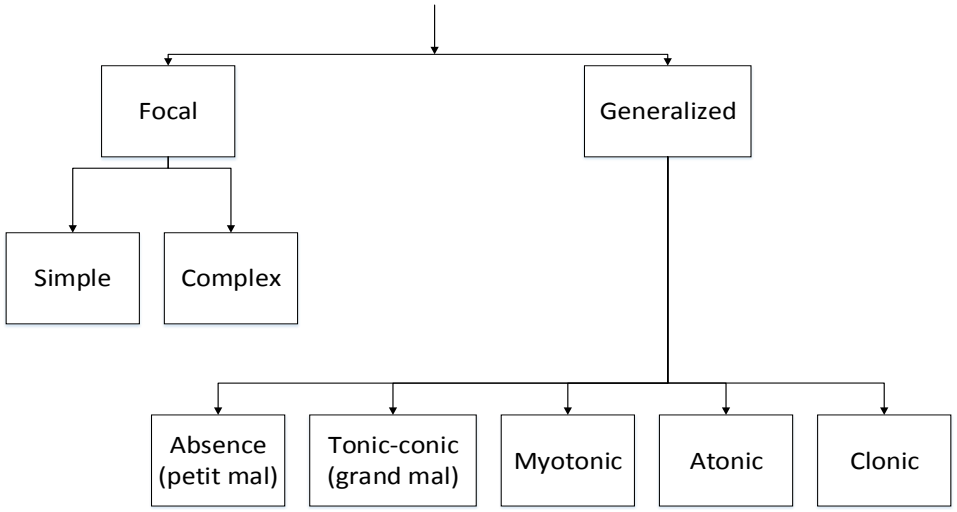


Figure 2.2: Seizure types

A focal seizure (also called partial- or localized seizure) originates in one location in the brain, either a cerebral hemisphere, or in part of a lobe of the cerebrum. When the patient retains consciousness during the partial seizure, then it is said to be a simple partial seizure. In contrast, if the consciousness of the patient is impaired, then the seizure is said to be a complex partial seizure.

Symptoms vary based on the location of the focal seizure (Epilepsy Foundation, 2013). In the frontal lobe, symptoms include a wave-like sensation in the head and unusual eye and body movements. In the parietal lobe, the feeling that the body is distorted, or that limbs are missing or foreign, difficulty in understanding language or doing simple maths, numbness, a tingling sensation or some other sensations are all symptoms. In the occipital lobe the symptoms are: visual disturbance, hallucination, uncontrollable eye or eye-lid movements (or the sensation thereof). In the temporal lobe a feeling of *déjà vu*, confusion,

and a difficulty in speaking. The connection between the symptom and its lobe of origin can be understood when considering the primary functions of each lobe as discussed in Section 2.1.1.

A generalized seizure distorts electrical activity of the whole, or a large portion of the brain. The types of generalized seizures include: absence, tonic-clonic, myoclonic, atonic, and clonic seizures. Absence (*petit mal*) seizures are characterized by a brief loss and return of consciousness. The tonic-clonic (*grand mal*) seizure are seizures that affect the entire brain and results in stiffening of the muscles, and then convulsions (repeated muscle jerking). The tonic-clonic seizures are perhaps the most well-known seizure type, since its effects are dramatic compared to all other relatively subdued seizure types. Myoclonic seizures are characteristically manifested clinically as brief jerks of muscles with the patient remaining fully conscious. Atonic seizures are a type of seizure that cause brief alteration in muscle tone, which can lead to loss in motor control for balance. Clonic seizures result in convulsions (clonus-phase), but unlike tonic-clonic seizures, there are no preceding stiffening of the muscles (tonus-phase).

Seizures may also be further classified according to the method of termination. There are two categories: self-limited epileptic seizures, and status epilepticus. Self-limiting (or self-terminating) seizures last up to several minutes and is terminated by natural mechanisms. Status epilepticus is a life-threatening condition whereby prolonged repetitive seizures continue for 30 minutes or more, since the natural mechanisms responsible for seizure termination fail.

2.2 Electroencephalography

Electroencephalography (EEG) is a method to record electrical activity in areas of the brain. This is achieved by placing non-invasive electrodes on the scalp of the patient. The device that measures and processes the data is called an electroencephalograph. These devices are capable of providing excellent temporal resolution with decent spatial resolution. An electroencephalogram (EEG) is the graphical output of the electrical activity of the brain from the electrodes.

Whereas a routine short (up to approximately 1 or 2 hours) duration brain-wave recording is called an EEG, long-term (longer than routine) recordings are termed continuous EEG (cEEG or CEEG). Despite this differentiation, cEEG is often simply referred to as EEG. Confusingly, EEG is also sometimes referred to as 'conventional EEG' and abbreviated cEEG or CEEG. The terms EEG and cEEG/CEEG refer strictly to recordings with electrodes placed on the scalp. In this work, the term EEG shall be used to denote scalp EEG (whether

short- or long-term). When electrodes are placed inside the scull, it is called intracranial EEG (iEEG or IEEG) or electrocorticography (ECoG).

2.2.1 Mechanisms of measurement

The electrophysiological signals measured by scalp electrodes are weakened due to obstructing structures of the skull, skin, and meninges that separate the electrodes from the neuronal layers. The electrodes can only measure signals that are of sufficient duration and strength such that it is observable. For neurons it is typically only the excitatory- and inhibitory post-synaptic potentials (EPSP and IPSP, respectively) that can readily fulfil those criteria and is the most significant source of EEG potentials (Olejniczak, 2006). EEG has a poor measure of the subcortical parts of the brain, because it is even further obstructed by the cerebral cortex. For this reason it is mainly the neural activity from the cortex (cerebrum) that is measured.

The primary excitation neurons found in the cerebrum are called pyramidal neurons (or pyramidal cells). These neurons have a triangular shaped soma, a large apical dendrite, multiple basal dendrites, and dendritic spines. What makes pyramidal cells special is the abundance of Na^{+}, Ca^{2+}, and K^{+} ion channels in its dendrites, their synchronous timing and parallel alignment with one another, and the fact that they are superficial to the skull, so they emit the strongest electrophysiological signal of the neuron types. When depolarization begins at the dendritic end of the neuron, repolarization occurs at the axonal end. The change in polarity causes a dipole over the neuron, thus conducting a current. The activity of one single neuron is not strong enough to record, but when thousands or millions of neurons are active at the same time, then the combined signal is observable by EEG electrodes. When an electrophysiological signal is generated by the summed electric current flowing from neurons within a small volume of nervous tissue, the small volume is referred to as a *local field potential* (LFP). It can then be said that EEG electrodes measure mostly the currents that flow during synaptic excitations of the dendrites of pyramidal neurons in a LFP on the cerebral cortex (Teplan, 2002).

When currents flow in the aforementioned LFP, it is as a result of ion pumping across neuronal membranes. When the positively charged ions move out of the neuron, they repel other ions in their neighbourhood, which in turn repel other ions in its neighbourhood, and so on, in a wave. This process is referred to as volume conduction. An LFP is large enough such that when ions are pushed out of the neuron, the volume conduction wave reaches up to the scalp, which pushes or pulls electrons in the electrode metal. The magnitude of the push or pull on electrons is different for all electrodes, and so the difference in push/pull between any two electrodes can be measured using a voltmeter. The peak-to-peak voltages for such recordings are typically in the microvolt

(μV) range, despite the fact that actual neuronal voltages are in the millivolt (mV) range. The EEG recording of a typical adult lies between 10 to 100 μV in amplitude, while the actual neuronal signal is 10 to 20 mV (Aurlien *et al.*, 2004). When the electrode recording is plotted over time, the graphical output is called the EEG. The graphical lines themselves are often referred to as 'brain waves' or 'waveforms'.

2.2.2 Artefacts

Artefacts are features in the EEG which do not originate from the cerebrum. Artefacts may therefore be considered as noise in the EEG. Artefacts may be divided into physiological artefacts, and non-physiological artefacts. The origins of artefacts as described by Fisch (1999) are summarized next.

Physiological artefacts are due to movements, bioelectric potentials, or skin resistance changes. Movements of the head, body, scalp, or other skeletal muscles generate electrical activity called Electromyography (EMG). Bioelectrical potentials are induced by moving electrical potentials (which can be caused by the eye, tongue and pharyngeal muscle movement) or by electrical potentials from the scalp muscles, heart, or sweat glands. Electrical activity caused by heart beats are termed electrocardiography (EKG or ECG). Skin resistance changes are due to sweat gland activity, perspiration, and vasomotor activity.

Non-physiological artefacts arise mainly due to external electrical interference, or internal electrical malfunctioning. Noise is generated by alternating current at the frequency of mains electricity and is called mains (or electric- or power line-) hum/noise. Mains noise can be caused by nearby appliances, transformers, or wiring and is set at 50 Hz or 60 Hz, depending on local power line frequencies. Internal electrical malfunctioning of the recording system arise from faults with the electrodes, electrode positioning method, or amplifiers.

Fortunately artefacts can, in many instances, be identified immediately by applying the following two spatial analysis rules:

1. Medium to high amplitude potentials that occur at only one electrode are classified as artefacts.

2. Repetitive, irregular or rhythmical waveforms that appear simultaneously in separate regions on the scalp are classified as artefacts.

2.2.3 Brain wave frequency bands

Neural oscillation at various frequencies are present in brain waves. Most rhythmic activities in the cerebrum fall in the range 1 to 20 Hz. Oscillations above or below this bracket are likely to be artefacts (Section 2.2.2). The

frequency of EEG activity is often divided into a number of frequency bands. There is no universal convention as to exactly how many bands there are and what the exact ranges are for each band should be, and as such the ranges will vary slightly between sources (Niedermeyer and Lopes da Silva, 2012; Noachtar *et al.*, 1999). Only some of the most common thresholds for these frequency bands are given.

- The delta (δ)-band ranges from 0 Hz to 4 Hz.
- The theta (θ)-band ranges taken from 4 Hz to 7 or 8 Hz.
- The alpha (α)-band ranges from 7 or 8 Hz to 12, 13, 14, 15, or 16 Hz.
- The beta (β)-band ranges from the upper limit of the α-band to an upper limit usually taken from somewhere between 30 to 40 Hz.
- The gamma (γ)-band begins from the upper limit of the β-band, but its upper limit is unbounded. The upper limit is sometimes taken as 70 Hz, since artefacts are encountered above this frequency.

In the δ-band, for practical purposes, the lower frequency limit is sometimes taken as 0.5 Hz, since DC potential differences are not monitored in conventional EEG (Noachtar *et al.*, 1999). The lower limit is alternatively taken as 1 Hz, since the bandwidth below this is described either as noise, or the K complex (Steriade *et al.*, 1993; Niedermeyer and Lopes da Silva, 2012). The β and γ bands are sometimes combined and simply referred to as the β band.

For the purposes of this study, a simplistic frequency band division is presented in Table 2.1.

Table 2.1: Brainwave frequency bands

Band	**Lower limit [Hz]**	**Upper limit [Hz]**
δ	> 0	4
θ	> 4	8
α	> 8	16
β	> 16	32
γ	> 32	∞

2.2.4 Instrumentation

The modern digital electroencephalograph device always has the following components: electrodes, input box (also called jack box or electrode/input board), calibrator, filters, amplifiers, A/D (Analog to Digital) converter, and recording device (with write out). Figure 2.3 shows a flowchart for a typical

modern digital system.

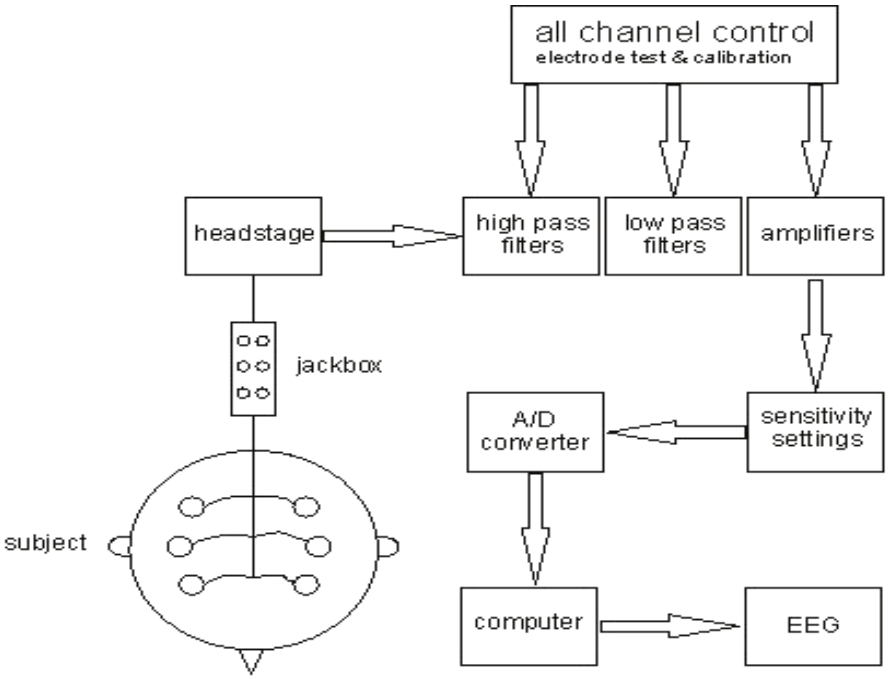


Figure 2.3: Digital EEG instrumentation (Karmos and Dombovári, 2011)

The *electrodes* are shown on the head of the subject in Figure 2.3. The function of the electrodes are to measure the brain waves from the scalp of the subject. The terminal end of the electrodes are inserted into the *input box*. In many modern devices, a pre-filter amplifier is included in the input box in order to mitigate external electrical interference. The role of the *calibrator* is to supply a precise preselected voltage, which lie within the range of EEG signals, to the input of all amplifiers.

Adjustable high- and low-pass *filters* are used to suppress the signals outside the frequency band of interest. Signals outside this band are typically noise. In routine EEG recording, it is suggested that the high-pass filter should be no higher than 1 Hz, and the low-pass filter should be no lower than 70 Hz, to prevent loss of useful information (Ebner *et al.*, 1999). Noise is generated by alternating current at the frequency of mains electricity, either 50 or 60 Hz (depending on local power-line frequencies). This noise is called mains (or electric- or power line-) noise. Mains noise can be caused by nearby appliances, transformers, or wiring. Mains noise should ideally be removed from EEG signals. A 50 Hz or 60 Hz notch filter should be used only if no other measure is sufficient, because the notch filter can distort sharply contoured components in the EEG. In fact, filters will distort to a degree both amplitude

and inter-channel phase of signals, so it is ideal to minimize the use of filters (Ebner *et al.*, 1999). In modern EEG amplifiers, the task of filtering is done by the amplifier discussed next.

It has been mentioned that the electrical signals recorded by electrodes are in the order of microvolts (μV). These signals must be magnified by an *amplifier* such that they may be digitized accurately and made compatible with display, recorder, and A/D converter devices (Teplan, 2002). The differential amplifier is used to measure the potential difference between two (active) electrodes. This refers to a subtraction of one signal (relative to the reference electrode) from another signal (also relative to the reference), and then amplification of that difference. What the differential amplifier achieves is to suppress signal variations that are common (common modes) to both electrodes, this is called common mode rejection (CMR). This is useful for instance in the filtering of the electric hum. Two electrodes will both have their signal distorted in almost identical fashion by the electric hum, but the effect of this will be reduced by CMR. When two similar input channels are subtracted from each other ($In1 - In2$), the result is a straight line with spikes where the inputs differ. By the polarity rule convention, when the spike is an upward deflection, then In1 is negative with respect to In2. When the spike is a downward deflection, then In2 is negative with respect to In1. In such an unintuitive convention, it may be useful to think of the term 'negative' as meaning 'large'.

Modern digital EEG systems take the amplified signal and digitize it using an *A/D converter*. The A/D converter is interfaced to a computer system where each sample is saved in memory. The *recording device* stores or records the data such that it may be used for write out. In modern EEG devices, the recording device is a computer and the write out is done on a monitor.

For a more detailed description of the recommended standards and specifications for EEG device components, refer to Ebner *et al.* (1999) and Teplan (2002).

2.2.5 Electrode placement and montages

The benefits of subtracting two signals to implement CMR is discussed in Section 2.2.4. EEG output waves displayed on a monitor are thus always the magnitude of a particular signal relative to another. A channel is therefore one continuous line of EEG recording, which is the difference between the two signals connected to the differential amplifier. A derivation is a description of the relation between the two electrodes (for example In1 - In2). Deciding which electrode signals to subtract from each other depends on the desired EEG montage (set of derivations), and each montage is associated with its own set of advantages and disadvantages. To understand the definitions of mon-

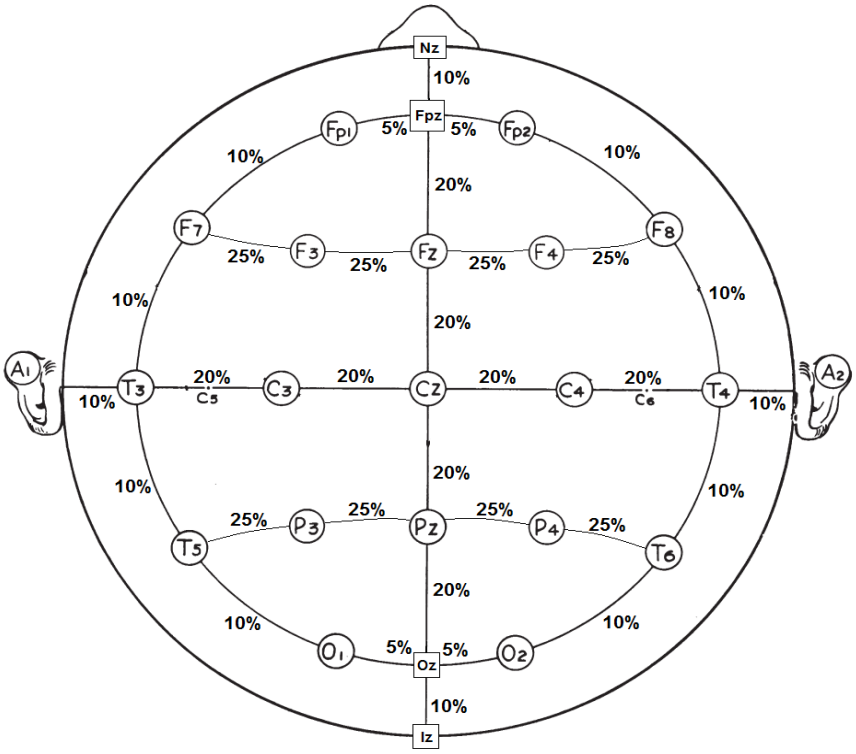
tages, the location that electrodes are placed on the scalp must be discussed first.

The international 10-20 system is the standard for the placement of scalp electrodes for EEG applications. The placement locations for electrodes are based on known skull landmarks, so that the standard could be applied to any person. The landmarks used are the nasion, inion and the left and right pre-auricular points. The pre-auricular points are felt as depressions at the root of the zygoma, just anterior to the tragus. The placement method is distance dependent and is described using percentages of the distance between landmarks. A detailed guide for placement is prescribed by the International Federation of Clinical Physiology (IFCP) (Klem *et al.*, 1999), but will not be repeated here.

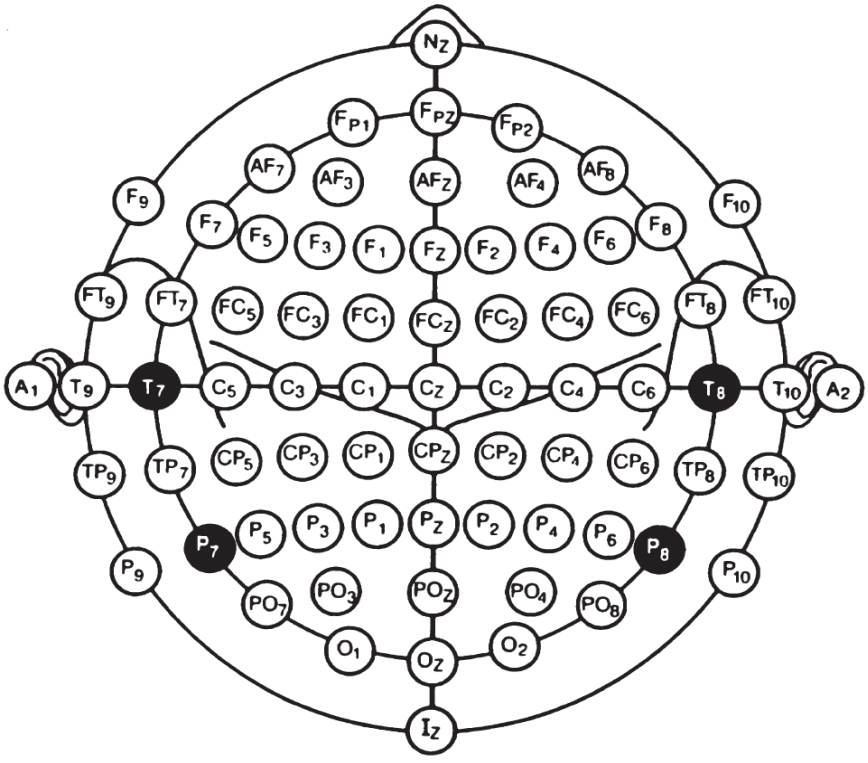
The international 10-20 system electrode location and relative distances are shown in Figure 2.4a. In total 21 electrodes are used for the international 10-20 system. Note that the points that are enclosed by a rectangle are not electrode positions, they are merely landmarks. The international 10-20 system requires the electrode designation to be expressed in terms of what area measurement is done. In Figure 2.4a the locations are denoted Fp (Fronto polar), F (Frontal), C (Central), P (Parietal), O (Occipital), T (Temporal), and A (Auricular). The auricular electrodes are placed on the earlobes. Some designations are followed by the letter 'z', which denotes 'zero', as these are vertex electrodes. The landmarks Nz and Iz are the nasion and inion respectively. When additional electrodes are required, the international 10-10 system (also called 10 % system or extended 10-20 system) may be used, as shown in Figure 2.4b. The new locations are halfway between locations of the 10-20 system. Note that for the 10-10 system, the positions highlighted in black have different designations compared to the top illustration. These alternate designations are also sometimes used in the 10-20 system.

The most common EEG montages are (in descending order): Bipolar (also called Sequential) montage, Referential (also called Common electrode reference) montage, Average (also called Common average) reference montage, and Laplacian montage.

In the Bipolar montage, each channel represents the difference between two adjacent electrodes. The standard direction for bipolar montages are the interior-posterior (longitudinal) direction. Considering only the electrodes in the 10-20 system: when the adjacent electrodes of Figure 2.4a are connected by a hypothetical line, it forms a grouping of electrodes, called a chain. The longitudinal direction has 5 chains (left to right by convention): [Fp1, F7, T3, T5, O1], [Fp1, F3, C3, P3, O1], [Fz, Cz, Pz], [Fp2, F4, C4, P4, O2], and [Fp2, F8, T4, T6, O2]. The derivations for chain 1 are Fp1-F7 (channel 1), F7-T3 (channel 2), T3-T5 (channel 3), T5-O1 (channel 4). Derivations and channels are derived in similar fashion for all 5 chains, in the sequence given, such that



(a) The 10-20 system (adapted from Klem *et al.* (1999))



(b) The 10-10 system (Klem *et al.*, 1999)

Figure 2.4: Placement of EEG electrodes (Klem *et al.*, 1999)

there are a total of 18 channels (channel 1 to channel 18). A less common direction in which to take the bipolar montage is the transverse (coronal) direction. The transverse bipolar montage also has 5 chains (top to bottom by convention): [Fp1, Fz, Fp2], [F7, F3, Fz, F4, F8], [A1, T3, C3, Cz, C4, T4, A2], [T5, P3, Pz, P4, T6], and [T5, O1, O2, T6]. This montage will have 19 channels if using only the electrodes in the 10-20 system.

2.3 Seizure analysis

2.3.1 Seizure analysis terminology

Some basic seizure analysis and machine learning terminology used in this research must be clarified:

- An *epoch* is a continuous segment of EEG record.
- The *ictal* epoch is the epoch in which the seizure itself occurs.
- The *inter-ictal* epoch is the time between successive ictal epochs, *i.e.* the non-seizure state.
- *Preictal-* and *postictal*-epochs refer to the time directly before and after a seizure, respectively.
- In *seizure detection*, the aim is to discriminate EEG signals in the ictal state from the signals in the inter-ictal state.
- In *seizure prediction*, the aim is to discriminate EEG signals in the pre-ictal state from the signals in the inter-ictal state.
- *Seizure analysis* refers collectively to the problems of seizure detection and seizure prediction.
- *Focus channel* is the electrode location that exhibits the earliest evidence of ictal activity or, if this is simultaneous in more than one channel, the channel in which the activity is maximal in amplitude.
- A *class-imbalance problem* in machine learning applications refers to applications with datasets where one class constitutes a small fraction relative to other classes. Seizure analysis problems are an example of a class-imbalance problem, since it is estimated that ictal data represent less than 0.05 % of all data (Gardner, 2004).
- A *binary classification* problem in machine learning refers to problems where data must be classified between one of two classes only. The binary classes are often termed as the ‘positive’ and ‘negative’ class.

- The *positive class* is all ictal data. The *negative class* is all inter-ictal data.
- A *positive declaration* refers to the event whereby the classifier declares a sample of data as ictal (positive class). A *negative declaration* refers to the event whereby the classifier declares a sample of data as inter-ictal (negative class).
- *Windowing* refers to the process of segmenting data into discrete data 'blocks'.
- In a *block processing* approach, data is windowed and advanced in time.
- Typically, data is imported from a pool of data in *frames*. A frame is the smallest data window used in an application.
- The *training set* is a portion of all available data that is used to create the optimal classifier.
- The *testing set* is the remaining portion of all available data. The test set is used to evaluate the optimal classifier. The performance of the classifier as applied to the test set is reported.

More Machine Learning terminology is provided in Section 2.5.1.

2.3.2 Performance metrics

In this subsection the performance metrics used in seizure detection/prediction algorithms are detailed. When a classifier labels an epoch (as ictal or inter-ictal), the classification is in reality one of the following:

- True Positive (TP): The classifier declares an ictal interval that overlaps with an actual ictal (pre-ictal for prediction) interval.
- True Negative (TN): The classifier declares an interval as inter-ictal, and that interval is indeed inter-ictal.
- False Positive (FP): The technique declares an ictal interval, which is actually inter-ictal.
- False Negative (FN): The technique declares an interval as inter-ictal, but that interval is actually ictal.

The classification outcomes could be visualized by the confusion matrix given in Figure 2.5. In statistics, a FP is also called a *Type 1 error* and a FN is also called a *Type 2 error*.

	Detected/ predicted label = ICTAL	Detected/ predicted label = INTER-ICTAL
Actual label = ICTAL (or pre-ictal)	TP	FN
Actual label = INTER-ICTAL	FP	TN

Figure 2.5: Confusion matrix

In seizure analysis problems, if a single positive declaration is made for a given seizure, then the entire seizure is considered detected and a single TP is counted. Additional positive declarations for the same seizure are not counted as additional TPs. Similarly, if no positive declaration is made within the seizure duration, then the seizure is missed, and a single FN is counted. In contrast to this, every FP is counted, regardless of vicinity to other FPs. This harsher rule is used, because each time a positive declaration is made the monitor alarm would go off to notify ICU staff.

In continuous EEG evaluation, there is a difficulty with using the TN metric. If the TN is counted in the same way as the FP, *i.e.* every sample of data can be declared TN, then the TN value would often be incredibly large and offer little performance information. This is a common in class-imbalance problems. In methods that use short-sample evaluation (see Section 2.3.3), data is partitioned into short, equal duration epochs. Only if the entire epoch has no positive declarations is a TN counted. An issue with this is that the selection of the epoch duration is still arbitrary, so selecting a smaller epoch size will invariably yield high TN. Another option is to select TN as the entire epoch between successive seizures, but this would likely yield high TN for patients with seizures in short succession, and does not add valuable performance information. These options confound comparison of methods, which is the entire purpose of performance metrics. Instead, publications in literature that test their method on long-term continuous data to simulate clinical conditions often take TN as a **duration**. This implies that the TN for a given patient is the duration of time for which the detector had TN declarations. Since TN is taken as a duration, it can only be added to other durations. Metrics TP, FP, or FN are taken as durations when added to TN.

Based on the labels described above, the performance of a classifier can be measured by creating performance metrics. There is unfortunately no standard

name or symbol for the metrics in seizure analysis literature. The most common performance metrics are given:

- True Positive Rate (TPR), also called *Sensitivity* or Recall or Probability of detection: The percentage of existing seizures that are correctly detected/predicted.

$$TPR = \frac{TP}{TP + FN} \tag{2.1}$$

  This metric is most often referred to as sensitivity in seizure analysis publications, and is likely the most consistently reported metric.

- *False Positive Rate* (FPR), also called Fall-out or Probability of false alarm: The percentage of inter-ictal epochs incorrectly classified as ictal.

$$FPR = \frac{FP}{TN + FP} \tag{2.2}$$

  Since TN is taken as a duration, this metric estimates the number of times that a classifier would produce a false positive over the entire inter-ictal recording duration. FPR is often reported as a rate per hour, and is a very often reported metric.

- Positive Predictive Value (PPV), also called *Selectivity* or Precision: The percentage of detections that are indeed seizures.

$$PPV = \frac{TP}{TP + FP} \tag{2.3}$$

  This metric is most often referred to as selectivity in seizure analysis publications.

- *Detection latency* (t_{dl}), also called Detection delay: The time lag associated with seizure detection.

$$t_{dl} = t_d - t_o \tag{2.4}$$

  where t_d is the time at which a seizure is detected and t_o is the time at which the onset of the seizure actually occurs according to the EEG data. Prediction has occurred if $t_{dl} < 0$. This metric is commonly used in seizure detection applications where it is reported in seconds.

- *Prediction horizon* (t_{ph}), also called Prediction time: The amount of time before seizure onset that a seizure was predicted.

$$t_{ph} = t_o - t_d \quad \equiv \quad -t_{dl} \tag{2.5}$$

  where t_d is the time at which a seizure is detected and t_o is the time at which the onset of the seizure actually occurs according to the EEG data. Prediction has occurred if $t_{ph} > 0$. This metric is commonly used in seizure prediction applications where it is reported in minutes.

- There are some other statistical measures that are used in seizure analysis literature. Defining each of these is not in the scope of this thesis, however they are listed below:
  - True Negative Rate (TNR), also called Specificity
  - False Negative Rate (FNR), also called Miss-Rate
  - Negative Predictive Value (NPV)
  - Accuracy (ACC)
  - False Omission Rate (FOR)
  - False Discovery Rate (FDR)
  - Prevalence
  - Positive Likelihood Ratio ($LR+$)
  - Negative Likelihood Ratio ($LR-$)
  - Diagnostic Odds Ratio (DOR)

Publications in literature most often report only sensitivity (TPR), false positive rate (FPR), and detection latency (t_{dl}) or prediction horizon (t_{ph}).

Ideally, the entire confusion matrix should be reported for each patient. The confusion matrix metrics should be reported in terms of count and duration (where applicable). Along with the confusion matrix, the detection latency (or prediction horizon) should be reported. Using only these metrics, any other metrics can be formed. The problem with **only** reporting more refined metrics such as TPR, FPR, and PPV is that researchers use various metrics to evaluate the efficacy of their algorithm. Comparison of algorithm performance is then hindered. Furthermore, if only mean performance metrics are reported (instead of reporting metrics for each individual patient), then the inter-patient variability of the algorithm performance cannot be assessed. The issue with inter-patient variability is that the algorithm is not robust, implying that for some patients, the solution is not reliable. This is undesirable in clinical technology. Medical staff prefer having knowledge of the accuracy of medical technology before even using it, as this can greatly influence decision-making.

A given classifier will output metrics TP, TN, FP, and FN based on its parameter settings. To optimize the parameters, a grid-search approach can be used to generate classifier performance over many parameter settings. The model parameters can be used to 'tune' the classifier toward good metrics. In order to decide whether one set of classifier parameters are superior to the next, the metrics must be combined into a single metric or *objective function*, such that a single measure of performance can be compared.

The formation of the objective function is an open-ended problem in seizure analysis applications. There are no universal standards for creating such a function, and so each researcher is left to define or select one for themselves. Some researchers define new objective functions that is formed by some new combination of a few of the performance metrics described. There are a number of common combinations used in research. One metric that has become more popular in recent years in the seizure analysis community is the F_1 score (also called F-score or F-measure). F_1 is the harmonic mean of precision (PPV) and recall (TPR). The harmonic mean for positive real numbers $x_1, x_2, ..., x_n$ is given by:

$$H = \frac{n}{\sum_{i=1}^{n} \frac{1}{x_i}} \tag{2.6}$$

The F_1 is the harmonic mean of PPV and TPR:

$$F_1 = 2 \cdot \frac{1}{\frac{1}{PPV} + \frac{1}{TPR}} = 2 \cdot \frac{PPV \cdot TPR}{PPV + TPR} \tag{2.7}$$

The harmonic mean is used when the average of two rates (also called ratios) are required. Since PPV and TPR are both rates, the harmonic mean is the preferred method of determining the average.

For example, consider the seizure analysis problem: two different models are compared. Model A outputs TPR=0.4 and PPV=0.6. The arithmetic mean is 0.5 for this model. Model B yields TPR=1 and PPV=0. The arithmetic mean is also 0.5 for this model. Clearly the first model is superior, since the second model merely labels all data as seizure. Using the harmonic mean, model A and B would have F_1 = 0.48 and F_1 = 0, respectively. Clearly the harmonic mean is a preferred averaging technique. The benefit of the F-score is that both TPR **and** PPV have to be high in order to obtain a high F_1. This metric is particularly convenient since it scales [0,1].

2.3.3 Types of methods and data

The types of classifier train-validate-test methods in seizure analysis vary significantly and the results from each have a unique meaning. Furthermore, the data used to accomplish these methods also play a role in the final interpretation or statistical power of the results. It is acknowledged that publications in literature have various intended applications. For example, some publications have the aim to develop implantable iEEG patient-specific devices, and others seek to develop offline post-monitoring tools.

Application

In this study, the *application* of the method is the intended use of the final algorithm. In seizure analysis publications, the application can be described as:

- A *patient-specific* method: data from the patient to be evaluated is required to train the classifier, and/or model parameters need to be tuned for the specific patient(s) to be evaluated.
- A *seizure-specific* method: morphologies (templates) of seizures of a given type and /or location is needed to train the classifier to detect similar seizures.
- A *generalized* method: strictly neither patient-specific nor seizure-specific.

Patient-specific and seizure-specific methods are intended for use by known epileptics, after a some prior EEG data have been collected from the patient. Generalized methods can be applied to a new patient, without requirement for any prior data collected.

Test type

Some methods in literature extract short samples of ictal and inter-ictal data to train, validate, and test their classifier. Commonly all seizures, and relatively short samples of non-seizure data is extracted from the training and testing sets. After classifier optimization using the training samples, the classifier would be evaluated using the short samples of the test set. Other methods in literature test their classifiers on long-term continuous EEG data to generate performance metrics. The long-term evaluation is considered a more reliable and statistically significant test of a seizure analysis classifier.

In machine learning data is commonly split into a training set and a testing set. The training set is used for classifier training and tuning. After the classifier is optimized, the classifier is applied to the independent (unseen) testing set. If the method does not keep its training and testing sets separate, *i.e.* there is significant overlap between the sets, then the machine learning classifier performance reported cannot be considered entirely reliable.

Data used

Researchers often collected data from the local ICU or hospital at great expense. If the data corpus is not made available to the public, it is referred to as a 'private' corpus. A data corpus that is made publicly available (whether free of charge or not) is termed a 'public' corpus. There are surprisingly few public corpora. Table 2.2 summarizes the public corpora used in seizure analysis.

Table 2.2: Public data corpora

Abbr	Details (at the time of writing)
Bonn	The University of Bonn's "Klinik für Epileptologie" EEG time series database (Andrzejak *et al.*, 2001). Contains 5 subjects, each with 100 single channel EEG records of 23.6 sec duration. 5 seizures are recorded. The segments are selected to contain no artefacts. The corpus is free and available online (Universität Bonn, 2017).
FSPP	The Freiburg Seizure Prediction Project (Winterhalder *et al.*, 2003) Contained 21 iEEG patients, 88 seizures. 128 channels. 509 hours of data is stored. The corpus was free and available online (FSPP, 2017), but it is now part of the EU database.
FHS	The Flint Hills Scientific, L.L.C., Public ECoG Database (Flint Hills Scientific, 2017*a*) Contains 1419 hours of continuous iEEG recordings for 10 patients, with 59 seizures, and 48 to 64 electrodes per patient. Over 1400 hours of EEG is recorded. The corpus was free and available online (Flint Hills Scientific, 2017*b*), however it is now unavailable.
CHB-MIT	The Children's Hospital of Boston - Massachusetts Institute of Technology data corpus (Shoeb, 2009; Goldberger *et al.*, 2000). Contains 24 EEG cases with 198 seizures and a total duration of 982.9 hours. Most records have 23 channels. The corpus is free and available online (PhysioNet, 2016).
KU	The Karunya University EEG database (Selvaraj *et al.*, 2014). Contains 176 records, each of duration 10 seconds. 16 channels of EEG is provided. The corpus is free and available online (Karunya University, 2017).
EU	The European Epilepsy Database was developed in the EPILEPSIAE project (Ihle *et al.*, 2010). Contains more than 250 epilepsy patients, 50 of which have iEEG taken, the remainder are EEG. There are 2400 seizures in the corpus. The total duration is more than 40 000 hours. The corpus is not free of charge, but is available (EPILEPSIAE, 2017).
TUH	The Temple University Hospital EEG Data Corpus (Obeid and Picone, 2016). The completed corpus comprises 16 986 sessions from 10 874 unique subjects. Minimum 24 channels. Corpus is free and available online (Picone, 2017).

Data corpora that are publicly available, but not free of charge can be purchased by anyone. The prices for such corpora are quite high, which impedes their use. Only publicly available data corpora that are free of charge may be downloaded directly from the source website. These are the more popular corpora.

From Table 2.2 it can be seen that the Bonn and KU corpora are only suited for short-sample methods. At the time of writing [1], the CHB-MIT and EU datasets are the only remaining available databases for long-term continuous EEG evaluation. Between these two corpora, only the CHB-MIT corpus is available free of charge.

Either EEG or iEEG/ECoG data is used. iEEG have several advantages over EEG: high-frequency signals are not attenuated by the skull and scalp, the signals are furthermore not heavily contaminated by EMG artefacts, and smaller clusters of neurons can be monitored. This makes iEEG much cleaner than EEG. The use of iEEG is, however, limited to clinical necessity due to its invasive nature. The use of iEEG is generally restricted to use for planning/management of epilepsy surgery. EEG is non-invasive, relatively inexpensive, quick and easy to apply, and provides excellent temporal resolution. For these reasons, EEG is the only promising recording method for immediate monitoring of patients.

The size of the database used also influences the statistical power of the results. In machine learning it is known that simpler models with larger amounts of training data can show better performance than a complex model with less training data. A model trained with too little training data may not yield performance true to its potential. Furthermore, a classifier evaluated on only few different patients cannot claim strong generalization ability as can a classifier evaluated on many different patients. For these reasons the volume and specifics of the database used is relevant. Researchers should ideally report the number of patients, number of seizures, and total duration of data used for the training set and for the testing set.

Other factors

There are a number of other factors that may influence the interpretation of the results of a given method. Often the classifier is only trained and tested on seizures of a certain type(s) and/or location(s). The classifier performance on non-selected seizure types/locations is therefore unknown.

In other methods, only a single electrode or small cluster of electrodes are used. Often *a priori* knowledge of the focus channel is used to select channels, since the focus channel often provides the easiest and earliest differentiation between ictal and inter-ictal states. The classifier performance on non-selected

[1] The TUH EEG database was released in 2016. Unfortunately the author only discovered this resource late into 2017. The 'TUH EEG Seizure Corpus' is a subset of the entire TUH corpus that groups the EEG of patients suffering seizures. This subset can be used for automated seizure detection evaluation, and is conveniently separated from the rest of the corpus on the website (Picone, 2017).

electrodes is unknown, since they are not evaluated. In the case where the focus channel is selected, the performance of the classifier will likely only be sustained if the focus channel of the independent data is known *a priori*. One cannot simply train a classifier on one channel and then apply that classifier to any other channel. A classifier trained on a patient with focal seizures in, for example, the occipital lobes, would be trained on features from occipital electrodes since these would be the best at detecting the seizure. These classifiers would perform poorly if the patient being monitored had a seizure in, say, the frontal lobes (Fergus *et al.*, 2015). It is unfortunately impossible to know on which channel a focal seizure can be detected before performing the test. The problem is then that seizure detection/prediction cannot be generalized across multiple subjects with only one focus channel.

Some methods cannot be applied to real-time (online) applications. If any data after the actual seizure is required for classification, then the method is applicable only for offline use. If any form of data preselection (incl. manual artefact removal) is performed, then the method can obviously not be applied in real-time.

2.3.4 Publications in literature

Seizure analysis through computational methods was first introduced by Gotman *et al.* (1973). Since then a multitude of research papers regarding seizure analysis have been published. Discussing each publication would require an entire dedicated chapter in this study. Instead the extensive literature study is summarized in Table 2.3. Although mentioning all relevant papers is not within the scope of this study, Table 2.3 contains many of the most significant publications in the field of seizure analysis. The publications are listed according to the date released.

In Table 2.3, sensitivity (TPR), mean detection latency (t_{dl}) in seconds, and False Positive Rate (FPR) per hour are tabulated if reported. A negative t_{dl} indicates prediction. Further details of the methods are encoded (PS, SS, Gen, S, L, O, I, $1 \rightarrow 3$) in Table 2.3. The implications of these details were discussed in Section 2.3.3. The number of cases (patients), the number of seizures, and the total duration, in hours, of the data used (training and testing sets together) is given if reported.

Table 2.3: Literature Review

Publication	**Application*** [A]	**Test type*** [B]	**Data [cases, seizures, hours]**	TPR (%)	FPR (/h)	t_{dl} (sec)
Gotman (1990)	Gen* [3]	L,I	Private (EEG) Private (iEEG) All: [49,244,5303]	73 83	0.84 1.35	
Qu and Gotman (1995)	PS & SS	L,I	Private (mixed) All: [17, 77, ≈ 65]	100	0.21	9.6
Gabor (1998)	Gen	L,I	Private (EEG) [65, 181, 4550]	92.8	1.35	
Osorio *et al.* (1998)	Gen* [1]	S,I	Private (iEEG) [16, 125,]	100	0	2.1
Le Van Quyen *et al.* (1999)	Gen* [1]	S,I	Private (iEEG) [13, 23,]	83		-345
Le Van Quyen *et al.* (2001)	SS* [1,3]	S,O	Private (EEG) [23, 8,]	96		-420
Iasemidis *et al.* (2003)	Gen PS PS	L,O L,O L,I	Private (iEEG) [5, 55, 433]	82 84 82.6	0.16 0.12 0.17	-4302 -4464 -6018
D'Alessandro *et al.* (2003)	PS* [1]	L,I	Private (iEEG) [4, , ≈160]	62.5	0.2775	
Hively and Protopopescu (2003)	Gen* [1]	L,O	Private (EEG) [41, 40, ≈261]	87.5	0.021	-2100
Khan and Gotman (2003)	Gen	L,I	Private (iEEG) [22, 163, 451.5]	85.6	0.3	
Acır and Güzeliş (2004)	Gen	S,I	Private (EEG) [25, , ≈4.25]	90.3		
Gardner (2004)	Gen* [1,2]	L,O	Private (iEEG) [16, 118, ≈1078]	97.85	1.74	-13.3
Shoeb *et al.* (2004)	PS	S,I	Private (EEG) [36, 139, 60]	94	0.25	8
Wilson *et al.* (2004)	Gen	L,O	Private (EEG) [459, 670, 1514]	76	0.11	
Grewal and Gotman (2005)	Gen PS	L,I	Private (iEEG) [49, 312, 796.1]	86 89.4	0.47 0.22	16.2 17.1
Jia *et al.* (2005)	Gen* [1,2]	L,O	Private (iEEG) [2, 24, 14]	72.25	0.4	≈-690

Le Van Quyen *et al.* (2005)	PS & SS*[1,2]	L,O	Private (iEEG) [5, 52, 305]	70		- 11220
Saab and Gotman (2005)	Gen PS	L,I	Private (EEG) [28, 126, 652]	77.9 76	0.86 0.34	9.8 10
Meier *et al.* (2008)	SS	L,O	Private (EEG) [57, 91, 1403]	96	0.5	1.6
Aarabi *et al.* (2009)	Gen*[2,3]	L,O	FSPP (iEEG) [21, 78, ≈302]	98.7	0.27	11
Khamis *et al.* (2009)	Gen*[1,2]	L,I	Private (EEG) [12, 101, 1939]	91.57		
Kuhlmann *et al.* (2009)	Gen*[3]	L,I	Private (EEG) [21, 88, 525]	81	0.6	16.9
Patel *et al.* (2009)	Gen*[2] PS*[2]	S,O	FSPP (iEEG) [13, ,]	90.9 94.2		
Shoeb (2009)	PS	L,I	CHB-MIT (EEG) [23, 163, 844]	96	0.13	4.6
Minasyan *et al.* (2010)	PS*[1,3]	L,I	Private (EEG) [25, 86, ≈ 625]	100	≡ 0.04	≡ -26.8
Polychronaki *et al.* (2010)	Gen*[1,2,3]	L,I	Private (EEG) [8, 55, 553.14]	100	0.42	8.82
Park *et al.* (2011)	PS*[2,3]	S,I	FSPP (iEEG) [18, 80, 433.2]	97.5	0.27	
Williamson *et al.* (2011)	PS*[2]	S,I	FSPP (iEEG) [19, 83, ≈500]	90.8	0.094	
Yuan *et al.* (2012)	PS*[2]	S,I	FSPP (iEEG) [21, 86, 179.57]	91.72		
Fergus *et al.* (2015)	Gen*[3]	S,O	CHB-MIT (EEG) [24, 171, 5.7]	93		

Table with sensitivity (TPR), false positive rate (FPR) per hour, and mean detection latency (t_{dl}) in seconds. A negative t_{dl} indicates prediction.
*[A]The application is Patient-Specific (PS), Seizure-Specific (SS), or Generalized (Gen). Generalized solutions are strictly neither PS nor SS.
*[B] The test-type (final evaluation) is on discontinuous short-samples (S) or long-duration continuous data (L). The testing set significantly overlaps (O) or is independent (I) of the training set.
Some methods used data that *[1] have only 1 seizure-type or that *[2] have 1 channel/group of channels.
*[3] Some methods cannot be applied in real-time as they are developed for offline use or include some form of manual data or channel preselection.

2.3.5 Discussion of publications

In this study, the desired seizure monitoring algorithm is one that is designed for immediate application in the clinical environment. The desired system should begin monitoring as soon as a patient is rushed into the ICU and the scalp EEG system is connected properly. It is acknowledged that not all publications in literature have the aim of developing such a monitor.

Reliability and clinical application

Methods that are generalized (Gen), involve no limitations ($1 \rightarrow 3$) indicated in Table 2.3, and are evaluated on long-term (L) scalp EEG with independent (I) training and testing sets are considered more promising in achieving the aims of this study. With such prerequisites, very few methods may be considered promising.

From Table 2.3, it is evident that some of the best algorithm performances reported in literature are from publications that are less promising in achieving the aims of this work. As part of this study, some of these high performing methods were evaluated on a test that simulate clinical conditions. The purpose was to investigate the degree by which the excellent reported performance would degrade under such a test. In general it was found that the performance metrics of these methods degraded heavily when applied to evaluations that simulate clinical conditions. The original report (Volschenk *et al.*, 2017*a*) is summarized in terms of its overall results:

The 1-class ν-SVM method by Gardner (2004) was implemented, however with no set overlap, and no *a priori* focus channel knowledge. Good TPR with very poor FPR was obtained. The 2-class SVM method by Shoeb (2009) was implemented, however no patient-specific training was allowed. Unsurprisingly, very poor TPR and FPR was yielded. The 2-class SVM method by Fergus *et al.* (2015) was implemented, with no set overlap, and additional postprocessing procedures for FPR reduction. Good TPR with very poor FPR was obtained. In these evaluations, the LIBSVM 3.21 (Chang and Lin, 2011) software package was used. Finally, the promising Saab and Gotman (2005) method was implemented, with results similar to those reported in the original publication. These findings support the notion that methods that involve any of the limitations indicated in Table 2.3 are to be considered less promising in achieving the aims of this work.

Of the methods presented in Table 2.3, the Saab and Gotman (2005) and Gabor (1998) methods are considered most promising for application in this study. The reliability of the Saab and Gotman (2005) method is reinforced since the method was repeated with good results on an entirely different data

corpus (Kuhlmann *et al.*, 2009). Choosing between the methods is essentially a trade-off between sensitivity and FPR. Although the Gabor (1998) method yields an impressive TPR=92.8 %, its FPR=1.35 /h implies that the detector will have a false alarm 32.4 times in a single day. The Saab and Gotman (2005) method yields a more modest TPR=77.9 %, however its FPR=0.86 /h yields a more promising 20.6 false alarms per day. Evidently neither of these methods can be applied clinically, since the high false alarm rate would likely result in the monitor being switched off by the ICU staff. Due to its significantly lower FPR at the cost of $\approx$ 15 % sensitivity, the Saab and Gotman (2005) method shall be considered more applicable, and therefore the 'state-of-the-art', in this study.

The Saab & Gotman (2005) method

The Saab and Gotman (2005) method has been highlighted as promising for clinical application. The method is described only briefly here, but is clarified throughout the remainder of Chapter 2 as well as Chapter 3.

Clinical as well as sub-clinical seizures were included, only seizures that showed no EEG manifestation were excluded from analysis. Their method begins with filtering out frequencies below 0.5 Hz and frequencies above 70 Hz. Their data was sampled at 200 Hz, and represented in either 24 or 32 channel bipolar montage.

Data is segmented into 2 second frames (no overlap). For each channel in the frame, the signal is split into frequency bands using a 5-level db2 wavelet transform. Frequency bands 50-100, 25-50, 12-25, 6-12, 3-6, and 0-3 Hz were formed. These bands correspond to wavelet coefficients (scales) D1, D2, D3, D4, D5, and A5, respectively.

For each channel, only scales D3,D4,D5 are used for feature extraction. For each scale, the Relative Average Amplitude (RAA), Relative Scale Energy (RSE), and the Coefficient of Variation of Amplitude (CVA) are calculated.

A Naïve Bayes Classifier (NBC) is used for training and testing. Seizure training data was extracted and processed frame by frame. For each feature (in each scale in each channel), a histogram with 5 bins are constructed, where each bin has the same number of samples. The bins of all three scales for a given channel-scale are combined. When a frame is imported then, a channel-scale can fall in one of 5^3 = 125 bins. The ratio of these bin counts to the total number of training samples serve as *a priori* probabilities in the NBC. The same bin ranges determined from the seizure training data is then used to generate probabilities for the non-seizure training data.

EEG with high α-band activity that is uncorrupted by artefacts were searched for visually. Low α-band data without seizures was also searched for manually. These datasets were stored separately. The same formulation used to characterize seizure and non-seizure activity using the NBC was used to characterize the high α and low α data. The bin ranges are based off the high α training data. In this process, only features from D4 were used.

During classification, when a frame is imported, it is rejected if it contains significant non-physiological artefact corruption (abnormally high signal amplitude, main noise, or phase reversal). To determine the mains noise, spectral analysis was used. If the frame is not rejected, it undergoes the wavelet transform and feature extraction. For each channel scale (D3,D4,D5), the NBC outputs a probability. The probability that a seizure is detected in a channel is found by summing the probabilities of the three scales for that channel. The 6 channels with the highest probabilities are then summed to obtain P_{SEZ_EPOCH}. The probability that high α-activity is present is calculated in the same manner, except the probability for a given channel is equal to the probability from D4 only. The 6 channels with the highest probabilities are summed to obtain P_{ALP_EPOCH}.

The seizure probability is scaled by the EMG activity $P_{SEZ\_SCALED} = P_{SEZ\_EPOCH}(1 - EMGAmpRatio)$. Here $EMGAmpRatio$ is the ratio of the sum of amplitudes in D1 and D2 to the sum of amplitudes in D1,D2,D4,D5. The final detection variable P_{SEZ} is found by summing the P_{SEZ_SCALED} of the last three frames. The P_{ALP} is found by summing the P_{ALP_EPOCH} of the last three frames. The thresholds P_{TH} and A_{TH} are now used. If $P_{SEZ} > P_{TH}$ and $P_{ALP} < A_{TH}$, then the frame is declared seizure. If $P_{SEZ} > P_{TH}$ and $P_{ALP} > A_{TH}$ then the frame is declared seizure only if the top 6 channels summed for P_{SEZ_EPOCH} are not from different hemispheres of the brain. If a single frame is declared seizure, then a detection is considered declared.

2.4 Signal processing

Spectral analysis (also called spectrum analysis) refers to signal processing methods that allow a signal to be characterized by its spectrum of frequencies or related quantities such as energies, and eigenvalues, among others. In this study, the mains noise (either 50 or 60 Hz) need to be evaluated for each signal in each frame of data being processed. Furthermore, signals in each frame need to be separated by its frequency content. To obtain the mains noise from a signal, and to separate a signal into frequency-bands, spectral analysis must be done.

2.4.1 The Fourier Transform

One method of to perform spectral analysis is to apply the Fourier Transform to the signal. Brain-wave signals obtained *via* EEG are discrete signals. The mains noise in the EEG is caused by alternating current (AC), which is cyclical/periodic in nature. The component of interest is therefore part of the *discrete-periodic* signal class. A Fourier Transform applied to a discrete-periodic signal is called a Discrete Fourier Transform (DFT).

By far the most efficient method for DFT application is the Fast Fourier Transform (FFT), and there are a number of FFT algorithms. The FFT is one of the most complicated digital signal processing algorithms, however due to its incredible speed and accuracy relative to other methods, it is an indispensable algorithm widely used in science, engineering, and mathematics. To properly understand the FFT, a lengthy explanation is required. Such an explanation is not within the scope of this study. Instead the higher-order workings of the FFT shall be described in this section, based off the guide by Smith (2003).

The FFT performs three main tasks:

1. Decompose an n point time domain signal into n time domain signals (each is therefore composed of one point).
2. For each of the n time domain signals, calculate its individual frequency spectrum.
3. Synthesize the n frequency spectra into a single frequency spectrum.

Time domain decomposition

The time domain decomposition is shown with an example in Figure 2.6. In the figure, an example 16-point signal is decomposed. Four stages are required to break the signal into n=16 signals. When a signal is split in two, an *interlaced decomposition* is applied. This means the signal is separated into its even and odd numbered samples. Consider the signal {0 2 4 6 8 10 12 14} obtained after the first decomposition stage, as shown in Figure 2.6. The sample originally in position 14 is now in position 7 (numbering starts at 0). This sample is grouped with other samples with odd indices: 2, 6, 8, 10. Using this simple heuristic, a signal of any length n can be decomposed. What the decomposition achieved, in essence, is to exchange samples whose indices are binary *reversals* of one another. For example the final position of sample 1 (0001) is exchanged with the final position of sample 8 (1000).

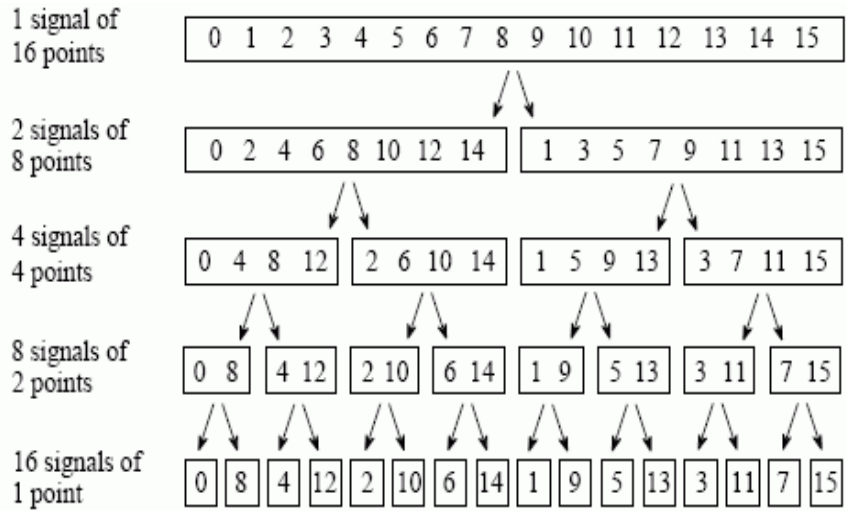


Figure 2.6: FFT time domain decomposition (Smith, 2003)

Determine frequency spectra

The FFT procedure must determine the frequency spectrum for each of the n time domain signals. Fortunately this is trivial, the frequency spectrum of a 1 point signal is simply equal to itself. There are now n frequency spectra.

Synthesize frequency spectra

The FFT must now recombine the n frequency spectra into a single frequency spectrum. This is done in the reverse order that the time decomposition took place. In the example, 16 frequency spectra (1 point each) are synthesized into 8 frequency spectra (2 points each). This synthesis procedure is repeated until the final 16 point frequency spectrum is formed. The procedure to synthesize frequency spectra is where the FFT algorithm becomes complex. An explanation of procedure that the FFT takes to synthesize frequency spectra is therefore not within the scope of this study. Further clarification is provided in the sourcebook by Smith (2003).

Output

The output of the Fourier Transform is the frequency domain data. The Fourier Transform essentially approximates the periodic time domain input signal using a set of sinusoids. When the sinusoids undergo summation, an approximation of the original signal is formed. The frequency domain data is typically plotted with the sinusoidal amplitude on the y-axis, and the frequency of the sinusoid on the x-axis. This plot over frequencies may be referred to as the frequency *spectrum*. An example is given in Figure 2.7. In this simple example, the input time domain signal shown is almost square-wave in appearance (on the left side of the image, in red). The Fourier Transform approximates the signal using,

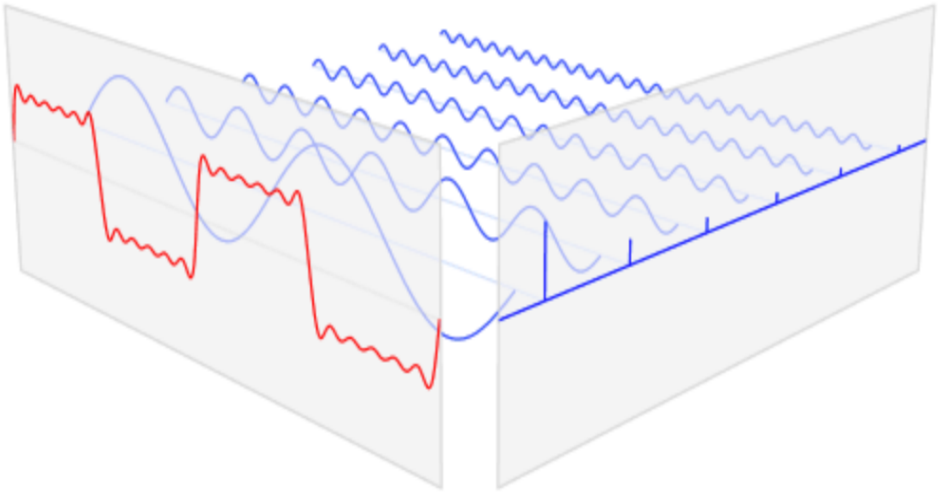


Figure 2.7: Time and frequency domains (Barbosa, 2013)

in this example, only 6 sine-wave sinusoids. In this example, the frequency domain is illustrated with the sinusoidal peak amplitude (also called half the sinusoidal peak-to-peak amplitude) as a function of the sinusoid frequency (on the right side of the image, in blue).

An EEG signal is a far more complex time domain signal than the simple signal in Figure 2.7. An EEG will be approximated by many more sinusoids, and so the frequency spectrum will have a full range of frequency components. When applying an FFT to an EEG signal, the amplitude of the mains noise (either 50 or 60 Hz) can be evaluated.

2.4.2 The Wavelet Transform

The Fourier Transform family is not the only method by which spectral analysis may be performed. The Wavelet Transform family is a popular alternative used in digital signal processing. To understand why a wavelet method may be preferred over the highly efficient FFT, some limitations of the Fourier Transform must be discussed. An overview based on the guide by Valens (1999) will describe briefly the motivation and procedures of the Wavelet Transform.

Real world signals, including EEG, often exhibit slowly changing *trends*, or *oscillations*, interspersed with abrupt changes, called transients. Transients are non-periodic waveforms. Transients within a signal often contain valuable information about the signal in question. The Fourier Transform represents data as a sum of infinite sinusoids, so we can determine every frequency present in the signal, but not *when* the signal is present. It is said that the Fourier Transform is well localized in frequency, but not well localized in time.

One method to overcome this is to window the signal by using a Windowed Fourier Transform (WFT). The problem is that the WFT causes frequency components to smear over the frequency axis. The phenomena is attributed to Heisenberg's uncertainty principle, which (in signal processing terms) states that it is impossible to know the exact frequency and time simultaneously in a signal. This principal leads to difficult trade-off between time- and frequency resolution. A solution to this problem is the Wavelet Transform. To understand how the problems faced by the Fourier Transform can be overcome, some basic Wavelet Transform concepts must first be discussed.

Properties

Wavelets are a class of functions that are well localized in both frequency and time, so they can be used to study local behaviour of a signal, like discontinuities or spikes. Wavelet functions shall be represented with $\phi(t)$. The most important properties of wavelets are the admissibility- and regularity conditions.

The *admissibility condition* implies that the Fourier Transform of $\phi(t)$ vanishes at zero frequency. This means that wavelets have a bandpass-like spectrum. It further implies that wavelets are wave-like oscillations, with zero mean in the time-domain. The *regularity condition* implies that the wavelet should have some smoothness and concentration in both the time- and frequency domains. Wavelets should be a rapidly decaying (*i.e.* compactly supported) function with some smoothness. Because of its rapid decay, wavelets exist for a finite duration. The regularity of a signal is a complex concept, and the theory of vanishing moments could be used to to help clarify regularity, however such an in depth explanation is not within the scope of this study.

Scaling

Scaling a signal refers to stretching or shrinking the signal in time. Scaling factor, s, corresponds to the degree with which a signal is scaled in time. The notation is $\phi(t/s)$, where $s > 0$. The factor s is inversely proportional to the signal's frequency magnification. For example, a 6 Hz sine wave scaled with s=2 will have a new frequency of 3 Hz. For a wavelet, there is a reciprocal relationship between the scale and the frequency, with a constant of proportionality called the *center frequency* (C_f) of the wavelet. Unlike the sinusoid, the wavelet has a band-pass characteristic in the frequency domain. The equivalent frequency (F_{eq}) is given by:

$$F_{eq} = \frac{C_f}{s \cdot \Delta} = \frac{C_f \cdot f_s}{s} \tag{2.8}$$

where Δ is the sampling interval/period, and f_s is the sampling frequency. We may say that scaling a wavelet by $s = 2$ reduces the F_{eq} by 2. Stretched

wavelets are used to capture the slow varying oscillations in a signal, whereas shrunken wavelets are used to capture abrupt changes.

Shifting

Shifting (or translating) a wavelet refers to delaying/advancing the onset of the wavelet along the length of the signal. The notation is $\phi(t-\tau)$, where the wavelet is centred at τ. The wavelet is shifted to align to the feature in the signal that need to be represented.

The Continuous Wavelet Transform

The Continuous Wavelet Transform (CWT) uses a fully scalable modulated window to solve the windowing problem. A window is shifted along the signal to determine the frequency spectrum at every instance. The process is then repeated many times with a slightly shorter (or longer) window for each iteration. The stretching/shrinking of the window is referred to as the scale. The collection of time-frequency representations is referred to as the time-scale representations in wavelet analysis. The CWT is given by Equation 2.9.

$$\Gamma(s,t) = \int f(t) \cdot \psi^*_{s,\tau}(t)\, dt \tag{2.9}$$

The equation shows how function $f(t)$ is decomposed into a set of basis functions, $\psi_{s,\tau}(t)$, called wavelets. The * operator indicates complex conjugation.

The wavelets that $f(t)$ is decomposed into are generated from a single template wavelet, called the *mother wavelet*. The wavelets $\psi_{s,\tau}(t)$ are generated from the mother wavelet $\phi(t)$ by scaling (s) and translation (τ) as shown in Equation 2.10.

$$\psi_{s,\tau}(t) = \frac{1}{\sqrt{s}} \cdot \psi\left(\frac{t-\tau}{s}\right) \tag{2.10}$$

Discrete Wavelets

The CWT is not an efficient method for three reasons:

1. There are a large number of redundancies in the procedure of continuously shifting a continuously scalable function over the signal to determine correlation.

2. There are an infinite number of wavelets in the Wavelet Transform.

3. Most signals will have no analytical solutions, they can only be calculated numerically or using an optical analog computer.

To overcome these problems, discrete wavelets must be used. The wavelet is therefore scaled and translated discretely. The scale is set using $s = s_0^j$. The translation is set using $\tau = k \cdot \tau_0 \cdot s_0^j$. Note that j and k are integers, and τ_0 is the translation factor. Equation 2.10 is modified:

$$\psi_{j,k}(t) = \frac{1}{\sqrt{s_0^j}} \cdot \psi\left(\frac{t - k \cdot \tau_0 \cdot s_0^j}{s_0^j}\right) \tag{2.11}$$

Usually $s_0 = 2$ is selected. The use of base 2 for this procedure is often referred to as *dyadic sampling*. Furthermore $\tau_0 = 1$ is usually selected such that we also have dyadic sampling of the time axis. To completely remove the final redundancies of the CWT method, the discrete wavelets must be made orthonormal. This procedure can only be done on discrete wavelets. To make wavelets orthonormal to their own scales and translations, special choices for the mother wavelet must be made such that:

$$\int \psi_{j,k}(t) \cdot \psi^*_{m,n}(t)dt = \begin{cases} 1 & \text{if } j = m \text{ and } k = n \\ 0 & \text{otherwise} \end{cases} \tag{2.12}$$

The redundancy of the CWT procedure is therefore removed.

There are still an infinite number of scaling and translations to calculate the Wavelet Transform. The number of allowable translations and scales must be minimized. The translations of the wavelets are limited by the length of the signal analysed, so the upper bound on translation is set. The question is then, how many scales are needed to adequately represent the signal?

One of the properties of wavelets are that they have a bandpass-like spectrum. Fourier theory states that compression in time is equivalent to stretching the frequency spectrum and shifting it upwards. This means that a time compression of the wavelet by a factor a will stretch the frequency spectrum of the wavelet by a factor a and also shift the frequency components up by a factor a. This means that the finite frequency spectrum of the signal analysed can be covered with the spectra of scaled wavelets. This is analogous to how the time-domain is covered by translated wavelets. To adequately cover the entire signal spectrum, the scaled wavelet spectra should overlap slightly. This can only be done by properly designing the wavelets. The first scale will be low, such that the top half of the frequency spectrum is covered. Subsequent wavelets are stretched to cover lower frequency spectra. One problem is that every time the wavelet is stretched, only half the remaining spectrum is covered. This means that an infinite number of wavelets would be needed to cover the frequency spectrum to zero. The solution is to simply not cover the entire spectrum with wavelets. Instead a low-pass spectrum, also called an *averaging filter*, is used to cover the lower frequencies. The infinite set of wavelets at lower frequencies is replaced by what is properly

referred to as the *scaling function*, also called *father wavelet*, and is denoted by ϕ. Note scale information in frequencies covered by the scaling function is lost. For this reason the wavelet spectra should reach low enough to capture scale information in low frequencies. The extent of this is application-specific. As an example, if the entire EEG frequency range needs to be analysed, then frequencies below ≈ 1 Hz may be covered by the scaling function since signal content below 1 Hz is likely noise (Section 2.2.2). Figure 2.8 shows how the signal spectrum is covered by wavelet frequency bands and a scaling function.

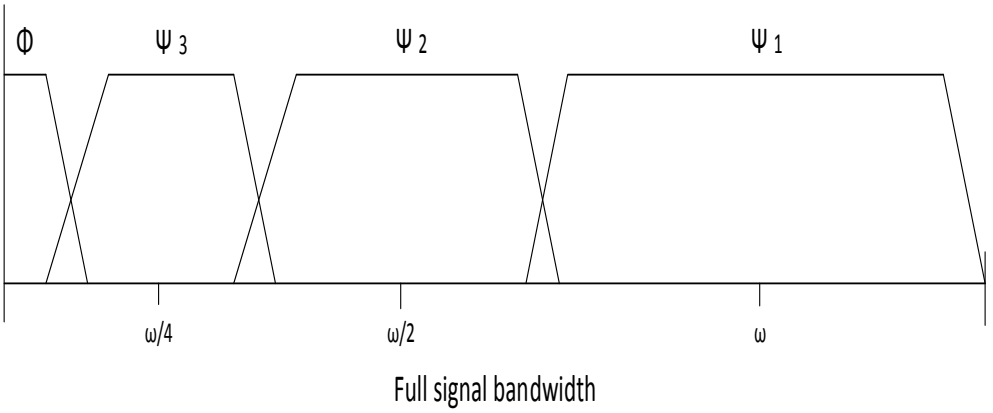


Figure 2.8: Covering the frequency spectrum

The scaling function is valuable as it replaces the infinite wavelets required beyond scale j. There are now therefore a **finite** number of wavelets. In summation then, a series of scaled wavelets together with a single scaling function acts as a series of bandpass filters with one low pass filter, respectively. The wavelets with the scaling function acts as a filter-bank.

Subband coding

Applying a filterbank to a signal is referred to as *subband coding*. The Wavelet Transform is an example of subband coding, more specifically, it is an iterated filterbank. The signal is split into high-pass and low-pass parts. Each split is subsampled by a factor 2. The low-pass part has some valuable information, so it is split again. This process is iterated until the information output by the low-pass part is not useful. By doing this, the Wavelet Transform is implemented as an iterative filterbank, and the wavelets do not need to be explicitly specified.

The Discrete Wavelet Transform

In many practical applications the signal analysed is a discretely sampled signal. To implement the discrete wavelet on such a signal, the Wavelet Transform

must be discretized. Currently the discrete wavelets are only discrete with scale and translation, but not time-discrete. The wavelet filterbank **can** however simply be implemented as a digital filterbank. In this study, the EEG channels in each frame need to be separated into frequency-bands.

The Discrete Wavelet Transform (DWT) is commonly used to implement the filterbank. Figure 2.9 illustrates the procedure with a simple example. The original signal, $x(t)$ is subband-coded using 4 levels of decomposition.

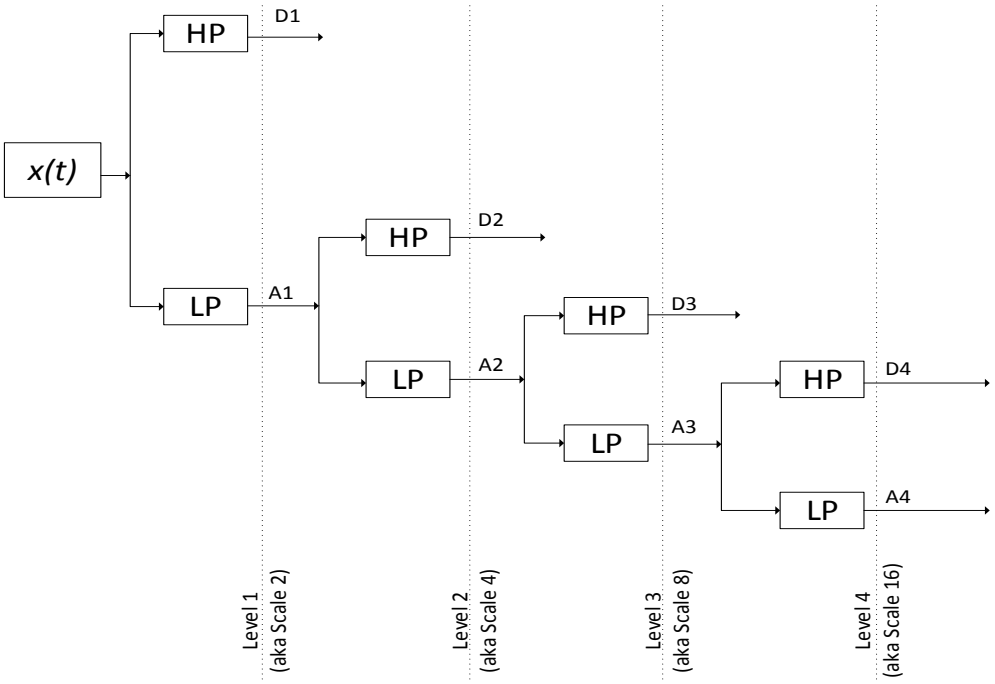


Figure 2.9: Discrete Wavelet Transform

Let $x(t)$ be a 200 Hz signal. According to the Nyquist-Shannon sampling theorem, a signal with sampling rate 200 Hz can have frequencies from 0 to 100 Hz resolved. The frequencies from the level 1 high-pass filter is 50 - 100 Hz. The scale used is $s = 2^1$ for wavelets. Since wavelets are used to describe this band, the output is detailed coefficients of level 1, denoted D1. The remaining spectra is covered by the scaling function, so wavelets are not used. The remaining spectra 0 - 50 Hz is referred to as the approximate coefficients A1. Both A1 and D1 are subsampled by a factor of 2, each now a 100 Hz signal. The next step is to apply level 2, *i.e.* scale $s = 2^2$. Here the signal from A1 undergoes the same process as level 1. This 100 Hz signal can have frequencies 0 - 50 Hz resolved. Detail coefficients D2 for 25 - 50 Hz are obtained from the high-pass filter and approximate coefficients A2 for 0 - 25 Hz are obtained from the low-pass filter. Again the resulting signals are subsampled by a factor of 2. The signal A1 is not kept, since the to half of its spectrum has

been 'detailed' by wavelets. This process is repeated to yield D3 (12.5 - 25 Hz), D4 (6.25 - 12.5 Hz), and A4 (0 - 6.25 Hz). Since no additional levels are implemented, the signal A4 is not discarded. The final output of the 4-level DWT is D1, D2, D3, D4, and A4.

2.4.3 Application

The strengths and weaknesses of the FFT and DWT algorithms have been discussed. In our application each signal in each frame must be checked for mains noise. If no signal in the frame is rejected, then each signal must be sent into a filterbank to form subbands.

It is evident that in order to separate each signal into its frequency bands without heavy loss of information, the DWT will need to be used, This is important, since the information output from the DWT is used in further characterization of the signal.

The speed of the FFT could be used to evaluate the periodic mains component in each signal in the frame. This procedure is computationally inexpensive, so if the frame is corrupted with noise, then the frame is rejected without being processed by the more computationally expensive wavelet filterbank.

2.5 Machine learning

2.5.1 Machine learning terminology

Some machine learning terminology is already provided in Section 2.3.1. Some additional explanations are provided below:

- In the training phase, some parameters ('A') are selected and a model is created with which classification can be done. The training set data is used in this process, except for the hold-out data.
- In the validation phase, some other parameters ('B') are selected that are applied to the classifier. The parameters 'A' may also influence the classifier. The classifier is then used to evaluate the held-out training set data. From evaluation, some new parameters 'A' and 'B' may be selected.
- The parameters 'A' may significantly influence parameters 'B', but parameters 'A' cannot be evaluated directly from the validation phase, they are required for the model creation step. The parameters denoted 'A' are appropriately termed *hyperparameters*, whereas the parameters denoted 'B' are simply termed *parameters*.

- After the parameters and hyperparameters are optimized, the optimized classifier is used in the testing phase. In the testing phase the independent testing set is evaluated by the optimized classifier. No further tuning or optimization is allowed. The performance results are reported.
- The training and validation phases may be repeated over and over again. These phases are often collectively termed the *offline phase*.
- The testing phase may only be performed once, as if the data is streaming in real-time. This phase is alternatively termed the *online phase*.

2.5.2 Naïve bayes classifier

The Naïve Bayes Classifier (NBC) is also referred to as the simple- or independence Bayes classifier in some literature. The NBC refers to a family of simple probabilistic classifiers that are based on the Bayes theorem, where a naïve assumption is made of independence between features. In probability theory, two events are (statistically/stochastically) independent if the occurrence of one does not affect the probability of occurrence of the other.

The NBC uses a conditional probability model. For a given sample of data represented by features (also called evidence) $\mathbf{x} = \{x_1, x_2, ...x_n\}$, the n features are assumed independent. The sample may be classified in one of k classes C $= \{C_1, C_2, ...C_k\}$. The probability that the sample is in class 1 is given by $P(C_1|\mathbf{x})$. The notation is read as "P of C_1 given $\mathbf{x}$". This refers to the probability of C_1 being true, given observations $\mathbf{x}$.

More generally, the array of probabilities for an array of classes C, given $\mathbf{x}$, are referred to as the *probability distribution* of C given $\mathbf{x}$, denoted $P(C|\mathbf{x})$. To determine $P(C|\mathbf{x})$, *Bayes' theorem* is used:

$$P(C|\mathbf{x}) = \frac{P(\mathbf{x}|C) \cdot P(C)}{P(\mathbf{x})} \tag{2.13}$$

In Equation 2.13 the following constituents are used:

- $P(C|\mathbf{x})$ is the probability distribution over C, given $\mathbf{x}$. This term is also referred to as the *posterior*.
- $P(\mathbf{x}|C)$ is the probability distribution over $\mathbf{x}$, given C. This term is also referred to as the *likelihood function*.
- $P(C)$ is the probability distribution over C. This term is also referred to as the *prior*.
- $P(\mathbf{x})$ is the probability distribution over $\mathbf{x}$. This term is also referred to as the *normalization constant*.

The prior may be described as the probability distribution of observing each class. The normalization constant may be described as the probability distribution that each given feature is observed in the training data. Both of these can be determined from the training data. The NBC may classify a sample by comparing the posterior to some threshold value.

One convenient aspect of the NBC is that its terms can be updated immediately as soon as new evidence is presented. No computationally-expensive iterative optimization is required. The ease with which probability distributions can be updated bodes well for dynamic learning applications.

Chapter 3

Methodology

3.1 Data preparation

3.1.1 Data corpus

The CHB-MIT data corpus described in Shoeb (2009) was retrieved electronically from PhysioNet (2016) (website by Goldberger *et al.* (2000)). The corpus was collected by a team of researchers from Children's Hospital Boston (CHB) and the Massachusetts Institute of Technology (MIT). Data was collected on site at CHB and consists of scalp EEG recordings from subjects with intractable seizures several days after withdrawal from AEDs.

This dataset was selected for a number of reasons: the large size of the dataset makes it statistically significant, the large number of channels available is good for optimal feature selection, and the fact that the data is presented raw and contains all the most common artefacts of EEG recording (Shoeb, 2009) makes it a good data corpus for estimating clinical performance of a classifier. The corpus contains male and female patients, with a wide range of ages, and with seizures of different types and locations. This diversity presents a thorough challenge to seizure detection algorithms. Finally, the fact that the corpus is publically available means that the results obtained in this study is reproducible. A short description of the most important aspects of the corpus is detailed on PhysioNet (2016) and is summarized next.

Recordings are grouped into 24 cases and were collected from 23 subjects. Case number designations range from **chb01** to **chb24** as shown in Table 3.3. Case **chb01** is the same subject as **chb21**, where **chb21** is a recording 1.5 years after **chb01**. Since **chb01** and **chb21** are significantly separated from each other chronologically, their data characteristics should not be assumed to be similar. For this reason **chb01** and **chb21** shall be considered and referred to as separate subjects/patients for machine learning purposes in this research.

For each case there are multiple EEG recording files. The beginning '[' and end ']' of each seizure is annotated in the **.seizure** annotation files by the CHB-MIT project experts. In total there are 198 seizures and 982.9 hours of EEG data in the corpus. Information about the montage used for each recording and the elapsed time in seconds from the beginning of each **.edf** file to

the beginning and end of each seizure is given in the **chb*nn*-summary.txt** files.

All signals were sampled at 256 samples per second with 16-bit resolution. Data is contained in continuous **.edf** files. Files are usually 1 hour in duration, but for cases **chb04**, **chb06**, **chb07**, **chb09**, and **chb23** the files are usually 4 hours in duration, and for case **chb10** the files are usually 2 hours in duration. Hardware limitations resulted in short gaps between consecutively-numbered **.edf** files, during which signals were not recorded. Generally the time gaps are about 10 seconds or less, but some are much longer.

The majority of files contain 23 EEG channels in the bipolar montage, and the international 10-10 system is used. Channel derivations and labels according to the CHB-MIT convention are given in Table 3.1. In some records

Table 3.1: CHB-MIT channels

Label	**Derivation**	**Label**	**Derivation**	**Label**	**Derivation**
1	$FP1$-$F7$	9	FZ-CZ	17	$T8$-$P8$
2	$F7$-$T7$	10	CZ-PZ	18	$P8$-$O2$
3	$T7$-$P7$	11	$FP2$-$F4$	19	$P7$-$T7$
4	$P7$-$O1$	12	$F4$-$C4$	20	$T7$-$FT9$
5	$FP1$-$F3$	13	$C4$-$P4$	21	$FT9$-$FT10$
6	$F3$-$C3$	14	$P4$-$O2$	22	$FT10$-$T8$
7	$C3$-$P3$	15	$FP2$-$F8$	23	$T8$-$P8$
8	$P3$-$O1$	16	$F8$-$T8$		

non-EEG signals such as ‘ECG’, ‘VNS’, or dummy signals (named “-") were interspersed among EEG channels, but these can be ignored. Channel 19 is the inverse derivation of channel 3, and that channel 23 is the inverse derivation of channel 17, so channels 19 and 23 offer no additional information.

3.1.2 Data set formation

Channel selection

Although most files in the CHB-MIT corpus contain at least the 23 channels given Table 3.1, a significant number of files do not have channels 19 through 23. These files are summarized in Table 3.2. The files in Table 3.2 only have the 18 standard bipolar channels of the international 10-20 system. In order to use the same channels from each record it was decided to only use channels 1 through 18 that are common to all records in the bipolar montage.

Table 3.2: Files with only 18 channels

Case	Files	Number of seizures
chb13	04→16, 18, 24, 30, 36→40, 47	2
chb15	01	0
chb16	18,19	2
chb17	c_13	0
chb18	01	0
chb19	01	0
TOTAL		**4**

Exclusions

In the CHB-MIT corpus there are some records that are not in the bipolar montage. In case **chb12** files 27→29 are not in the bipolar montage. These files were excluded from the data set used in this study. In total 13 seizures (within 3 hours of data) are excluded from analysis.

Dividing the data sets

It was decided to partition the corpus into a ratio of 80:20 for training and testing sets, respectively. For the testing set it was desirable to select cases with a single long continuous recording session. A single *recording session* is defined to be consecutive recording files. Records are consecutive if the start time of one file is no more than 6 minutes after the end time of the previous record. Table 3.3 uses the definition of recording session to tabulate data for each case after exclusions (Section 3.1.2) were removed. The corpus holds diverse seizure types, including Simple Partial (SP), Complex Partial (CP), and Generalized Tonic-Clonic (GTC) seizures from various locations over the brain. These locations include the Frontal (F), Temporal (T), and Occipital (O) lobes on either hemisphere of the brain (Nasehi and Pourghassem, 2013).

As shown in Table 3.3, there are 5 cases that are only 1 continuous session. These cases are assigned to the testing-set. In the testing phase, all data will be processed sequentially in order to simulate clinical conditions. The training set simply consists of all patients that are not in the testing set. During the train-validate phase, data from some patients in the training set is used for training the classifier, and the data of the remaining patients is used for validating the classifier performance. The division of the training set is explained in Section 3.6.4. Although it has been decided to consider cases **chb01** and **chb21** as separate patients, the fact that these are both assigned to the training set further assures no possibility of patient-specific tuning between the train-validate and test phases.

Table 3.3: Case information after exclusions

Case	Gender	Age [years]	Seizure type	Seizure origin	Recording sessions	Duration [hours]	Seizures
chb01	F	11	SP,CP,GTC	T	3	40.6	7
chb02	M	11	SP,CP	F	5	35.3	3
chb03	F	14	SP,CP,GTC	T	6	38	7
chb04	M	22	SP,CP,GTC	T, O	5	156	4
chb05	F	7	SP,CP	F	1	39	5
chb06	F	1.5	SP,CP,GTC	T	3	66.7	10
chb07	F	14.5	SP,CP,GTC	T	1	67.1	3
chb08	M	3.5	SP,CP,GTC	T	3	20	5
chb09	F	10	SP,CP	F	1	67.9	4
chb10	M	3	SP,CP,GTC	T	6	50	7
chb11	F	12	SP,CP	F	8	34.8	3
chb12	F	2	SP,CP	F	5	20.7	27
chb13	F	3	CP, GTC	T, O	10	33	12
chb14	F	9	SP,CP,GTC	T	11	26	8
chb15	M	16	CP,GTC	F, T	12	40	20
chb16	F	7	SP,CP,GTC	T	1	19	10
chb17	F	12	SP,CP,GTC	T	7	21	3
chb18	F	18	CP, GTC	T, O	9	35.6	6
chb19	F	19	SP,CP	F	8	29.9	3
chb20	F	6	SP,CP,GTC	T	8	27.6	8
chb21	F	13	SP,CP,GTC	T	8	32.8	4
chb22	F	9	CP, GTC	T, O	6	31	3
chb23	F	6	SP,CP	F	3	26.6	7
chb24	?	?	?	?	1	21.3	16
TOTAL						979.9	185

Table 3.4 details the final data set division. The desired train:test ratio of 80:20 was achieved, since the testing set has 5 of 24 cases (20.8 % of patients), 38 of 185 seizures (20.54 % of seizures), and 214.3 of 979.9 hours of data (21.87 %).

Table 3.4: Data sets

Data set cases	Number of seizures	Number of data hours
Training : 01, 02, 03, 04, 06, 08, 10, 11, 12, 13, 14, 15, 17, 18, 19, 20, 21, 22, 23	147	765.6
Testing : 05, 07, 09, 16, 24	38	214.3

3.2 Experimental design

3.2.1 Experimental layout

Figure 3.1 illustrates the layout of the experimental design in this research by means of pseudo-code. The individual steps, parameters, and phases that are

Figure 3.1: Experimental Design

```
1  while Π_h(r) optimization not converged do
2  |  Select Π_h(r).
3  |  while Π_h(m) optimization not converged do
4  |  |  Select Π_h(m)
5  |  |  Training phase        //generate models
6  |  |  while Π_p optimization not converged do
7  |  |  |  Select Π_p
8  |  |  |  Validation phase        // obtain performance metrics
9  |  |  end
10 |  end
11 end

12 Training phase        // using entire training set
13 Testing phase        //obtain final performance of classifier
```

given in the experimental layout pseudo-code are detailed in this section, as well as throughout the remainder of Chapter 3. The $\Pi_{h(r)}$, $\Pi_{h(m)}$, and Π_p sets are defined in Section 3.6.2. The training-, validation-, and testing phases are subroutines that are discussed next.

The training, validation, and testing phases are detailed by means of pseudo-code in Figure 3.2. The steps and parameters given in Figure 3.2 are detailed in this section, as well as throughout the remainder of Chapter 3. The train-phase is further detailed in Section 3.4. The classification procedure used in the validation and testing phases are detailed in Sections 3.4.5, and 3.5. Performance metrics generated are discussed in Section 3.6.1. The rationale behind using 'hold-out' cases for cross-validation is discussed in Section 3.6.4. Note that in the testing phase, the non-'hold-out' cases are the entire training set.

3.2.2 Software architecture

In Figure 3.2, the **train steps** were given in the training phase. The **validate-** and **test steps** were given in the validation- and testing phases, respectively. The train and validate/test software architectures used in this study is illustrated

Figure 3.2: Experimental phases

```
Function Training phase
    for ictal, then α do
        for each non-'hold-out' case in the training set do
            train step        //store ictal/α features
        end
        determine bin ranges
        for each non-'hold-out' case in the training set do
            train step        // Determine ictal/α likelihood model
        end
        for each non-'hold-out' case in the training set do
            train step        //store inter-ictal/low-α features
        end
        for each non-'hold-out' case in the training set do
            train step        // Determine inter-ictal/low-α likelihood model
        end
    end
    for each non-'hold-out' case in the training set do
        Determine and store ictal posterior model
        Determine and store high-α posterior model
    end
end

FunctionFunction Validation phase
    validate step        // hold-out case only
                         // Generate performance metrics
end

FunctionFunction Testing phase
    for each case in the testing set do
        test step        // generate final performance metrics
    end
end
```

in Figures 3.3 and 3.4. Note that the software architectures are always repeated

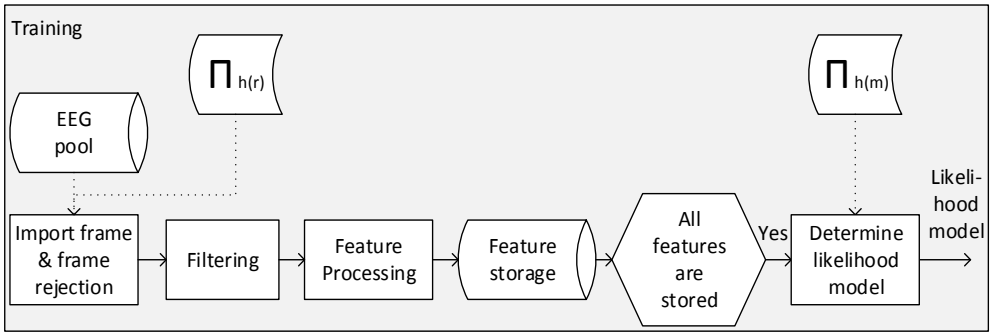


Figure 3.3: Training software architecture

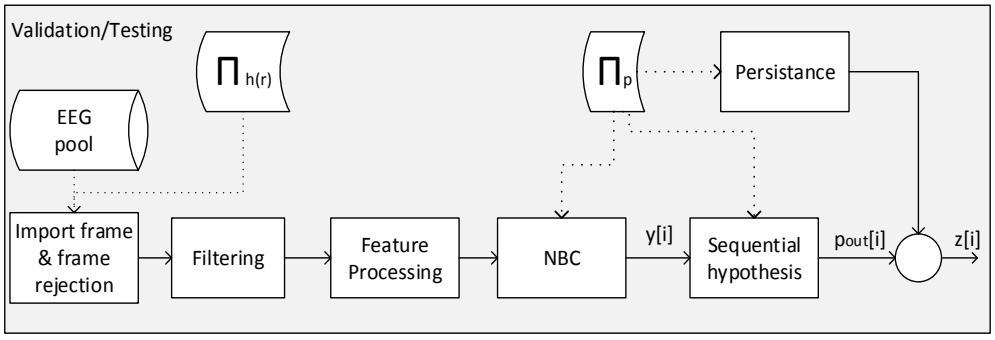


Figure 3.4: Validation/testing software architecture

in a loop until all data in the 'EEG pool' have been imported and processed. The meanings of each block function, signal, and parameter in Figures 3.3 and 3.4 are explained throughout the remainder of Chapter 3.

3.3 Preprocessing

3.3.1 Block processing and frame rejection

In this study the frame size is selected to be 2 seconds in duration with no frame overlap. A frame imported contains 2 seconds of data from channels 1 through 18.

After a frame is imported, it is evaluated for data corruption. If the data of a given frame is corrupted, then the entire frame is rejected instead of being passed to the next process. A data frame is considered corrupted if the data indicates electrode failure due to high levels of non-physiological artefacts. The non-physiological artefacts considered in this step are:

- Mains noise, caused by nearby appliances, transformers, or wiring.
- Abnormal channel amplitude, caused by electrode movement, amplifier disconnection and reconnection, recording phase faults, or physiological interference.
- Phase reversal in channels containing the same loose electrode.

An example each non-physiological artefact is given in Appendix A.

Mains noise

The mains noise in the CHB-MIT corpus is at 60 Hz, since the corpus is taken from the United States of America (Boston, MA). If the 60 Hz component in any channel exceed some threshold x_{mains} then the frame should be rejected. In order to determine the 60 Hz component in a given channel, a spectral analysis can be performed. For each channel in the frame, the Fast Fourier Transform (FFT) is used to obtain spectral information. If the 60 Hz component is greater than x_{mains} (given in μV) then the frame is rejected.

Abnormal channel amplitude

If the amplitude of any channel in the frame equals or exceeds the maximum channel-amplitude threshold x_{high} (given in μV) at any point in time, then the frame is rejected.

Inspection of the CHB-MIT data showed that, for some cases, records had instances of missing data of various durations, originally represented as dashes. The dashes are stored and displayed as zeros in MATLAB. Although no explanation is given for these dispersed clusters of zeros in the data, possible reasons could include recording phase faults, or physiological symptoms such as body movement or sweat interference with electrodes.

The "high signal amplitude" check proposed by Saab and Gotman (2005) was added to in this study to reject missing data in a frame that would otherwise not be rejected. If the amplitude of any channel in the frame equals **exactly** zero at any point in time, then the frame is rejected.

Phase reversal

When an electrode is loosened, the channels using that electrode will display phase reversal. When two channels (A and B) share an electrode, the channels form a *channel-pair*. Channel-pairs (A-B) are {1-5, 1-2, 2-3, 3-4, 4-8, 5-6, 6-7, 7-8, 9-10, 11-15, 11-12, 12-13, 13-14, 14-18, 15-16, 16-17, 17-18}. The order of channels A and B in the given electrode-pairs is arbitrary.

To detect phase reversal, the following heuristic is performed for each channel-pair:

$$\text{If } \; mean(|A+B|) < mean\left(\left|\frac{A}{x_{phase}}\right|\right) \; \text{ then reject frame} \tag{3.1}$$

A new parameter x_{phase} is used to set the strictness of the phase-reversal heuristic. If a single channel-pair in the frame has status 'reject', then the entire frame is rejected.

3.3.2 Filtering

Band-pass filter

In order to shed a large portion of artefacts, each channel in the frame is filtered to remove low- and high-frequency content. A 2^{nd} order Butterworth bandpass filter is applied to each channel between 0.5 to 70 Hz. The 2^{nd} order filter offer good transition band characteristics at low coefficient orders and as such they can be implemented efficiently. Experimenting with higher order filters showed serious instability around 0.5 Hz.

Wavelet Filter-bank

A filter-bank is used to separate each channel in the frame into frequency bands. For each channel, a 5-level wavelet transform is applied using a Daubechies 4-tap (also called D4, also called DAUB4) wavelet. Since the DAUB4 wavelet has 2 vanishing moments, it is also often denoted db2. The db2 wavelet is recommended in Khan and Gotman (2003) and well documented in the use of non-stationary signals, including EEG (Gabor *et al.*, 1996; Gabor, 1998; Osorio *et al.*, 1998; D'Alessandro *et al.*, 2003; Shoeb *et al.*, 2004; Saab and Gotman, 2005).

With signal frequency of 256 Hz, the 5-level wavelet transform produces decomposition scales D1, D2, D3, D4, D5, and A5. Since the bandpass filter have been applied to the 0.5-70 Hz range, the decomposition scales have bandwidths D1 (64-70 Hz), D2 (32-64 Hz), D3 (16 - 32 Hz), D4 (8-16 Hz), D5 (4-8 Hz), and A5 (0.5-4 Hz). Decomposition scales D1, D2, D3, D4, D5, and A5 approximate the brain-wave frequency bands γ(high), γ(low), β, α, θ, and δ, respectively. Figure 3.5 illustrates the filtering procedure.

Scale A5 is discarded due to high levels of activity in that band observed during inter-ictal sleep. Furthermore, seizures with content in a very low frequency range often have some content in a higher frequency range as well (Saab and Gotman, 2005). According to Gotman (1982), seizure activity ranges primarily between 3 and 29 Hz. For this reason, seizure activity is

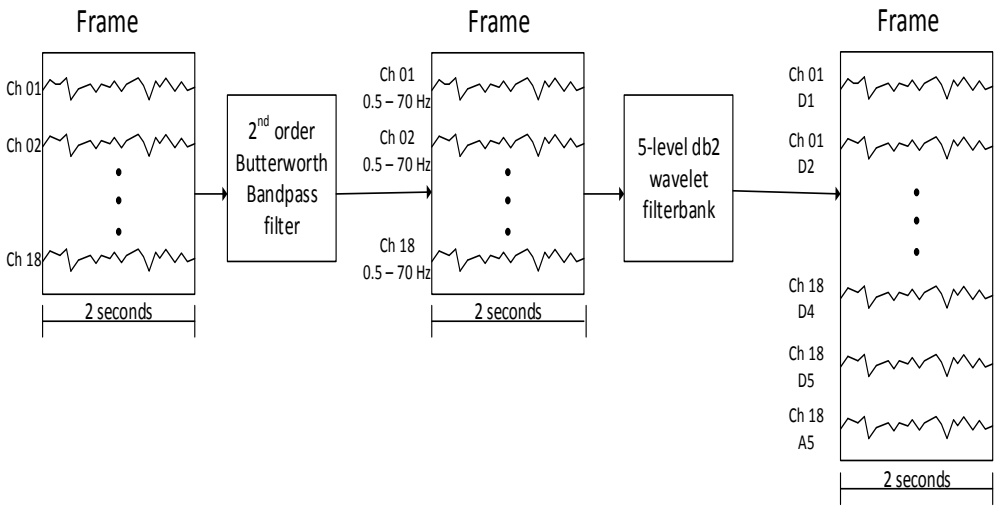


Figure 3.5: Filtering procedure

characterized by scales D3, D4, and D5. According to O'Donnell *et al.* (1974), EMG artefacts are not significant below 14 Hz. The EMG is characterized by scales D1 and D2. Scale D3 is omitted to avoid overlap between seizure and EMG characterization. Scale D4 is also used to represent the α-band activity in this study.

3.3.3 Feature processing

Only scales D3, D4, and D5 are required in the feature processing stage. The frame output from the filtering stage therefore has 18 channels with 3 scales each, forming 54 signals at 256 Hz each with 2 second duration. The 54 signal frame is input to the feature processing stage. For each of the 54 signals, three wavelet-scale-based features are computed:

- Relative Average Amplitude (RAA)
- Relative Scale Energy (RSE)
- Coefficient of Variation of Amplitude (CVA)

Relative Average Amplitude

For a given signal, the RAA is the ratio of the mean of the peak-to-peak amplitudes in the current epoch to the mean of the peak-to-peak amplitudes in the background epoch. The current epoch is in the current frame, and the background epoch is a 30 second block ending 1 minute earlier than the current epoch. Likewise it could be stated that the background epoch begins 45 frames before the current epoch, and ends 30 frames before the current epoch. Figure 3.6 clarifies this process by means of pseudocode.

Figure 3.6: Relative Average Amplitude

localMax = peaks($signal_k$) // for a signal in current frame (k)
localMin = peaks(-1 × $signal_k$)
for *each max-min pair* **do**
  amp_k(i) = localMax(i) - localMin(i) // peak-to-peak amplitude
end
$$RAA = \frac{mean(amp_k)}{mean\left(\ history_{[k-45]\ to\ [k-30]}\ \right)} \equiv \frac{mean(amp_k)}{mean\left(\ history_{[t-90]\ to\ [t-60]}\ \right)}$$
add $mean(amp_k)$ to end of history

The rationale is that the first minute of potential ictal EEG is compared to EEG that is highly likely inter-ictal. By moving the *background* window with the current frame ensures that the comparison is made to the recent state of the patient's EEG.

Relative Scale Energy

The RSE of the signal described by scale j in channel i is the ratio of the energy (E) of the given scale to the energy in all scales in the channel. Scales j = {D3,D4,D5} are used for channels $i = \{1 \rightarrow 18\}$. The simple formulae are given below for an n-sample discrete-time signal $x(n)$:

$$E_{ij} = \sum_{n=0}^{n} |x_{ij}(n)^2| \tag{3.2}$$

$$RSE_{ij} = \frac{E_{ij}}{\sum E_i} \tag{3.3}$$

where $\sum E_i$ sums energies of all scales (D1 → D5 and A5) in channel i.

Coefficient of Variation of Amplitude

The CVA of a signal uses the peak-to-peak amplitude (amp_k) used in the RAA heuristic as shown in Figure 3.6. The formula for CVA is given by:

$$CVA = \left(\frac{\sigma_{amp}}{\mu_{amp_k}}\right)^2 \tag{3.4}$$

where σ_{amp_k} is the standard deviation of the peak-to-peak amplitudes of the signal and μ_{amp_k} is the mean of the peak-to-peak amplitudes of the signal.

Feature vector

EEG signal amplitudes are influenced by the patient, the electrodes used, the data acquisition and amplification system, among others. The

features extracted are all relative to the patient and the time period. This serves to eliminate inter-patient and inter-system variability, which implies that a more representative training model can be created from the features.

The 54 signals (18 channels, 3 scales) in a frame each have 3 features extracted, which equates to 162 features. The benefit of feature extraction is that further processing is performed only with the 162 single-value features, instead of the frame of 54 signals of 512 (256 Hz, 2 seconds) samples each. For a given frame, a 162 parameter *feature vector* is therefore created. The transformation of frame to feature vector is illustrated in Figure 3.7.

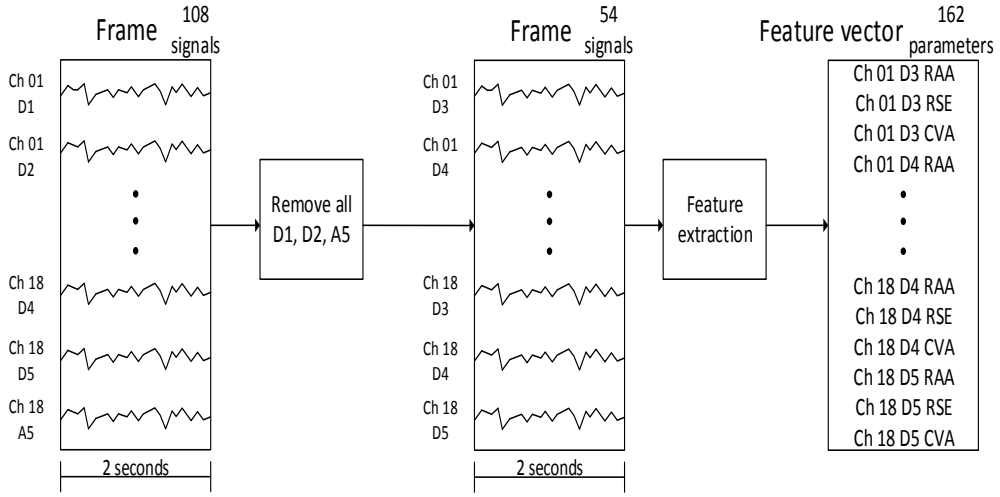


Figure 3.7: Feature vector

3.4 Naïve bayes classification and training

The method requires classification of a frame as either ictal or inter-ictal. The frame may also need to be labelled as high α or low α activity. The Naïve Bayes Classifier (NBC) can label a frame if the following Bayes' formulae are generated:

$$P(ictal|features) = \frac{P(features|ictal) \cdot P(ictal)}{P(features)} \tag{3.5}$$

$$P(high\ \alpha|features) = \frac{P(features|high\ \alpha) \cdot P(high\ \alpha)}{P(features)} \tag{3.6}$$

Equation 3.5 is used to determine whether the frame is ictal and Equation 3.6 is used to determine whether the frame contains high α activity. The posteriors $P(ictal|features)$ and $P(high\ \alpha|features)$ are referred to as the ictal-model

and high-α-model respectively in this work. Posteriors will collectively be referred to as 'models'.

The so called 'training phase' of the NBC involves determining the posteriors in Equations 3.5 and 3.6. Only data from the training-set is used in this phase. The training phase was given in Figure 3.2.

3.4.1 Ictal and inter-ictal training

Store ictal features

First the ictal likelihoods must be determined. Individually for non-'hold-out' case in the training set, the records containing ictal data is processed to store the feature-vector of every ictal frame. A maximum duration of 60 seconds of ictal data per seizure is used. This is done to avoid having longer seizures overwhelm shorter seizures in the likelihood model, and to avoid using ictal data as background data (Section 3.3.3). As per Figure 3.3, the EEG pool in this instance is records containing ictal data, and the procedure is continued until all ictal records have been processed. Keep in mind that the ictal feature-vectors of each case is stored separately.

Determine bin ranges

In this study, a new weighing procedure is introduced. Each case in the training set is assigned a weight. The case with the most training samples is assigned a weight of 1. This is referred to as the largest case. Then other samples have weights based on the inverse of their number of training samples, normalized to the largest case. For example, if the largest case has 200 samples, and case A has 153 samples, then case A will have a weight of $200/153 \approx 1.3072$. Each case is assigned a weight normalized to the largest case. The number of samples in a case multiplied by its weight shall be referred to as the number of weighted-samples in this study. The total number of weighted-samples is equal to the number of cases in the training set (*i.e.* 19) multiplied by the number of samples in the largest case.

The total number of weighted-samples is divided by 5, the output is the number of weighted-samples allowable in each of the 5 histogram bins. Now a histogram for each feature may be formed. Weights are divided evenly into 5 histogram bins. This can be done by first sorting all values for a given feature. Features are added to a bin from lowest to highest. When a value is assigned to a bin, the weight of that value is counted towards the bin count. As soon as the number of weighted-samples is equal to or exceed the number of weighted samples allowed per bin, then the next bin gets filled. When the next bin exceeds the number of weighted samples allowable, then the

next is filled, *etc.* Of course this means that not all bins will have the same number of actual samples, however the number of weighted samples will be approximately equal. The procedure effectively creates 162 histograms, with each of the 5 bins containing approximately equal weight. Histogram bin ranges should cover the entire spectrum $(0, \infty)$. The boundaries of higher and lower bins are selected by taking the mean of the highest value in the lower bin and the lowest value in the higher bin. Only the lower limit of the lowest bin and the higher limit of the highest bin are set to 0 and ∞, respectively.

Only by assigning weights to the samples, are the ranges made more representative of the entire training set instead of being biased towards the 'larger cases'. Figure 3.8 illustrates this procedure with a simplified example. In the example, only 2 cases are used. The figure only shows how ranges

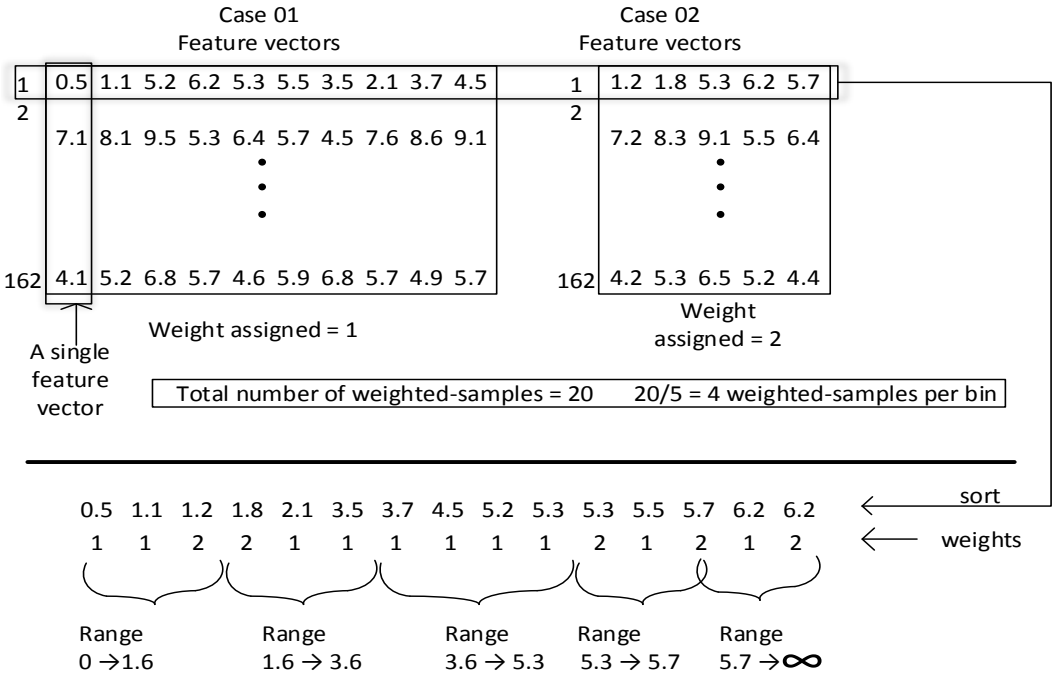


Figure 3.8: Finding ranges

for the first feature is determined. Case 01 has the most samples. Case 02 has a weight of $10/5 = 2$. The simplified example shows how bin ranges are determined for the first feature. This procedure is performed for every feature in the feature-vector. Note how the mean is taken between the highest value in the lower bin and the lowest value of the higher bin at each bin boundary.

Determine ictal likelihood model

The features (RAA, RSE, CVA) in each scale in each channel may now be used to determine a joint probability. Each feature has 5 bins into which

a signal may fall. For each scale in each channel there are $5^3 = 125$ unique combinations of the three feature bins, these unique combinations shall be called 'joint-bins' in this study. Since there are 18 channels and 3 scales, and 125 joint-bins for every scale in every channel, there are $125 \times 3 \times 18 = 6750$ joint-bins, of which $3 \times 18 = 54$ are true per frame.

The ictal data is reprocessed to determine the number of times that each of the 6750 joint-bins are true. Note that in the train step of the train-validation phase, data from the 'hold-out' patient is not reprocessed. In the train step of the testing phase, data from every case in the training set is reprocessed. The likelihood, $P(features|ictal)$, in Equation 3.5 is then the count in each joint bin divided by the number of frames (also called samples) stored for ictal training during reprocessing.

Inter-ictal feature storage and likelihood model

The bin ranges of the ictal class is used directly. The inter-ictal data is processed to generate a count for each of the 6750 joint bins. The counts are divided by the number of frames stored for inter-ictal training to determine $P(features|inter-ictal)$.

3.4.2 High α and low α selection

Rationale of α data selection

The overlap of frequency ranges for α-activity and seizures result in a large number of false alarms (Saab and Gotman, 2005). To prevent labelling of high-α data as ictal, the probability that a given frame contains high α data is implemented as given in Equation 3.6.

First the high-α and low-α data must be separated. This task is non-trivial. Saab and Gotman (2005) visually inspected their data corpus to extract 56 minutes of α activity that is free of artefacts from 6 of 28 patients, non-α activity data free of seizures was extracted from 13 of 28 patients. They used only features from scale D4. This process is tedious and labour intensive. Kuhlmann *et al.* (2009) aimed to address this by proposing a semi-automatic high-α and low-α selection heuristic. If the RSE in scale D4 was lower than some threshold in a subset of channels, then the data was stored as 'low-α' data. If the RSE in scale D4 was greater than some threshold in a subset of channels, then the data is stored for review. The stored data was visually inspected for artefacts. If no significant artefacts were found, then the data was held out for use in training. The threshold and number of channels considered was patient-specific for each patient in the training set.

In this study, a fully automated 'high-α and low-α selection heuristic' is proposed. If the heuristic is successful, it holds further promise in dynamic learning applications. As in Kuhlmann *et al.* (2009), the RSEs of scale D4 in each channel is considered. The EMG level is characterized by its Amplitude Ratio (amp_{EMG}). First the amplitudes of scales D1, D2, D4, and D5 must be determined. For example, consider scale D1: the sum over the absolute amplitudes of all samples in the D1 scale signal is obtained for channels 1 through 18. These are then summed together to obtain one value for scale D1 over all channels. The value is denoted EMG_{D1}. The EMG amplitude ratio is then

$$amp_{EMG} = \frac{EMG_{D1} + EMG_{D2}}{EMG_{D1} + EMG_{D2} + EMG_{D4} + EMG_{D5}} \tag{3.7}$$

Procedure for α data selection

First all data from every patient in the training set is processed and the $(x_{chn})^{th}$ highest D4 RSE as well as the amp_{EMG} for each frame is stored in memory. The x_{chn} th highest D4 RSE is denoted D4_{RSE}. High-α data without high EMG needs to be stored for training. For a given patient, consider only data with amp_{EMG} below the $(x_{EMG})^{th}$ percentile of all data **for that patient**. Within the low EMG data, only data where the D4_{RSE} is above the top $(x_{ALP})^{th}$ percentile of all data **for that patient** is considered. Within this remaining data, only data that are inter-ictal are used for the training phase (to avoid using ictal data in the α-classifier). In this study low-α data is stored when a frame has a D4_{RSE} below the $x_{NAL}{}^{th}$ percentile without regard for EMG level, or whether a seizure takes place. Note that percentiles x_{EMG}, x_{ALP}, and x_{NAL} are determined by excluding rejected frames.

The rationale for using the percentiles relative to the patient's own data is to avoid having to select a unique threshold for each patient, as was done by Kuhlmann *et al.* (2009). The proposed method is therefore generalized, as it finds optimal thresholds x_{EMG}, x_{ALP}, and x_{NAL} over all patients. All training set data is processed once in order to determine the D4_{RSE} values, and amp_{EMG} value that corresponds to the percentiles x_{EMG}, x_{ALP}, and x_{NAL}. Note that the percentiles are not-patient specific, but the values are. The values are needed such that data may be processed and stored. These values are used when processing data from a case, so that features for the 'low EMG, high α' and 'low α' classes may be stored. This is discussed next.

3.4.3 High α and low α training

The procedure outlined for generating likelihood models for ictal and inter-ictal classes is repeated for high-α and low-α classes. In this instance the posteriors $P(high\ \alpha|features)$ and $P(low\ \alpha|features)$ of Equation 3.6 must be generated

as the training step.

First, all data is processed sequentially. When a frame is 'low EMG, high α' and it is inter-ictal, then it is stored. Only features from scale D4 are stored such that the feature vector is a (18 channels, 1 scale, 3 features) 54 parameter feature vector. A histogram is made for each of the 54 parameters, each with 5 bins, in the same way as done with the ictal class. Again each frame is characterized jointly by the three features in the wavelet scale. Again there are 125 joint bins per scale, but since there is only one scale per channel, the total number of joint-bins are only 2250. Data is reprocessed so that a counter is made for each time a frame falls within one of the joint-bins. The likelihood $P(features|high\ \alpha)$ is determined by dividing the bin counts by the number of low-α frames stored for training.

Data is reprocessed again. When a frame is 'low α' then it is stored. Again only features from scale D4 are stored. The bin ranges determined in the high-α procedure are used to generate bin counts for the low-α class. The likelihood $P(features|low\ \alpha)$ is then determined by dividing the bin counts by the number of high-α frames stored for training.

3.4.4 Posterior model creation

Ictal model

Next the prior, $P(ictal)$, of Equation 3.5 may be determined by taking the ratio of the number of ictal frames over all frames stored. Similarly prior $P(inter-ictal)$ is found by taking the ratio of the number of inter-ictal frames over all frames. Finally the normalization constant, $P(features)$ is determined using:

$$P(features) = (features|ictal) \cdot P(ictal) + \tag{3.8}$$

$$P(features|inter-ictal) \cdot P(inter-ictal) \tag{3.9}$$

With all other terms determined, posterior $P(ictal|features)$ is generated. Note that in the train-validate phase a unique posterior is formed for each 'hold-out' case, since terms used to determine the posterior is unique for each case (since the 'hold-out' case data is not used). In the testing phase only one posterior model is formed since there is no 'hold-out'.

High-α model

The prior $P(high\ \alpha)$ is determined by taking the ratio of the number of high-α frames stored over all frames stored (high-α and low-α). Similarly prior $P(low\ \alpha)$ is determined by taking the ratio of the number of low-α frames

stored over all frames stored (high-α and low-α). Finally the normalization constant $P(features)$ used for Equation 3.6 is determined as:

$$P(features) = (features|high\ \alpha) \cdot P(high\ \alpha) + \tag{3.10}$$
$$P(features|low\ \alpha) \cdot P(low\ \alpha) \tag{3.11}$$

With all other terms determined, posterior $P(high\ \alpha|features)$ is generated. As for the ictal model, a unique posterior is generated for each 'hold-out' case in the train-validate phase, whereas only one posterior is created for the testing phase.

3.4.5 Classification

Classification occurs in the validation phase, as well as the testing phase. Once posterior models $P(ictal|features)$ and $P(high\ \alpha|features)$ have been obtained, classification of frames can be performed by the ictal- and the α-NBCs. Note that a rejected frame is automatically assigned inter-ictal (y = -1). For rejected frames the NBCs are not used, they is completely bypassed. Note further that a frame is only processed by the NBCs if the frame has enough non-rejected history. The last 6 minutes of frame history is stored. If the current frame is i, then it is only processed by the classifiers if there is at least 90 seconds of non-rejected frames in the 6 minutes before frame i. The feature RAA requires 90 seconds of history (Section 3.3.3). The 90 seconds need not be consecutive, however frames from more than 6 minutes before the current frame shall be considered as too far back in history. If the frame was not assigned a label of y=-1 for either of these reasons, then the frame is sent to the NBCs. The entire NBC classification procedure is illustrated by means of pseudo-code in Figure 3.9.

The frame is classified into a joint-bin based on the ictal bin ranges from the training-phase. The probability (for each scale in each channel) that the frame is ictal is taken directly from the posterior $P(ictal|features)$. The probability that a given scale j in a given channel i represents ictal activity is denoted $P_{ICT}(i,j)$. The probability that channel i represents ictal activity is given by $P_{ICT\_CHN}(i) = \sum_{j=3}^{5} P_{ICT}(i,j)$, since only scales 3, 4, and 5 are used for seizure characterization. The probability that the current frame is ictal, P_{SEZ_FRAME}, is obtained by summing the top x_{chn} channels with highest probability. The rationale is that limiting the number of channels that contribute to the frame probability to x_{chn} increases the chance for focal seizures to be detected.

The probability that a frame contains high-α is determined similarly. Recall that only scale 4 is used. The probability $P_{ALP}(i)$ that scale 4 in channel i represents high α activity is taken directly from the posterior

Figure 3.9: Classification heuristic

if *(frame was rejected) OR (< 45 frames in previous 6 minutes)* **then**
    y = -1
else
    frame is sent to the ictal-NBC
    Determine ictal joint-bin of all 162 signals in frame
    for *each channel i and scale j* **do**
        Obtain $P_{ICT}(i,j)$ directly from $P(ictal|features)$
    end
    for *each channel i* **do**
        $P_{ICT\_CHN}(i) = \sum_{j=3}^{5} P_{ICT}(i,j)$
    end
    sort(P_{ICT_CHN}) // in descending order
    $P_{SEZ\_FRAME} = \sum_{i=1}^{x_{chn}} P_{ICT\_CHN}(i)$ // top x_{chn} summed
    $P_{ICT\_scaled} = P_{ICT\_FRAME} \cdot (1 - amp_{EMG} \cdot x_{NTH})$
    $P_{ICT} = \sum_{n=i}^{i+1-x_{es}} P_{ICT\_scaled}$ // sum last x_{es} samples
    if $P_{ICT} < I_{TH}$ **then**
        y = -1
    else
        Determine high-α joint-bin of all 18 D4 signals
        for *each channel i in scale D4* **do**
            Obtain $P_{ALP}(i)$ directly from $P(high\ \alpha|features)$
            $P_{ALP\_CHN}(i) = P_{ALP}(i)$
        end
        sort(P_{ALP_CHN}) // in descending order
        $P_{ALP\_FRAME} = \sum_{i=1}^{x_{chn}} P_{ALP\_CHN}(i)$ // top x_{chn} summed
        $P_{ALP} = \sum_{n=i}^{i+1-x_{es}} P_{ALP\_FRAME}$ // sum last x_{es} samples
        if *If* $P_{ALP} < A_{TH}$ **then**
            y = 1
        else
            if $P_{ALP\_CHN}(1 \rightarrow x_{chn})$ *all from same hemisphere* **then**
                y = 1
            else
                y = -1
            end
        end
    end
end

$P(high\ \alpha|features)$. The probability that a given channel i represents high α activity is then simply $P_{ALP\_CHN}(i) = P_{ALP}(i)$. The probability that the current frame represents high α activity, P_{ALP_FRAME}, is obtained by summing the top x_{chn} channels with highest probability.

Since EMG activity often causes false positives (Saab and Gotman, 2005), the ictal probability for the frame is scaled by its EMG activity: $P_{ICT\_scaled} = P_{ICT\_FRAME} \cdot (1 - amp_{EMG} \cdot x_{NTH})$ where x_{NTH} is a scaling factor on amp_{EMG}. Finally, temporal context is introduced by summing P_{ICT_scaled} over the last x_{es} frames to determine the final detection variable P_{ICT}. The final detection variable for high α activity is obtained by summing P_{ALP_FRAME} over the last x_{es} epochs and denoted P_{ALP}. Finally, the final variable P_{ICT} is compared directly to ictal threshold I_{TH} and final variable P_{ALP} is compared directly to high α threshold A_{TH}.

When a frame is passed to the NBC, a heuristic is performed. The P_{ICT} at the current frame is determined. If $P_{ICT} < I_{TH}$ then the frame is immediately labelled as inter-ictal. The classifier output is $y = -1$. If however $P_{ICT} \geq I_{TH}$ then the frame is sent to the α-classifier. If $P_{ALP} < A_{TH}$ then the frame is labelled ictal, with classifier output $y = 1$. If $P_{ALP} \geq A_{TH}$ then the frame is labelled inter-ictal, unless all x_{chn} channels used to determine the current P_{ALP_FRAME} is within the same hemisphere, in which case the frame is labelled ictal. In the convention of this study, if the channels used to determine P_{ALP_FRAME} includes channels of only one hemisphere along with Fz, Cz, or Pz, then the channels are considered to be from the same hemisphere. The rationale here is that a seizure-like epoch is encountered, but there is α activity generalized over the brain, then the epoch is likely α activity that merely resembles seizure activity. This is done since α activity is more likely to be dual-hemispheric than seizures are (Saab and Gotman, 2005).

3.5 Postprocessing

3.5.1 Sequential hypothesis test

Despite every effort made, some instances of misclassification of data is bound to occur. Some ictal frames will be labelled inter-ictal, and vice-versa. In a system meant for monitoring a patient in an attempt to detect seizures and sound an alarm when it does, misclassification is always serious. A high False Negative count would render the monitor worthless for seizure detection, and a high False Positive count would cause a lot of work for hospital staff who need to check up on the patient after every alarm.

A new process is introduced whereby the the software architecture

does not declare an ictal event from just one frame label output, y, by the NBC. A higher form of data windowing is introduced, whereby the software architecture determines the probability that a seizure has taken place based off the last N frame labels. Parameter N is the number of frames used for probability estimation.

Number of frames

In order to limit misclassification, a simple *sequential hypothesis test* can be used (Gardner, 2004). The test is used to estimate the probability that a given window of N frames is an ictal event. For N sequential output frames labelled $y \in \{-1, +1\}$:

$$N = N_+ + N_- = \sum_{i=1}^{N} y[i] \tag{3.12}$$

Intuitively, N_+ is the number of frames labelled +1, and N_- is the number frames labelled −1. Again the window is advanced in time, this time the window overlap is $N - 1$ frames. This means when a new frame is labelled, the last N frames (including the current frame) is evaluated.

Probability threshold

The probability estimate, $\hat{p}$, that an ictal event occurred within the last N frames is given by the following equations:

$$\hat{p} = \frac{N_+}{N} \tag{3.13}$$

The decision function, $z[i]$, that declares a window of size N as *ictal* or *inter-ictal* is given by:

$$z[i] = sgn\left(\hat{p} - p\right) \tag{3.14}$$

where $z[i] = -1$ (inter-ictal) and $z[i] = 0$ or $+1$ (ictal). Parameter p is the detection threshold probability. Therefore, only when $\hat{p} \geq p$ will the frame be declared ictal. The window length N and the threshold p are parameters that need to be optimized.

The output label $z[i]$ is the label assigned strictly to the *last* (also called current i^{th}) frame in the size N window. This means that for metric t_{dl} calculation, the seizure is only considered detected on the last frame. An observation that can be made from this is that the detection latency t_{dl} will degrade if the window length N is large.

3.5.2 Persistence refractory period

There are instances where the seizure may be detected shortly before its annotated onset. To avoid labelling such early detections as False Positives,

detection persistence by means of a *refractory parameter*, T, is introduced. Parameter T specifies an interval during which the detector, if triggered, maintains its state and ignores subsequent triggers. This means that, after an ictal period is detected and the detector output set to $z[i] = +1$, the detector outputs remain $z[i] = +1$ for T seconds, regardless of incoming data. If the T second block of frames classified as ictal overlap with an actual ictal event, then only is the ictal event considered detected. The detection is made at the onset of the detection block.

Importantly, persistence will also diminish the amount of FP declarations, since a burst of FPs may be declared within a short time period. The persistence block can contain these FPs within one detection block and therefore label it as one FP. This 'trick' should be considered acceptable, since a single FP will notify ICU staff, but FPs that are immediately after will not cause additional disturbance to staff. As a trade-off, a large T will diminish the TN duration.

Parameter T has a real-world influence besides only the classifier performance. Consider the event where a detection is made. The ICU staff is alerted, and a nurse tends to the patient being monitored. For a very large T, subsequent detections would not be made, so the staff would not be alerted again for this duration. So then for the duration of T, the nurse should attend to the patient. Evidently a large T is problematic, however the diminishment of TN due to large T may not even register an effect compared to the advantages of a large T in the train-validate phase objective function. Some maximum threshold for T should be set. Gardner (2004) reports this same problem. In his work he obtains an increasingly improved objective function with higher T. He selected $T = 180$ seconds to be the maximum allowable persistence period. His choice of maximum T is applied in this work for practical considerations.

3.6 Optimization

3.6.1 Performance metrics and objective function

In this study the metrics that constitute the confusion matrix are reported. In this way comparison of performance between this work and the work of future researchers are facilitated, since any performance metric can easily be constructed with only the primary confusion matrix metrics (count and duration) TP, TN, FP, and FN, and the detection latency t_{dl}.

The objective function (OF) used in the train-validate phase shall be the F_1 score. The F_1 score is the harmonic mean of precision (PPV) and recall (TPR), given by Equation 2.7. To obtain PPV and TPR of Equation

2.7, the arithmetic mean (μ) over all cases is taken. The procedure is therefore as follows: the PPV and TPR for each patient is generated, the $\mu(TPR)$ and $\mu(PPV)$ over all patients are determined, the F1 score is determined as in Equation 2.7 using $\mu(TPR)$ and $\mu(PPV)$:

$$F_1 = 2 \cdot \frac{1}{\frac{1}{\mu(PPV)} + \frac{1}{\mu(TPR)}} = 2 \cdot \frac{\mu(PPV) \cdot \mu(TPR)}{\mu(PPV) + \mu(TPR)}$$

3.6.2 Parameter vectors

The parameters described so far that require optimization are listed:

- $\{x_{high}$, x_{mains}, and $x_{phase}\}$ from Section 3.3.1.
- $\{x_{EMG}$, x_{ALP}, and $x_{NAL}\}$ from Section 3.4.2.
- $\{x_{chn}\}$ from Sections 3.4.2 and 3.4.5.
- $\{I_{TH}$, A_{TH}, x_{NTH}, and $x_{es}\}$ from Section 3.4.5.
- $\{N$, p, and $T\}$ from Section 3.5.

The parameters listed all need to be optimized in this study.

For their corpus, Saab and Gotman (2005) found optimal parameters $\{x_{mains}, x_{high}, x_{phase}\} = \{20\,\mu V, 1000\,\mu V, 2\}$. Kuhlmann *et al.* (2009) determined optimal parameters: $\{x_{mains}, x_{high}, x_{phase}\} = \{300\,\mu V, 1000\,\mu V, 2\}$ for their data corpus. What this implies is that there is no universal optimum for these parameters, even when using the same methodology. These parameters are unfortunately unique for each recording equipment set, since the characteristics of the recording signal is unique for each set.

Parameters x_{mains}, x_{high}, and x_{phase} cannot be selected arbitrarily for the CHB-MIT corpus. These parameters must be optimized in the training-validation phases to ensure that they maximally reject artefacts, without rejecting clean data. The parameters x_{mains}, x_{high}, and x_{phase} will be selected over the entire database, since it is likely that the entire CHB-MIT corpus has been collected using the same recording equipment (Guttag, 2017). These parameters affect the seizure and α classifiers. In other words they affect the posteriors in Equations 3.5 and 3.6 in Section 3.4.

Parameters x_{EMG}, x_{ALP}, and x_{NAL} are novel, and are intended for the implementation of the automatic α data selection heuristic. These parameters affect the model created for the α-classifier. In other words these parameters affect the posterior in Equation 3.6. All remaining parameters $\{x_{chn}$, P_{TH},

A_{TH}, x_{NTH}, x_{es}, N, p, $T\}$ do not affect the model, however they do affect the classifier.

The parameter vector that is optimized within a single train-validate iteration is denoted $\Pi_p = \{I_{TH}, A_{TH}, x_{NTH}, x_{chn}, N, p, x_{es}, T\}$. The hyper-parameter vector in this study is denoted $\Pi_h = \{x_{mains}, x_{high}, x_{phase}, x_{EMG}, x_{ALP}, x_{NAL}\}$. A further distinction within Π_h is made. $\Pi_{h(r)} = \{x_{mains}, x_{high}, x_{phase}\}$ are the rejection hyper-parameters. $\Pi_{h(m)} = \{x_{EMG}, x_{ALP}, x_{NAL}\}$ are the model hyper-parameters. The distinction is made for a number of reasons. Firstly, $\Pi_{h(r)}$ will influence the model created using $\Pi_{h(m)}$. Secondly, $\Pi_{h(r)}$ is also used in the validate and testing phase in the frame rejection procedure, whereas $\Pi_{h(m)}$ influences **only** the (α) model creation step. The set $\Pi_{h(r)}$ can be thought of as some sort of hyper-hyper-parameters, as is evident from Figure 3.1.

3.6.3 Optimization schedule

As shown in Figure 3.1, parameters Π_p are optimized in a loop under $\Pi_{h(m)}$, which are optimized in a loop under $\Pi_{h(r)}$. A simple grid-search procedure is used to optimize each.

The model-creation procedure makes use of Π_h, and occurs in the training phase. For each new model that must be created, the entire training set needs to be processed multiple times. Creating multiple models simultaneously is possible, however, a separate feature-vector storage would be required for each model, which in turn significantly slows computational performance. The validation/testing phase is, compared to the training phase, time inexpensive. After model creation, the train/test set is processed once, however, multiple classifiers can be evaluated simultaneously, since only one feature storage is necessary. This means that only one variable for mean of the peak-to-peak amplitudes (amp) of the last 45 frames (Section 3.3.3), and only one variable for the rejection history of the last 180 frames (Section 3.4.5) are required, regardless of the number of classifiers. The only storage variables that increase with the number of classifiers are the sequential hypothesis window of size N (Section 3.5.1) and number of sets of performance metrics. For this reason, although the grid of 8 parameters where the number of values for each parameter is given by $\{n_1, n_2, ..., n_8\}$ will yield $\Pi_{i=1}^{8} n_i$ combinations, it can still be evaluated simultaneously with acceptable speed. Here Π refers to the product operator.

3.6.4 Cross-validation

Cross-validation is useful to limit overfitting, and to provide a more accurate metric to evaluate how well a classifier will generalize to an independent

(unseen) dataset. In short, cross-validation is used for more accurate evaluation of classifier performance.

It is standard practice in the fields of machine learning and data mining to perform *k-fold cross-validation*. First, data is partitioned into k equal sized subsamples. The train-validation phases are each repeated k times. In each iteration, of the k subsamples, only one subsample is kept for testing, and the other $k-1$ are used for training. The subsample kept for testing is termed the *hold-out* sample/patient/case. This means that each subsample will be used only once for validation. The number of results will therefore be k, and these results can be averaged or consolidated in some other way. Additionally, when k is equal to the sample size, the technique is specifically referred to as *leave-one-out cross-validation* (LOOCV).

In this study a special method of performing cross-validation is used. A 19-fold LOOCV method is used with each patient in the training-set being a single subsample. The purpose of the train-validate phase is to optimize parameters for performance on independent data, for this reason data from the validation case should not be used in the training phase. Furthermore the statistics presented for optimization are global statistics taken as the (arithmetic) mean over individual patients (see Section 3.6.1) rather than per seizure. This is done to avoid optimization bias toward training cases with many seizures. In effect the train-validate phase becomes non patient-specific, so the performance metrics in the validation phase is a more accurate representation of what the metrics in the testing phase will be, using the same parameters.

3.6.5 Resources used

The software for this research was coded in MATLAB® by Mathworks®, and tested successfully using the R2014a through R2017a releases. The University of Stellenbosch hosts a High Performance Computing (HPC) *cluster*, designated 'HPC1' (Universiteit Stellenbosch, 2017). Computational processing was performed on HPC1.

In order to diminish processing time, the HPC1 was used for parallel computing. In the training and validation phases, the 19 cases training-set is used. One node on the cluster is assigned for each case, and an additional node is assigned as the remote host for administration. A total of 20 nodes are therefore used in the training and validation phases. Similarly, in the testing phase, 6 nodes on the cluster are used for the 5 cases.

Chapter 4

Results

4.1 Offline evaluation

The offline evaluation refers to the train-validate phase iterations. In this section the offline evaluation is detailed.

The selections of $\Pi_{h(r)}$, $\Pi_{h(m)}$, and Π_p are given in Appendix B. Table B.4 gives the final confusion matrix and detection latency of the optimized algorithm. Table 4.1 presents some of the most common performance metrics for discussion, along with the mean seizure duration (D) of each patient.

Table 4.1: Offline performance metrics

Case	**D [sec]**	TPR	PPV	FPR
01	63.1426	0.5714	1	0
02	57.3333	0.6667	1	0
03	57.4285	0.5714	1	0
04	94.5	0.75	0.375	0.0321
06	15.3	0.1	0.0588	0.24
08	132.6	0.8	0.8	0.0505
10	63.8571	0.5714	1	0
11	98	1	0.4286	0.1155
12	36.6296	0.4074	0.6875	0.2636
13	44.5833	0.0833	0.1667	0.1519
14	21.1250	0	0	0.0769
15	99.6	0.5	0.7692	0.0764
17	97.6667	1	0.0698	1.9179
18	52.8333	0.6667	0.3333	0.2256
19	78.6667	0.6667	0.0909	0.6703
20	36.75	0.8750	1	0
21	49.75	0.5	1	0
22	68	1	0.6	0.0648
23	60.5714	0.4286	1	0

The optimal parameter sets $\Pi_{h(r)}$, $\Pi_{h(m)}$, and Π_p as determined in the offline phase are given below:

- $\Pi_{h(r)} = \{\text{NA,NA}, 1.6\}$
- $\Pi_{h(m)} = \{\text{NA, NA, NA}\}$
- $\Pi_p = \{1.45,\text{NA}, 1.1, 3, 15, 0.35, 2, 180\}$

The offline mean values of the common performance metrics of Tables B.4 and 4.1 are given for convenience:

- $TPR = 58.73$ %
- $PPV = 59.89$ %
- $FPR = 0.2045$ /h
- $t_{dl} = 23.1313$ sec

4.2 Online evaluation

In the online evaluation, the previously unused testing-set is evaluated. No further parameter selection or optimization is allowed in this phase. Each case is processed only once and performance metrics are generated.

The confusion matrix and detection latency for the online evaluation is given in Table B.6. Table 4.2 presents some of the most common performance metrics for discussion, with the mean seizure duration (D) of each patient. The online mean values of some common performance metrics of Tables B.6

Table 4.2: Online performance metrics

Case	**D [sec]**	TPR	PPV	FPR
05	111.6	0.8	0.3636	0.1804
07	108.3333	1	0.3333	0.0897
09	69	1	0.6667	0.0296
16	8.4	0	0	1.4211
24	40.9375	0.1250	0.6667	0.0471

and 4.2 are given for convenience:

- $TPR = 58.50$ %
- $PPV = 40.61$ %
- $FPR = 0.3536$ /h
- $t_{dl} = 29.4167$ sec

Chapter 5

Discussion

5.1 Evaluating contributions

The intended contributions listed in Section 1.6 are discussed in this section.

5.1.1 Automated α heuristic

The offline evaluation in Section 4.1 has indicated that the α-classifier yields no significant improvement. The optimal offline algorithm has no α detection whatsoever. It appears that the automatic α-training data selection heuristic lacks efficacy, since a review of ictal and α probabilities showed that high-α probabilities almost never coincided with high ictal probabilities from inter-ictal EEG. For this reason parameter set $\Pi_{h(m)}$ is NA (not applicable), as shown in Section 4.1.

5.1.2 Improving on the state-of-the-art

As per Section 2.3.5, the Saab and Gotman (2005) method is considered the state-of-the-art. They report performance of $TPR = 77.9$ %, and $FPR = 0.86$ /h. The algorithm presented in this study is based on their work. Previously unoptimized parameters (x_{phase}, x_{NTH}, x_{es}) are however optimized, and additional procedures are introduced. The final performance metrics reported in the online evaluation is $TPR = 58.5$ %, $PPV = 40.61$ %, and $FPR = 0.3536/h$. The method of this study captures 19.4 % less seizures, however, the FPR is reduced by more than half. The significance of this is made evident when comparing the FPR per day. The Saab and Gotman (2005) method registers $FPR = 20.64$ /day, whereas the method in this study registers $FPR = 8.4864$ /day. Although neither is likely to be used clinically, the algorithm presented here arguably better approaches clinical applicability. For this reason it is maintained that the algorithm presented in this study is an improvement to the Saab and Gotman (2005) method.

In Section 3.5.2 the rationale for setting some maximal threshold to avoid converging to a very large value for T was discussed. From Table B.3 it is evident that this threshold was necessary also in this work. More accurate algorithms should enable normal optimization of parameter T as a trade-off between FPs and TN. In algorithms such as the one presented in this work, a

higher T will almost inevitably lead to better results due to the high number of FPs that it covers.

The sequential hypothesis test that was introduced to the Saab and Gotman (2005) method in this study is utilized in the final classifier. What is however concerning is that patients with a short mean seizure duration (D) in the offline (Table 4.1) and online (Table 4.2) evaluations did not show good sensitivity (TPR). This is likely due to the large value number of frames (N=15) selected in the offline phase. The sequential hypothesis test therefore likely plays a large role in the inter-patient variability in the offline and online evaluations.

Besides the unexpected exclusion of the α-classifier, the optimal $\Pi_{h(r)}$ were also highly surprising. Section 4.1 indicates that no mains noise-(x_{mains}) or abnormally high channel amplitude (x_{high}) rejection is necessary for algorithm improvement. Parameter set $\Pi_{h(r)}$ therefore has two NA (not applicable) parameters, as shown in Section 4.1. This is a positive outcome, since rejecting a high portion of data would make the algorithm less implementable clinically. The ICU staff would be notified every time a continuous period of frame rejection takes place, so that the source of contamination could be mitigated. An algorithm that operates with minimal frame rejection demonstrates good robustness on noisy data. Clinical data may be quite noisy, and so the low levels of required data rejection is therefore a welcome result.

It is noteworthy that one parameter has not been optimized in this study, namely the number of histogram bins. If n histogram bins are used per feature, then n^2 joint-bins are created. Some further optimization could be performed, however the number of bins will correlate with the size of the training set of the data corpus. With more training data, the histograms can be made more precise, since there is enough samples for every joint-bin. To avoid such a corpus-specific parameter, a different method of creating probability distributions may be considered. A probability density function (PDF) is a continuous function used to represent probability. Constructing a PDF from a discrete set of points is non-trivial. To encode the probabilities associated with joint-bins, a 4-dimensional PDF is created, where 3 dimensions encode the 3 feature values, and 1 dimension encode the likelihood.

5.1.3 Use of a publicly available data corpus

Comparing the proposed method to any other method in literature proposes a problem. Unless the publication in question has used the CHB-MIT data corpus, a comparison of results to compare algorithm performance it not entirely appropriate. It was discussed in Section 2.3.3 that use of private data

corpora hinders algorithm comparison. The fact that Saab and Gotman (2005) have more data channels to analyse and that they have only adult patients (whereas the CHB-MIT corpus contains patients with age range from 1.5 to 22 years) may also benefit their final performance. This is substantiated by the fact that younger children (under the age of 5 years) have markedly different EEG to older children and adults (Page *et al.*, 2015). Finally, a comparison of detection latency t_{dl} is also meaningless in this context, since the seizures in the CHB-MIT corpus annotates seizure start and end times, whereas on the Saab and Gotman (2005) data corpus seizures are annotated to start when the EEG displays a clear seizure discharge, without returning to background. To be able to properly compare algorithms, the same data corpus should be evaluated. For this reason it is maintained that implementing this method on a public data corpus is a contribution to the field of seizure detection. Reporting the full confusion matrix in terms of counts as well as duration is also considered beneficial for future comparison.

Of course performing the online evaluation on only 5 cases is insufficient for making statements with strong statistical significance. It is recommended that more patients be sourced from other databases to evaluate the algorithm more thoroughly. The Temple University Hospital (TUH) data corpus (Section 2.3.3, Table 2.2) is a good candidate.

5.2 Achievement of the aims and objectives

5.2.1 Achievement of the aims

The aim of this study, as given in Section 1.4, was to develop a robust, completely automatic software solution intended for real-time whole-brain seizure monitoring that uses EEG data, and no patient- or seizure-specific tuning. The training and testing was to be performed using a large, publicly available data corpus. The current state-of-the-art is improved upon. The final deliverable of this research was the online performance data of the algorithm and a discussion that addresses each limitation in Section 1.2.

The solution proposed is not highly **robust** in terms of inter-patient variability of performance. This is evident from Table 4.2. The algorithm presented is however remarkably **robust** despite noisy data (see Section 5.1.2), requiring very little data rejection to achieve its peak performance. The method presented in this study does not require any manual data inspection or manipulation. It is therefore fully **automatic**.

Every effort was made during the programming of the software to ensure that the processing requirement is minimal. The processing requirement

is benchmarked using an Intel® Core™ i7 3.40 GHz processor. Using this machine, the speed at which software is processed is evaluated 5 times. Over 5 repetitions, it was determined that 3600 seconds of 256 Hz EEG data (all 18 channels) is processed in 167.3523 seconds (mean). Equivalently, 21.5115 seconds of EEG data is processed per second. The software speed benchmark implies that the method proposed could be run on computers of lower processing capacity if required. This is promising for **real-time** clinical application using dedicated devices.

The CHB-MIT corpus contains seizures of many locations and types (see Table 3.3). Seizures of all types and locations have been detected in the offline and online phases. This suggests that the procedure is capable of **whole-brain monitoring**. The methodology of this study ensures that the solution is **generalized** (neither patient-specific, nor seizure-specific). The training- and testing data were strictly separated, and no training or optimization was done using the test-set. Seizures of various types and locations are mixed indiscriminately into the training- and testing sets. Finally, the data corpus used for all phases in this study is the **publicly available** CHB-MIT corpus. This means that the results obtained in this study can be reproduced exactly. Future solutions may be compared to this solution, by evaluating using the same data corpus.

The proposed algorithm is discussed with regards to the limitations of Section 1.2:

1. The LOOCV procedure was used to minimize inter-patient variability of performance. From Table 4.2 it can be seen that despite this, there is still significant inter-patient variability. Inter-patient performance variability is an inherit problem for a number of patient-monitoring applications, including seizure analysis.

2. In this study the detection latency, along with the confusion matrix (both counts and durations) for the final offline (Table B.4) as well as online (Table B.6) evaluations are reported. This enables future researchers to easily compare any performance metric to this work.

3. No patient- or seizure-specific tuning is required for this procedure.

4. The technique was evaluated on long-term continuous (independent) EEG data in the online phase.

5. The data used is representative of clinical data.

6. Training and testing were done using the large CHB-MIT public data corpus.

7. Multiple channels (without prior focus channel knowledge) is used in training and testing. Multiple seizure types and locations are evaluated.

8. No manual removal of data intervals of ocular- or muscle artefacts or any other data pre-selection is required.

5.2.2 Achievement of the objectives

In this study, EEG-based seizure detection software was developed based off the current state-of-the-art Saab and Gotman (2005) method. The method was improved by automating the procedure and achieving improved performance through further optimization and introduction of additional heuristics. The system was trained and optimized offline, and then tested online with independent data. The final online performance metrics was reported in Table 4.2. Evidently the objectives of the study (Section 1.4) have been completed.

5.3 Recommendations for future research

The probability density function (PDF) replacement of the histogram method used in this study warrants further investigation. Furthermore, it has been mentioned that more patients need to be evaluated online for stronger statistical significance of the final performance metrics reported. These recommendations were discussed in Sections 5.1.2 and 5.1.3. More training data may also be valuable, as it may significantly contribute to improved model creation.

In Appendix C, Section C.1, the probability output history of variable P_{ICT} is provided for each case during the online evaluation. It is shown that selecting any global threshold I_{TH} is performs rather poorly in separating ictal from inter-ictal probabilities over multiple patients. Instead some dynamic threshold could be applied that changes with patient-data. This could potentially improve the algorithm performance.

A further method for improving both inter-patient variability, as well as overall algorithm performance, is to introduce some form of dynamic learning. With dynamic learning, data being monitored may be added to the classifier model in real-time such that the model is primed for improved performance on the patient being monitored. In Appendix C, Section C.2, a simplistic, unoptimized dynamic learning method is implemented. The technique (once further sophisticated and optimized) shows significant promise for improving the algorithm performance.

Ideally a forewarning of seizure onset should be given, that is, seizures

should ideally be predicted. In hospital settings, for the administration of fast-acting AEDs (commonly called Rescue Medication or Rescue treatment) the medication may be injected directly into the patient's bloodstream. A 2-minute prediction time is required (Schelter *et al.*, 2006) for fast-acting AEDs to take effect. Such a prediction time corresponds to a detection latency $t_{dl} = -120$ [sec] or equivalently a prediction horizon $t_{ph} = 120$ [sec]. Methods where the mean $t_{dl} < 0$ or the mean $t_{ph} > 0$ may be referred to as prediction-oriented methods. There are unfortunately no seizure prediction methods that may be considered reliable for application to clinical EEG as defined in this study. The development of a seizure prediction algorithm with $t_{ph} > 120$ is therefore needed.

The prediction classifier is desirable, since seizures should ideally be predicted. The detection classifier is valuable, since seizure prediction algorithms may not be able to predict or detect seizures with sudden onsets (as explained in Section 1.1). This novel combination of the detection and prediction classifiers may result in a superior new algorithm. An evaluation on the improvement of performance when implementing a seizure detector with a seizure predictor, as opposed to using these independently may be valuable.

Since the variation between ictal and inter-ictal EEG is more pronounced than the variation between pre-ictal and inter-ictal EEG, the seizure prediction classifier will likely benefit greatly from using patient-specific training data. To achieve this without patient-specific tuning, implies that dynamic learning for the predictor may need to be implemented.

Chapter 6

Conclusion

In this study a software solution intended for robust, completely automatic, real-time, whole brain EEG seizure monitoring with no patient- or seizure-specific monitoring. The aims listed here were addressed, however the algorithm could not perform robustly between patients, showing high inter-patient variability in performance. The algorithm was trained and evaluated on a publicly available data corpus, and the online performance (Table 4.2) and full confusion matrix (Table B.6) are reported. This facilitates future replication and comparison to this work.

The algorithm was trained offline using 19 cases (147 seizures, 765.6 hours of data) and tested online on an independent test set of 5 cases (38 seizures, 214.3 hours of data). The database contains patients with ages between 1.5 to 22 years of age. In the online evaluation a sensitivity (TPR) of 58.5 %, a selectivity (PPV) of 40.61 %, and a false positive rate (FPR) of 0.3536 /h was achieved. The results obtained here are more likely to be implementable clinically than the current state-of-the-art due to the lower FPR (See Section 5.1.2). For this reason it is maintained that this solution represents an improvement to the current state-of-the-art.

This solution, similar to those before it, cannot yet overcome all impediments to clinical applicability. In particular, the inter-patient variability and even the overall classifier performance is still not satisfactory. The requirements of a clinically applicable universal seizure detection monitor are not easy to solve, however promising solutions for future research are provided in Section 5.3.

Bibliography

Aarabi, A., Fazel-Rezai, R. and Aghakhani, Y. (2009 sep). A fuzzy rule-based system for epileptic seizure detection in intracranial EEG. *Clinical Neurophysiology*, vol. 120, no. 9, pp. 1648–1657. ISSN 13882457.
Available at: http://www.ncbi.nlm.nih.gov/pubmed/19632891http://linkinghub.elsevier.com/retrieve/pii/S1388245709004246

Abend, N.S., Mani, R., Tschuda, T.N., Chang, T., Topjian, A.a., Donnelly, M., LaFalce, D., Krauss, M.C., Schmitt, S.E. and Levine, J.M. (2011). EEG monitoring during therapeutic hypothermia in neonates, children, and adults. *American journal of electroneurodiagnostic technology*, vol. 51, no. 3, pp. 141–64. ISSN 1086-508X.

Acır, N. and Güzeliş, C. (2004). Automatic spike detection in EEG by a two-stage procedure based on support vector machines. *Computers in Biology and Medicine*, vol. 34, pp. 561–575.
Available at: http://www.intl.elsevierhealth.com/journals/cobm

American Cancer Society (2017). What Are the Key Statistics About Brain and Spinal Cord Tumors?
Available at: https://www.cancer.org/cancer/brain-spinal-cord-tumors-adults/about/key-statistics.html

American Heart Association and American Stroke Association (2012). What You Should Know About Cerebral Aneurysms.
Available at: http://www.strokeassociation.org/STROKEORG/AboutStroke/TypesofStroke/HemorrhagicBleeds/What-You-Should-Know-About-Cerebral-Aneurysms_UCM_310103_Article.jsp#.WLNdfNpNzIU

Andrzejak, R.G., Lehnertz, K., Mormann, F., Rieke, C., David, P. and Elger, C.E. (2001 nov). Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state. *Physical Review E*, vol. 64, no. 6, p. 061907. ISSN 1063-651X.
Available at: http://www.ncbi.nlm.nih.gov/pubmed/11736210https://link.aps.org/doi/10.1103/PhysRevE.64.061907

Aurlien, H., Gjerde, I.O., Aarseth, J.H., Eldøen, G., Karlsen, B., Skeidsvoll, H. and Gilhus, N.E. (2004). EEG background activity described by a large computerized database. *Clinical Neurophysiology*, vol. 115, no. 3, pp. 665–673. ISSN 13882457.

Barbosa, L.V. (2013). Fourier transform time and frequency domains.

Bergen, D.C. (2006). Do seizures harm the brain? *Epilepsy currents*, vol. 6, no. 4, pp. 117–8. ISSN 1535-7597.
Available at: http://www.ncbi.nlm.nih.gov/pubmed/17260030http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC1783429

Brisman, J.L., Song, J.K. and Newell, D.W. (2006). Cerebral Aneurysms. vol. 9, no. 31.

Bronen, R. (2000). The Status of Status: Seizures Are Bad for Your Brain's Health. *American Journal of Neuroradiology*, vol. 21, pp. 1782–1783.

Bruck, I., Antoniuk, S.A., Spessatto, A., Bem, R.S., Hausberger, R. and Pacheco, C.G. (2001 mar). Epilepsy in children with cerebral palsy. *Arquivos de neuro-psiquiatria*, vol. 59, no. 1, pp. 35–9. ISSN 0004-282X.
Available at: http://www.ncbi.nlm.nih.gov/pubmed/11299428

Busl, K.M. and Greer, D.M. (2010). Hypoxic-ischemic brain injury: Pathophysiology, neuropathology and mechanisms. *NeuroRehabilitation*, vol. 26, no. 1, pp. 5–13.

Chang, C.-C. and Lin, C.-J. (2011). LIBSVM: A Library for Support Vector Machines. In: *ACM Transactions on Intelligent Systems and Technology*.

D'Alessandro, M., Esteller, R., Vachtsevanos, G., Hinson, A., Echauz, J. and Litt, B. (2003 aug). Epileptic seizure prediction using hybrid feature selection over multiple intracranial eeg electrode contacts: a report of four patients. *IEEE Transactions on Biomedical Engineering*, vol. 50, no. 8, pp. 1041–1041. ISSN 0018-9294.
Available at: http://www.ncbi.nlm.nih.gov/pubmed/12769436http://ieeexplore.ieee.org/document/1213858/

Davis, E.A., Keating, B., Byrne, G.C., Russell, M. and Jones, T.W. (1997 jan). Hypoglycemia: incidence and clinical predictors in a large population-based sample of children and adolescents with IDDM. *Diabetes care*, vol. 20, no. 1, pp. 22–5. ISSN 0149-5992.
Available at: http://www.ncbi.nlm.nih.gov/pubmed/9028688

Ebner, A., Sciarretta, G., Epstein, C. and Nuwer, M. (1999). EEG instrumentation. *Guidelines of the IFCN (EEG suppl. 52)*, pp. 7–10.

EPILEPSIAE (2017). The European Epilepsy Database.
Available at: http://epilepsy-database.eu/

Epilepsy Foundation (2013). What Causes Epilepsy and Seizures? | Epilepsy Foundation.
Available at: http://www.epilepsy.com/learn/epilepsy-101/what-causes-epilepsy-and-seizures

Farrar, H.C., Chande, V.T., Fitzpatrick, D.F. and Shema, S.J. (1995 jul). Hyponatremia as the Cause of Seizures in Infants: A Retrospective Analysis of Incidence, Severity, and Clinical Predictors. *Annals of Emergency Medicine*, vol. 26, no. 1, pp. 42–48. ISSN 01960644.
Available at: http://linkinghub.elsevier.com/retrieve/pii/S0196064495702369

Fergus, P., Hignett, D., Hussain, A., Al-Jumeily, D. and Abdel-Aziz, K. (2015). Automatic epileptic seizure detection using scalp EEG and advanced artificial intelligence techniques. *BioMed Research International*, vol. 2015. ISSN 23146141.

Fisch, B.J. (1999). *Fisch & Spehlmann's EEG Primer*. 3rd edn. Elsevier, Amsterdam, the Netherlands.

Fisher, R., Salanova, V., Witt, T., Worth, R., Henry, T., Gross, R., Oommen, K., Osorio, I., Nazzaro, J., Labar, D., Kaplitt, M., Sperling, M., Sandok, E., Neal, J., Handforth, A., Stern, J., DeSalles, A., Chung, S., Shetter, A., Bergen, D., Bakay, R., Henderson, J., French, J., Baltuch, G., Rosenfeld, W., Youkilis, A., Marks, W., Garcia, P., Barbaro, N., Fountain, N., Bazil, C., Goodman, R., McKhann, G., Babu Krishnamurthy, K., Papavassiliou, S., Epstein, C., Pollard, J., Tonder, L., Grebin, J., Coffey, R. and Graves, N. (2010 may). Electrical stimulation of the anterior nucleus of thalamus for treatment of refractory epilepsy. *Epilepsia*, vol. 51, no. 5, pp. 899–908. ISSN 00139580.
Available at: http://doi.wiley.com/10.1111/j.1528-1167.2010.02536.x

Flint Hills Scientific (2017*a*). Flint Hills Scientific, LLC.
Available at: http://www.fhs.lawrence.ks.us/

Flint Hills Scientific (2017*b*). Flint Hills Scientific, LLC : ECoG database.
Available at: http://www.fhs.lawrence.ks.us/PublicECoG.htm

Forsgren, L. (2008). Epidemiology and Prognosis of Epilepsy and its Treatment. In: *The Treatment of Epilepsy*, pp. 21–42. Blackwell Science, Inc., Malden, Massachusetts, USA. ISBN 9780470752463.
Available at: http://doi.wiley.com/10.1002/9780470752463.ch2

FSPP (2017). Database description âĂŤ Seizure Prediction Project Freiburg.
Available at: http://epilepsy.uni-freiburg.de/database

Gabor, A., Leach, R. and Dowla, F. (1996 sep). Automated seizure detection using a self-organizing neural network. *Electroencephalography and Clinical Neurophysiology*, vol. 99, no. 3, pp. 257–266. ISSN 00134694.
Available at: http://linkinghub.elsevier.com/retrieve/pii/0013469496960010

Gabor, A.J. (1998 jul). Seizure detection using a self-organizing neural network: validation and comparison with other detection strategies. *Electroencephalography and clinical neurophysiology*, vol. 107, no. 1, pp. 27–32. ISSN 0013-4694.
Available at: http://www.ncbi.nlm.nih.gov/pubmed/9743269

Gardner, A. (2004). *A novelty detection approach to seizure analysis from intracranial EEG*. Ph.D. thesis, Georgia Institute of Technology.

Goldberger, A.L., Amaral, L.A.N., Glass, L., Hausdorff, J.M., Ivanov, P.C., Mark, R.G., Mietus, J.E., Moody, G.B., Peng, C.-K. and Stanley, H.E. (2000). PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals. *Circulation*, vol. 101, no. 23, pp. e215–e220.
Available at: http://circ.ahajournals.org/cgi/content/full/101/23/e215

Gotman, J. (1982 nov). Automatic recognition of epileptic seizures in the EEG. *Electroencephalography and clinical neurophysiology*, vol. 54, no. 5, pp. 530–40. ISSN 0013-4694.
Available at: http://www.ncbi.nlm.nih.gov/pubmed/6181976

Gotman, J. (1990 oct). Automatic seizure detection: improvements and evaluation. *Electroencephalography and clinical neurophysiology*, vol. 76, no. 4, pp. 317–24. ISSN 0013-4694.
Available at: http://www.ncbi.nlm.nih.gov/pubmed/1699724

Gotman, J., Skuce, D., Thompson, C., Gloor, P., Ives, J. and Ray, W. (1973). Clinical applications of spectral analysis and extraction of features from electroencephalograms with slow waves in adult patients. *Electroencephalography and Clinical Neurophysiology*, vol. 35, pp. 225–235.

Granerod, J. and Crowcroft, N.S. (2007 aug). The epidemiology of acute encephalitis. *Neuropsychological Rehabilitation*, vol. 17, no. 4-5, pp. 406–428. ISSN 0960-2011.
Available at: http://www.ncbi.nlm.nih.gov/pubmed/17676528http://www.tandfonline.com/doi/abs/10.1080/09602010600989620

Grewal, S. and Gotman, J. (2005). An automatic warning system for epileptic seizures recorded on intracerebral EEGs. *Clinical Neurophysiology*, vol. 116, pp. 2460–2472.
Available at: http://ac.els-cdn.com/S1388245705002671/1-s2.0-S1388245705002671-main.pdf?_tid=2d296136-533a-11e7-8f7a-00000aacb35e&acdnat=1497689679_596baef8618ae98e748936e2224d5a63

Guttag, J.V. (2017). personal email correspondence.

Hesdorffer, D.C., Hauser, W.A., Annegers, J.F. and Rocca, W.A. (1996 aug). Severe, uncontrolled hypertension and adult-onset seizures: a case-control

study in Rochester, Minnesota. *Epilepsia*, vol. 37, no. 8, pp. 736–41. ISSN 0013-9580.
Available at: http://www.ncbi.nlm.nih.gov/pubmed/8764811

Hill, M.W., Wong, M., Amarakone, A. and Rothman, S.M. (2000 oct). Rapid cooling aborts seizure-like activity in rodent hippocampal-entorhinal slices. *Epilepsia*, vol. 41, no. 10, pp. 1241–8. ISSN 0013-9580.
Available at: http://www.ncbi.nlm.nih.gov/pubmed/11051118

Hively, L. and Protopopescu, V. (2003 may). Channel-consistent forewarning of epileptic events from scalp EEG. *IEEE Transactions on Biomedical Engineering*, vol. 50, no. 5, pp. 584–593. ISSN 0018-9294.
Available at: http://ieeexplore.ieee.org/document/1198248/

Hoh, B.L., Nathoo, S., Chi, Y.-Y., Mocco, J. and Barker, F.G. (2011 sep). Incidence of Seizures or Epilepsy After Clipping or Coiling of Ruptured and Unruptured Cerebral Aneurysms in the Nationwide Inpatient Sample Database: 2002-2007. *Neurosurgery*, vol. 69, no. 3, pp. 644–650. ISSN 0148-396X.
Available at: http://www.ncbi.nlm.nih.gov/pubmed/21499155https://academic.oup.com/neurosurgery/article-lookup/doi/10.1227/NEU.0b013e31821bc46d

Hutcheon, J.A., Lisonkova, S. and Joseph, K. (2011). Epidemiology of pre-eclampsia and the other hypertensive disorders of pregnancy. *Best Practice & Research Clinical Obstetrics & Gynaecology*, vol. 25, no. 4, pp. 391–403. ISSN 15216934.

Iasemidis, L., Shiau, D.-S., Chaovalitwongse, W., Sackellares, J., Pardalos, P., Principe, J., Carney, P., Prasad, A., Veeramani, B. and Tsakalis, K. (2003 may). Adaptive epileptic seizure prediction system. *IEEE Transactions on Biomedical Engineering*, vol. 50, no. 5, pp. 616–627. ISSN 0018-9294.
Available at: http://ieeexplore.ieee.org/document/1198251/

Ihle, M., Feldwisch-Drentrup, H., Teixeira, C.A., Witon, A., Schelter, B., Timmer, J. and Schulze-Bonhage, A. (2010). EPILEPSIAE âĂŞ A European epilepsy database. *Computer Methods and Programs in Biomedicine*, vol. 106, pp. 127–138.
Available at: http://jeti.uni-freiburg.de/papers/Ihle_Datenbank.pdf

ILAE (1981). International League Against Epilepsy.
Available at: http://www.ilae.org/

Jia, W., Kong, N., Li, F., Gao, X., Gao, S., Zhang, G., Wang, Y. and Yang, F. (2005 oct). An epileptic seizure prediction algorithm based on second-order complexity measure. *Physiological Measurement*, vol. 26, no. 5, pp. 609–625.

ISSN 0967-3334.
Available at: http://stacks.iop.org/0967-3334/26/i=5/a=004?key=crossref.e0d238d54689ab975664699bb1942fe2

Karmos, G. and Dombovári, B. (2011). Electroencephalography (EEG) (Elektroenkefalográfia). Dialóg Campus Kiadó, Budapest.

Karunya University (2017). EEG Database of Seizure Disorders for Experts and Application Developers.
Available at: http://www.karunya.edu/research/EEGdatabase/public/index.php

Khamis, H., Mohamed, A. and Simpson, S. (2009 aug). Seizure state detection of temporal lobe seizures by autoregressive spectral analysis of scalp EEG. *Clinical neurophysiology : official journal of the International Federation of Clinical Neurophysiology*, vol. 120, no. 8, pp. 1479–88. ISSN 1872-8952.
Available at: http://www.ncbi.nlm.nih.gov/pubmed/19564130

Khan, Y.U. and Gotman, J. (2003 may). Wavelet based automatic seizure detection in intracerebral electroencephalogram. *Clinical neurophysiology : official journal of the International Federation of Clinical Neurophysiology*, vol. 114, no. 5, pp. 898–908. ISSN 1388-2457.
Available at: http://www.ncbi.nlm.nih.gov/pubmed/12738437

Klem, G.H., Lüders, H.O., Jasper, H.H. and Elger, C. (1999). The ten-twenty electrode system of the International Federation. *Electroencephalography and Clinical Neurophysiology*, vol. 10, no. 2, pp. 371–375. ISSN 00134694.

Kuhlmann, L., Burkitt, A.N., Cook, M.J., Fuller, K., Grayden, D.B., Seiderer, L. and Mareels, I.M.Y. (2009 oct). Seizure Detection Using Seizure Probability Estimation: Comparison of Features Used to Detect Seizures. *Annals of Biomedical Engineering*, vol. 37, no. 10, pp. 2129–2145. ISSN 0090-6964.
Available at: http://link.springer.com/10.1007/s10439-009-9755-5

Kurinczuk, J.J., White-Koning, M. and Badawi, N. (2010). Epidemiology of neonatal encephalopathy and hypoxic-ischaemic encephalopathy. *Early Human Development*, vol. 86, no. 6, pp. 329–338. ISSN 03783782.

Le Van Quyen, M., Martinerie, J., Baulac, M. and Varela, F. (1999 jul). Anticipating epileptic seizures in real time by a non-linear analysis of similarity between EEG recordings. *Neuroreport*, vol. 10, no. 10, pp. 2149–55. ISSN 0959-4965.
Available at: http://www.ncbi.nlm.nih.gov/pubmed/10424690

Le Van Quyen, M., Martinerie, J., Navarro, V., Boon, P., D'Havé, M., Adam, C., Renault, B., Varela, F. and Baulac, M. (2001 jan). Anticipation of epileptic seizures from standard EEG recordings. *The Lancet*, vol. 357, no.

9251, pp. 183–188. ISSN 01406736.
Available at: http://www.ncbi.nlm.nih.gov/pubmed/11213095http://linkinghub.elsevier.com/retrieve/pii/S0140673600035911

Le Van Quyen, M., Soss, J., Navarro, V., Robertson, R., Chavez, M., Baulac, M. and Martinerie, J. (2005 mar). Preictal state identification by synchronization changes in long-term intracranial EEG recordings. *Clinical Neurophysiology*, vol. 116, no. 3, pp. 559–568. ISSN 13882457.
Available at: http://www.ncbi.nlm.nih.gov/pubmed/15721070http://linkinghub.elsevier.com/retrieve/pii/S1388245704004626

Marieb, E. (2015). *Essentials of Human Anatomy & Physiology*. 11th edn. Pearson Education Limited, Essex, England.

Maschio, M. (2012 jun). Brain tumor-related epilepsy. *Current neuropharmacology*, vol. 10, no. 2, pp. 124–33. ISSN 1875-6190.
Available at: http://www.ncbi.nlm.nih.gov/pubmed/23204982http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3386502

McKeon, A., Frye, M.A., Delanty, N. and Mckeon, A. (2008). The alcohol withdrawal syndrome. *J. Neurol. Neurosurg. Psychiatry*, vol. 79, pp. 854–862.
Available at: http://jnnp.bmj.com/cgi/content/full/79/8/854http://jnnp.bmj.com/cgi/content/full/79/8/854#BIBLhttp://jnnp.bmj.com/cgi/eletter-submit/79/8/854

Meier, R., Dittrich, H., Schulze-Bonhage, A. and Aertsen, A. (2008). Detecting epileptic seizures in long-term human EEG: a new approach to automatic online and real-time detection and classification of polymorphic seizure patterns. *Journal of clinical neurophysiology : official publication of the American Electroencephalographic Society*, vol. 25, no. 3, pp. 119–131. ISSN 0736-0258.

Minasyan, G.R., Chatten, J.B., Chatten, M.J. and Harner, R.N. (2010 jun). Patient-Specific Early Seizure Detection From Scalp Electroencephalogram. *Journal of Clinical Neurophysiology*, vol. 27, no. 3, pp. 163–178. ISSN 0736-0258.
Available at: http://www.ncbi.nlm.nih.gov/pubmed/20461014http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC2884286http://content.wkhealth.com/linkback/openurl?sid=WKPTLP:landingpage&an=00004691-201006000-00002

Misra, U. and Kalita, J. (2009 oct). Seizures in encephalitis: Predictors and outcome. *Seizure*, vol. 18, no. 8, pp. 583–587. ISSN 10591311.
Available at: http://linkinghub.elsevier.com/retrieve/pii/S1059131109001319

Mormann, F., Andrzejak, R.G., Elger, C.E. and Lehnertz, K. (2007 feb). Seizure prediction: the long and winding road. *Brain*, vol. 130, no. 2, pp. 314–333.

ISSN 0006-8950.
Available at: http://www.ncbi.nlm.nih.gov/pubmed/17008335http://www.brain.oxfordjournals.org/cgi/doi/10.1093/brain/awl241

Morrell, M. (2006 apr). Brain stimulation for epilepsy: can scheduled or responsive neurostimulation stop seizures? *Current Opinion in Neurology*, vol. 19, no. 2, pp. 164–168. ISSN 1350-7540.

Nasehi, S. and Pourghassem, H. (2013 aug). A Novel Fast Epileptic Seizure Onset Detection Algorithm Using General Tensor Discriminant Analysis. *Journal of Clinical Neurophysiology*, vol. 30, no. 4, pp. 362–370. ISSN 0736-0258.
Available at: http://www.ncbi.nlm.nih.gov/pubmed/23912574http://content.wkhealth.com/linkback/openurl?sid=WKPTLP:landingpage&an=00004691-201308000-00007

Niedermeyer, E. and Lopes da Silva, F. (2012). *Electroencephalography: Basic Principles, Clinical Applications, and Related Fields*. 6th edn. Lippincott Williams & Wilkins, Philadelphia, PA. ISBN 978-0781751261.

Nirantharakumar, K., Marshall, T., Hodson, J., Narendran, P., Deeks, J., Coleman, J.J. and Ferner, R.E. (2012 jul). Hypoglycemia in Non-Diabetic In-Patients: Clinical or Criminal? *PLoS ONE*, vol. 7, no. 7, p. e40384. ISSN 1932-6203.
Available at: http://dx.plos.org/10.1371/journal.pone.0040384

Noachtar, S., Binnie, C., Ebersole, J., Mauguière, F., Sakamoto, A. and Westmoreland, B. (1999). A glossary of terms most commonly used by clinical electroencephalographers and proposal for the report form for the EEG findings. The International Federation of Clinical Neurophysiology. *Electroencephalography and Clinical Neurophysiology.*, vol. 52, pp. 21–41. ISSN 0424-8155.
Available at: http://www.clinph-journal.com/pb/assets/raw/HealthAdvance/journals/clinph/Chapter1-5.pdf

Obeid, I. and Picone, J. (2016). The Temple University Hospital EEG Data Corpus. *Frontiers in neuroscience*, vol. 10, p. 196. ISSN 1662-4548.
Available at: http://www.ncbi.nlm.nih.gov/pubmed/27242402http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4865520

Odding, E., Roebroeck, M.E. and Stam, H.J. (2006 jan). The epidemiology of cerebral palsy: Incidence, impairments and risk factors. *Disability and Rehabilitation*, vol. 28, no. 4, pp. 183–191. ISSN 0963-8288.
Available at: http://www.ncbi.nlm.nih.gov/pubmed/16467053http://www.tandfonline.com/doi/full/10.1080/09638280500158422

O'Donnell, R.D., Berkhout, J. and Adey, W. (1974 aug). Contamination of scalp EEG spectrum during contraction of cranio-facial muscles. *Electroencephalography and Clinical Neurophysiology*, vol. 37, no. 2, pp. 145–151. ISSN 00134694.
Available at: http://linkinghub.elsevier.com/retrieve/pii/0013469474900054

Olejniczak, P. (2006). Neurophysiologic basis of EEG. *Journal of clinical neurophysiology : official publication of the American Electroencephalographic Society*, vol. 23, no. 3, pp. 186–189. ISSN 0736-0258.

Osorio, I., Frei, M.G. and Wilkinson, S.B. (1998 jun). Real-Time Automated Detection and Quantitative Analysis of Seizures and Short-Term Prediction of Clinical Onset. *Epilepsia*, vol. 39, no. 6, pp. 615–627. ISSN 0013-9580.
Available at: http://doi.wiley.com/10.1111/j.1528-1157.1998.tb01430.x

Page, A., Sagedy, C., Smith, E., Attaran, N., Oates, T. and Mohsenin, T. (2015 feb). A Flexible Multichannel EEG Feature Extractor and Classifier for Seizure Detection. *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 62, no. 2, pp. 109–113. ISSN 1549-7747.
Available at: http://ieeexplore.ieee.org/document/6996043/

Park, Y., Luo, L., Parhi, K.K. and Netoff, T. (2011 oct). Seizure prediction with spectral power of EEG using cost-sensitive support vector machines. *Epilepsia*, vol. 52, no. 10, pp. 1761–1770. ISSN 00139580.
Available at: http://www.ncbi.nlm.nih.gov/pubmed/21692794http://doi.wiley.com/10.1111/j.1528-1167.2011.03138.x

Patel, K., Chua, C.-P., Faul, S. and Bleakley, C.J. (2009). Low power real-time seizure detection for ambulatory EEG. In: *Proceedings of the 3d International ICST Conference on Pervasive Computing Technologies for Healthcare*. ICST.
Available at: http://eudl.eu/doi/10.4108/ICST.PERVASIVEHEALTH2009.6019

Peberdy, M., Callaway, C., Neumar, R., Geocadin, R.G., Zimmerman, J.L., Donnino, M., Gabrielli, A., Silvers, S.M., Zaritsky, A.L., Merchant, R., Vanden Hoek, T.L. and Kronick, S.L. (2010). Part 9: Post-cardiac arrest care: 2010 American Heart Association Guidelines for Cardiopulmonary Resuscitation and Emergency Cardiovascular Care. *Circulation*, vol. 122, no. SUPPL. 3. ISSN 00097322.

PhysioNet (2016). PhysioNet: CHB-MIT Scalp EEG Database.
Available at: https://www.physionet.org/pn6/chbmit/

Picone, J. (2017). The TUH EEG Corpus.
Available at: https://www.isip.piconepress.com/projects/tuh_eeg/

Polychronaki, G.E., Ktonas, P.Y., Gatzonis, S., Siatouni, A., Asvestas, P.A., Tsekou, H., Sakas, D. and Nikita, K.S. (2010 aug). Comparison of fractal dimension estimation algorithms for epileptic seizure onset detection. *Journal of Neural Engineering*, vol. 7, no. 4, p. 046007. ISSN 1741-2560.
Available at: http://www.ncbi.nlm.nih.gov/pubmed/20571184http://stacks.iop.org/1741-2552/7/i=4/a=046007?key=crossref.bd07dfb407857367bde93a904743197c

Ponten, S.C., Ronner, H.E., Strijers, R.L.M., Visser, M.C., Peerdeman, S.M., Vandertop, W.P., Beishuizen, A., Girbes, A.R.J. and Stam, C.J. (2010). Feasibility of online seizure detection with continuous EEG monitoring in the intensive care unit. *Seizure*, vol. 19, no. 9, pp. 580–586. ISSN 10591311.

Qu, H. and Gotman, J. (1995 dec). A seizure warning system for long-term epilepsy monitoring. *Neurology*, vol. 45, no. 12, pp. 2250–4. ISSN 0028-3878.
Available at: http://www.ncbi.nlm.nih.gov/pubmed/8848202

Raichle, M.E. and Gusnard, D.a. (2002). Appraising the brain's energy budget. *Proceedings of the National Academy of Sciences of the United States of America*, vol. 99, no. 16, pp. 10237–10239. ISSN 00278424.

Rijsdijk, M., Leijten, F. and Slooter, A. (2008). Continuous EEG monitoring in the Intensive Care Unit.

Roozenbeek, B., Maas, A.I.R. and Menon, D.K. (2013 feb). Changing patterns in the epidemiology of traumatic brain injury. *Nature Reviews Neurology*, vol. 9, no. 4, pp. 231–236. ISSN 1759-4758.
Available at: http://www.nature.com/doifinder/10.1038/nrneurol.2013.22

Rutland-Brown, W., Langlois, J.A., Thomas, K.E., Xi, Y.L. and Lily, Y.X. (2006). Incidence of traumatic brain injury in the United States, 2003. *The Journal of head trauma rehabilitation*, vol. 21, no. 6, pp. 544–8. ISSN 0885-9701.
Available at: http://www.ncbi.nlm.nih.gov/pubmed/17122685

Saab, M. and Gotman, J. (2005 feb). A system to detect the onset of epileptic seizures in scalp EEG. *Clinical Neurophysiology*, vol. 116, no. 2, pp. 427–442. ISSN 13882457.
Available at: http://www.ncbi.nlm.nih.gov/pubmed/15661120http://linkinghub.elsevier.com/retrieve/pii/S1388245704003098

Sander, J.W. (2004 aug). The Use of Antiepileptic Drugs-Principles and Practice. *Epilepsia*, vol. 45, no. s6, pp. 28–34. ISSN 0013-9580.
Available at: http://doi.wiley.com/10.1111/j.0013-9580.2004.455005.x

Schelter, B., Winterhalder, M., Maiwald, T., Brandt, A., Schad, A., Timmer, J. and Schulze-Bonhage, A. (2006 nov). Do False Predictions of Seizures

Depend on the State of Vigilance? A Report from Two Seizure-Prediction Methods and Proposed Remedies. *Epilepsia*, vol. 47, no. 12, pp. 2058–2070. ISSN 00139580.
Available at: http://www.ncbi.nlm.nih.gov/pubmed/17201704http://doi.wiley.com/10.1111/j.1528-1167.2006.00848.x

Scheuer, M.L. (2002 jun). Continuous EEG Monitoring in the Intensive Care Unit. *Epilepsia*, vol. 43, no. s3, pp. 114–127. ISSN 00139580.

Selvaraj, T.G., Ramasamy, B., Jeyaraj, S.J. and Suviseshamuthu, E.S. (2014 oct). EEG Database of Seizure Disorders for Experts and Application Developers. *Clinical EEG and Neuroscience*, vol. 45, no. 4, pp. 304–309. ISSN 1550-0594.
Available at: http://journals.sagepub.com/doi/10.1177/1550059413500960

Sharifian, M. (2012). Hypertensive encephalopathy. *Iranian journal of child neurology*, vol. 6, no. 3, pp. 1–7. ISSN 1735-4668.
Available at: http://www.ncbi.nlm.nih.gov/pubmed/24665265http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3943026

Shellhaas, R. and Clancy, R. (2007). Characterization of neonatal seizures by conventional EEG and single-channel EEG. *Clinical Neurophysiology*, vol. 118, no. 10, pp. 2156–2161. ISSN 13882457.

Shoeb, A., Edwards, H., Connolly, J., Bourgeois, B., Ted Treves, S. and Guttag, J. (2004 aug). Patient-specific seizure onset detection. *Epilepsy & Behavior*, vol. 5, no. 4, pp. 483–498. ISSN 15255050.
Available at: http://www.ncbi.nlm.nih.gov/pubmed/15256184http://linkinghub.elsevier.com/retrieve/pii/S1525505004001593

Shoeb, A.H. (2009). *Application of Machine Learning to Epileptic Seizure Onset Detection and Treatment*. Ph.D. thesis, Massachusetts Institute of Technology.

Shorvon, S.D.S.D. (2005). *Handbook of epilepsy treatment : forms, causes, and therapy in children and adults*. Blackwell Pub. ISBN 1405131349.

Smith, S.W. (2003). *Digital Signal Processing: A Practical Guide for Engineers and Scientists*. 1st edn. Newnes, Burlington, MA.

Stein, A.G., Eder, H.G., Blum, D.E., Drachev, A. and Fisher, R.S. (2000 apr). An automated drug delivery system for focal epilepsy. *Epilepsy research*, vol. 39, no. 2, pp. 103–14. ISSN 0920-1211.
Available at: http://www.ncbi.nlm.nih.gov/pubmed/10759298

Steriade, M., Nunez, A. and Amzica, F. (1993). A novel slow (< 1 Hz) oscillation of neocortical neurons in vivo: depolarizing and hyperpolarizing components.

Journal of Neuroscience, vol. 13, no. 8.
Available at: http://www.jneurosci.org/content/13/8/3252.short

Tagliaferri, F., Compagnone, C., Korsic, M., Servadei, F. and Kraus, J. (2006 mar). A systematic review of brain injury epidemiology in Europe. *Acta Neurochirurgica*, vol. 148, no. 3, pp. 255–268. ISSN 0001-6268.
Available at: http://link.springer.com/10.1007/s00701-005-0651-y

Teasell, R., Bayona, N., Lippert, C., Villamere, J. and Hellings, C. (2007 jan). Post-traumatic seizure disorder following acquired brain injury. *Brain Injury*, vol. 21, no. 2, pp. 201–214. ISSN 0269-9052.
Available at: http://www.ncbi.nlm.nih.gov/pubmed/17364531http://www.tandfonline.com/doi/full/10.1080/02699050701201854

Tekgul, H., Gauvreau, K., Soul, J., Murphy, L., Robertson, R., Stewart, J., Volpe, J., Bourgeois, B. and du Plessis, A.J. (2006 apr). The current etiologic profile and neurodevelopmental outcome of seizures in term newborn infants. *Pediatrics*, vol. 117, no. 4, pp. 1270–80. ISSN 1098-4275.

Teplan, M. (2002). Fundamentals of EEG measurement. *Measurement Science Review*, vol. 2, pp. 1–11. ISSN 15353893.

Universität Bonn (2017). Klinik für Epileptologie.
Available at: http://epileptologie-bonn.de/cms/front_content.php?idcat=193&lang=3

Universiteit Stellenbosch (2017). High Performance Computer 1 (HPC1).
Available at: http://www.sun.ac.za/hpc

Valens, C. (1999). A Really Friendly Guide to Wavelets.
Available at: http://math.ecnu.edu.cn/$\sim$qgu/friendintro.pdf

van Putten, M.J.A.M. and Tavy, D.L.J. (2004). Continuous quantitative EEG monitoring in hemispheric stroke patients using the brain symmetry index. *Stroke*, vol. 35, no. 11, pp. 2489–2492. ISSN 00392499.

Volschenk, A.D., van der Merwe, J. and Fourie, P.R. (2017*a*). Efficacy of a machine learning based seizure detection method developed using short electroencephalographic training samples. *Medical & biological engineering & computing*, , no. Unpublished (in review).

Volschenk, A.D., van der Merwe, J. and Fourie, P.R. (2017*b*). Probabilistic seizure detection with artefact rejection. *Medical & biological engineering & computing*, , no. Unpublished (submitted).

WHO (2016). WHO | Cooling for newborns with hypoxic ischaemic encephalopathy. *WHO*.

Williamson, J.R., Bliss, D.W. and Browne, D.W. (2011 may). Epileptic seizure prediction using the spatiotemporal correlation structure of intracranial EEG. In: *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 665–668. IEEE. ISBN 978-1-4577-0538-0.
Available at: http://ieeexplore.ieee.org/document/5946491/

Wilson, S.B., Scheuer, M.L., Emerson, R.G. and Gabor, A.J. (2004 oct). Seizure detection: evaluation of the Reveal algorithm. *Clinical Neurophysiology*, vol. 115, no. 10, pp. 2280–2291. ISSN 13882457.
Available at: http://www.ncbi.nlm.nih.gov/pubmed/15351370http://linkinghub.elsevier.com/retrieve/pii/S1388245704002032

Winterhalder, M., Maiwald, T., Voss, H.U., Aschenbrenner-Scheibe, R., Timmer, J. and Schulze-Bonhage, A. (2003 jun). The seizure prediction characteristic: a general framework to assess and compare seizure prediction methods. *Epilepsy & behavior : E&B*, vol. 4, no. 3, pp. 318–25. ISSN 1525-5050.
Available at: http://www.ncbi.nlm.nih.gov/pubmed/12791335

Yuan, Q., Zhou, W., Liu, Y. and Wang, J. (2012 aug). Epileptic seizure detection with linear and nonlinear features. *Epilepsy & Behavior*, vol. 24, no. 4, pp. 415–421. ISSN 15255050.
Available at: http://linkinghub.elsevier.com/retrieve/pii/S1525505012003630

Appendix A

Non-physiological artefacts

In this Appendix, data corruption by non-physiological artefacts are illustrated. Abnormal signal amplitude, mains noise, and phase reversal are illustrated.

The seizure in Figure A.1 is taken from **chb18_35**. The seizure takes place at time [2196 → 2264] in the file. The EEG is contaminated by mains noise in channels 4, 8, 14, and 18. The seizure is rejected on the basis of high mains noise. The mains noise in channels 4 and 8 are around 67 μV and the mains noise in channels 14 and 18 are around 25 μV. From Figure A.1 it can clearly be seen that, compared to other channels, there is significant high frequency contamination in channels 14 and 18, while the mains noise in channels 4 and 8 are excessive.

Figure A.2 shows a segment of EEG that displays abnormally high amplitudes. Detail is provided in Figure A.3 of a select few channels. In the example given, the $x_{high} = 1000\mu V$ threshold caused frame rejection for time $t = [246 \rightarrow 266]$. The signals in these figures seem to be almost flat-lined before and after the high signal amplitudes, however this is only due to the scaling on the figures. Figure A.4 shows an example of missing data, represented as zero amplitude signals. Note that 'chb17_303' refers to 'chb17c_03'. The frame that spans $t = [2746 \rightarrow 2748]$ is the first frame rejected.

Figure A.5 shows phase reversal due to electrode $T8$. Clearly the signals based of this electrode exhibit phase reversal relative to each other. From Equation 3.1:

$$\text{If } \; mean(|A+B|) \; < \; mean\left(\left|\frac{A}{x_{phase}}\right|\right) \quad \text{then reject frame}$$

$$\text{for } \; x_{phase} = 2: \quad 21.4896 \; < \; 22.8545 \; = \; \text{reject !}$$

Evidently the frame that spans $t = [1130 \rightarrow 1132]$ should be rejected.

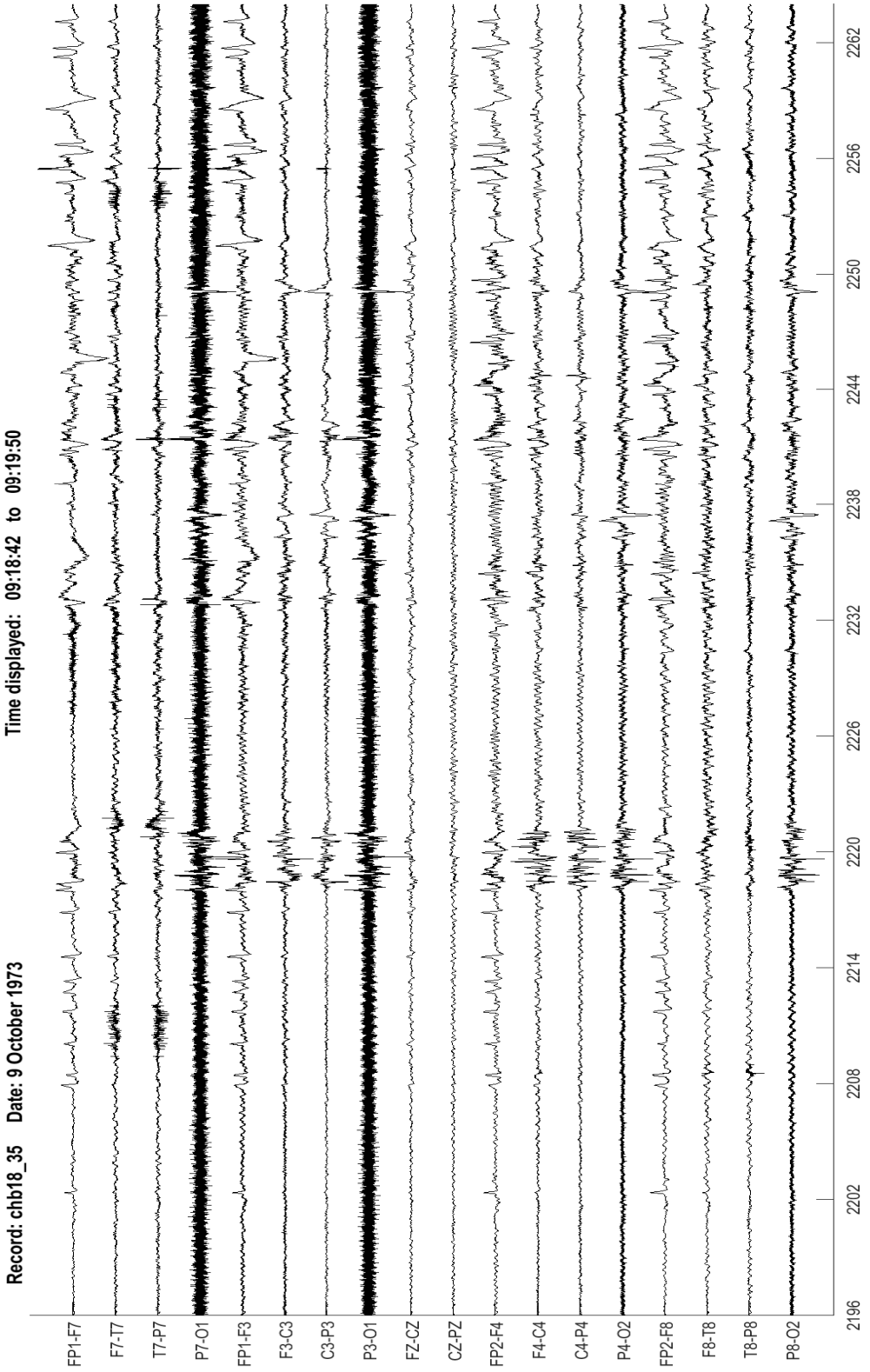


Figure A.1: Rejected seizure due to mains noise

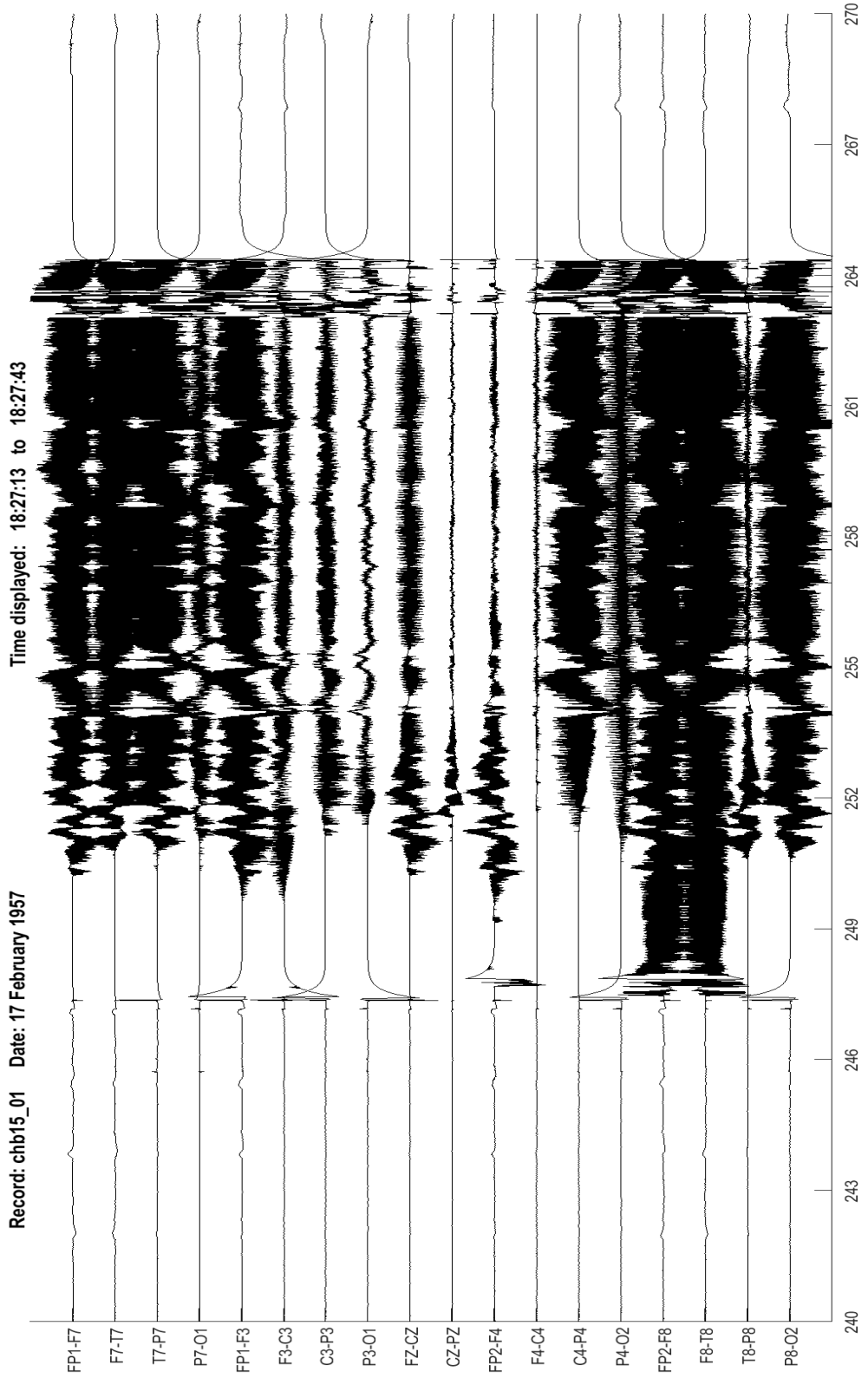


Figure A.2: High amplitude segment

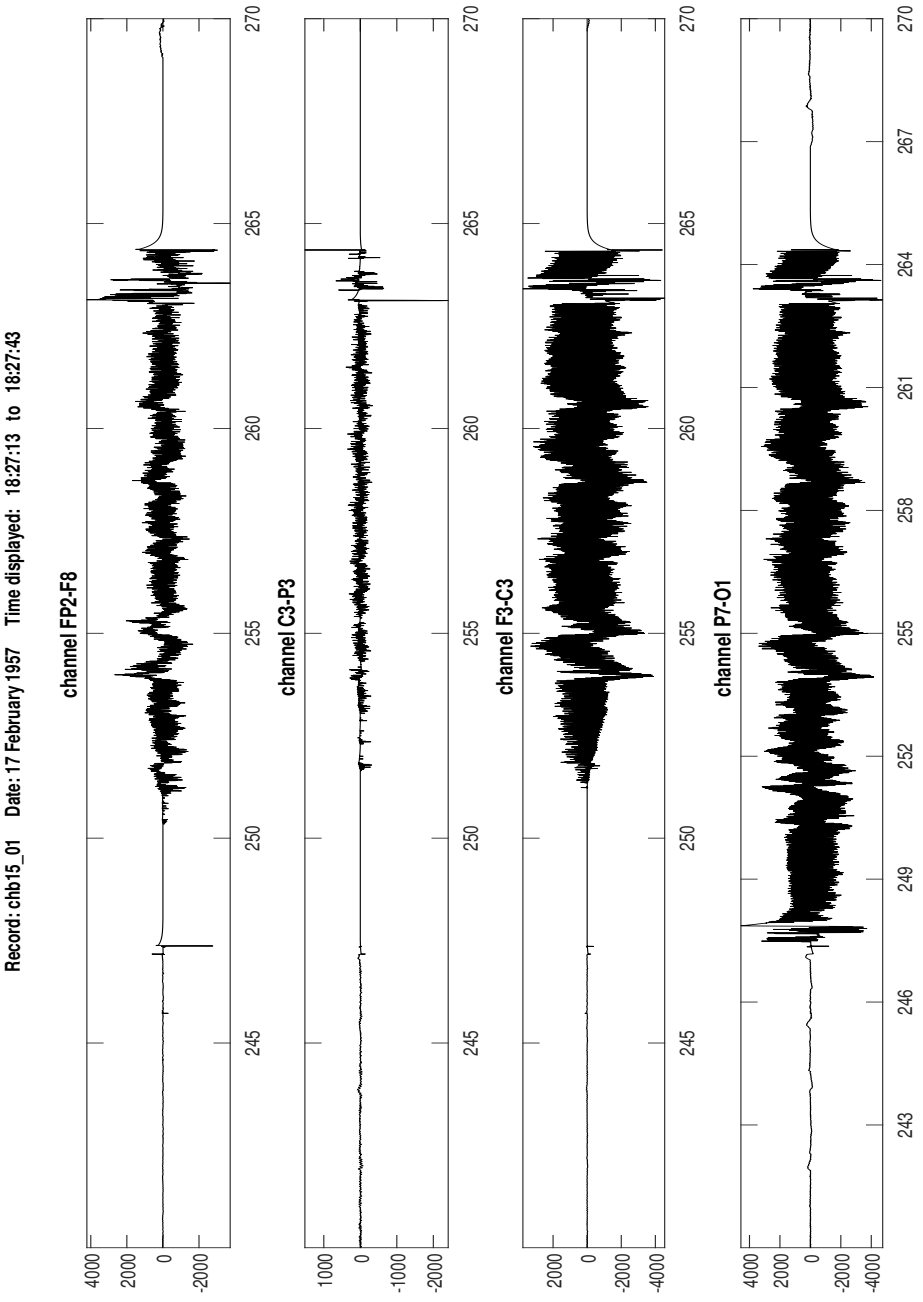


Figure A.3: High amplitude channels

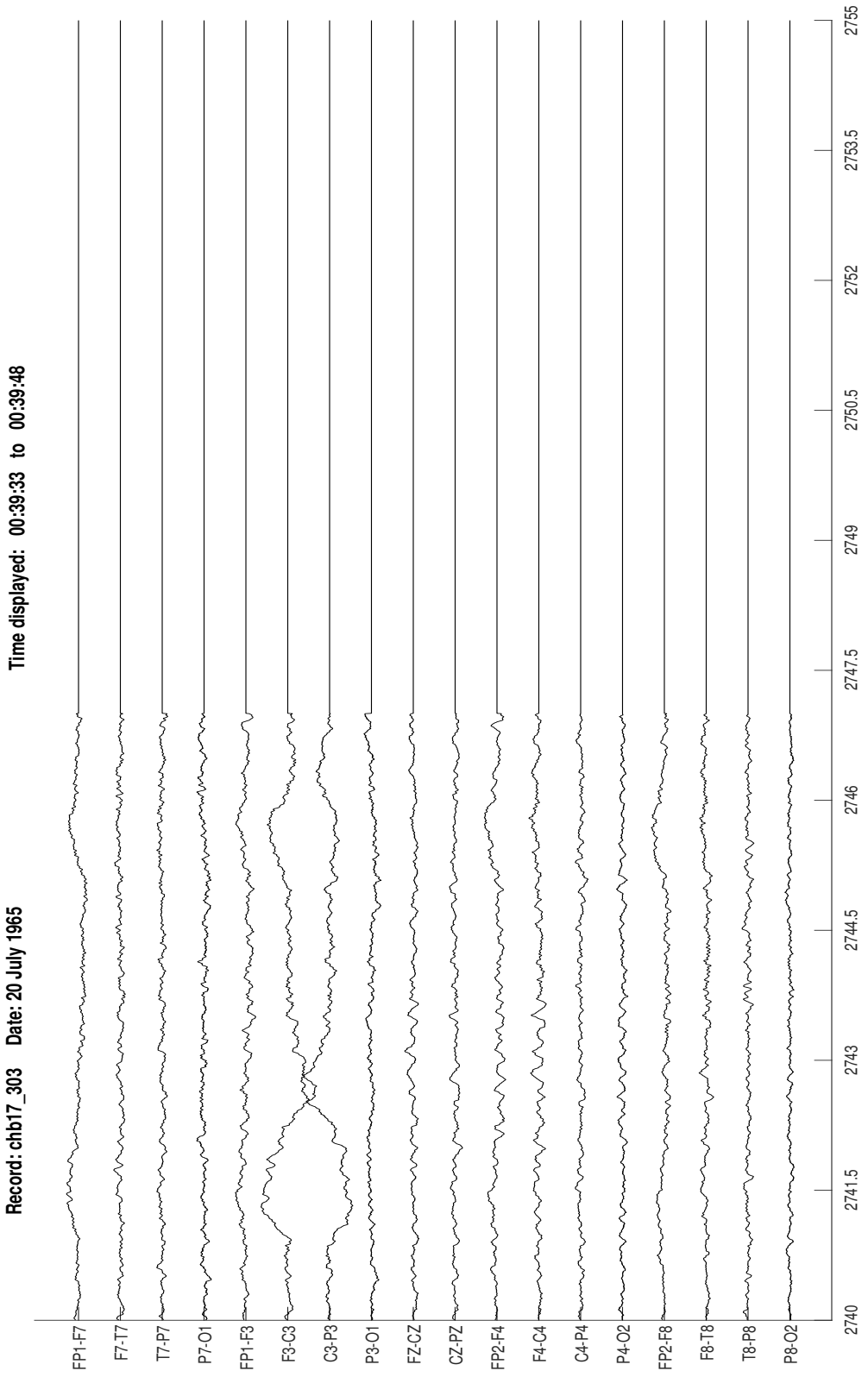


Figure A.4: Zero amplitude segment for chb17c_03

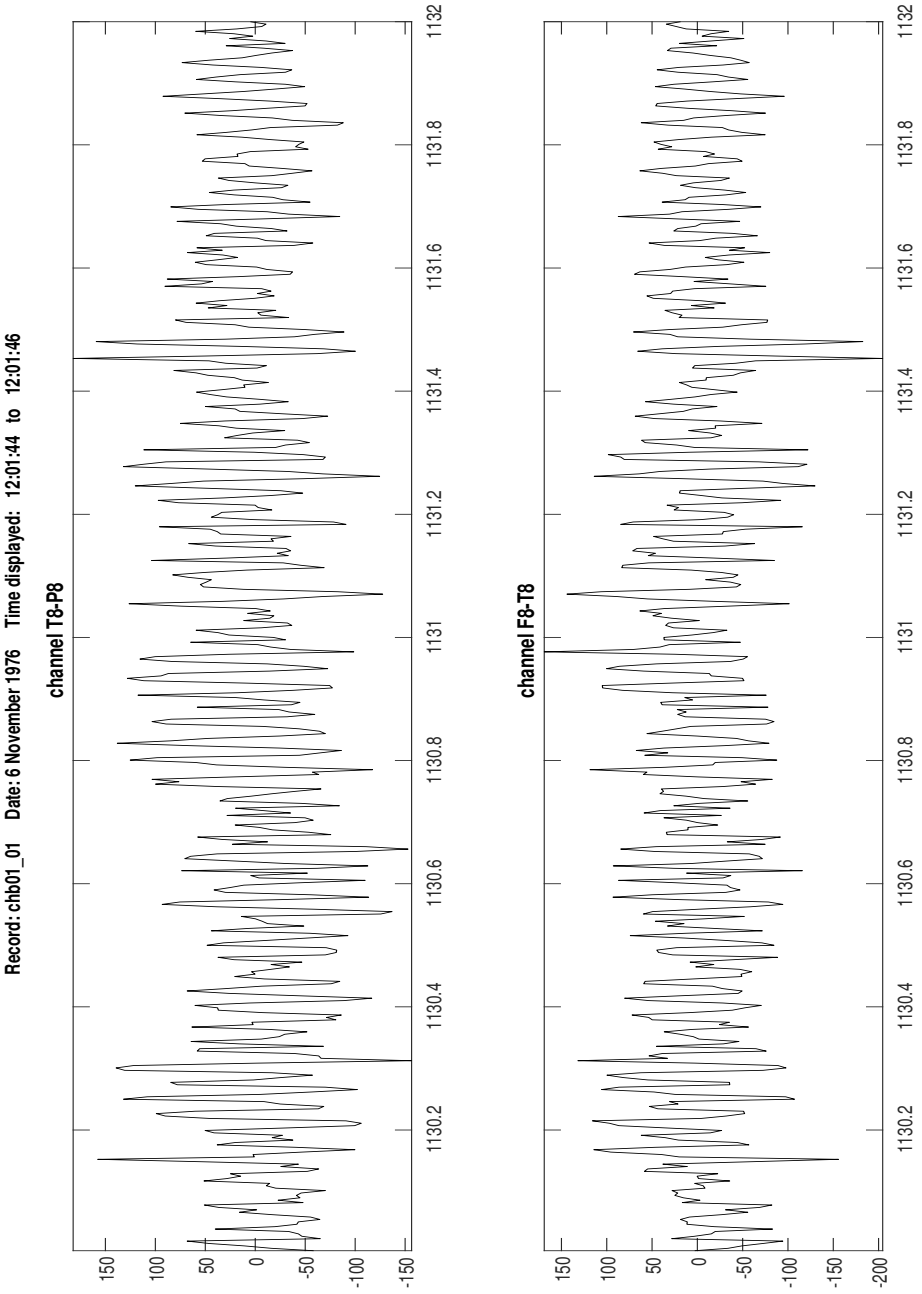


Figure A.5: Phase reversal segment

Appendix B

Iterations

This appendix contains the grid settings of $\Pi_{h(r)}$, $\Pi_{h(m)}$, and Π_p over every iteration in the train-validate phase. The optimal parameters are encircled. Furthermore the F_1 score along with common performance metrics sensitivity (TPR), selectivity (PPV), and false positive rate (FPR) per hour are given.

As indicated in Figure 3.1, a number of parameter selections must be made. These parameter sets $\Pi_{h(r)}$, $\Pi_{h(m)}$, and Π_p must first be initialized.

B.1 Initialization

B.1.1 Select rejection hyperparameters $\Pi_{h(r)}$

Some initial rejection hyperparameters $\Pi_{h(r)} = \{x_{mains}, x_{high}, x_{phase}\}$ needed to be selected. These parameters are used to set the strictness of frame rejection. It was decided that the first set of rejection hyperparameters should not encode a great deal of noise into the model, but should ideally not miss any seizures. The rationale is that some decent estimate of the other parameters ($\Pi_{h(m)}$ and Π_p) should be obtained before finally optimizing $\Pi_{h(r)}$. Too much data rejection or seizure rejection would mean that iterations are not adequate evaluations of the data, and too little rejection would mean that the models may be corrupted by noise.

To obtain some decent set, $\Pi_{h(r)}$, the data in the training set was evaluated. The selection by Saab and Gotman (2005) seemed appropriate also for the CHB-MIT corpus. The settings $\Pi_{h(r)} = \{20\mu V, 1000\mu V, 2\}$ did not reject a very high portion of the data. With these settings, only 1 seizure is rejected in the entire training set. The rejected seizure from patient 18 (file 35) is shown in Figure A.1 in Appendix A. The seizure is rejected on the basis of high mains noise. The mains noise in channels 4 and 8 are around 67 μV. No other seizures in the entire training set have mains noise above 20 μV. Since other seizures have much lower noise, it was considered acceptable to maintain the threshold at 20 μV. Since only 1 seizure is rejected for the first selection of $\Pi_{h(r)}$, there are still 146 seizures that are not rejected in the training set.

B.1.2 Select model hyperparameters $\Pi_{h(m)}$

The (α) model hyperparameters $\Pi_{h(m)} = \{x_{EMG}, x_{ALP}, x_{NAL}\}$ must be selected next. These parameters contribute to the α-classifier used to reject detections that are more likely α activity than seizures. As mentioned in Section 3.6.2, the model creation step is an expensive process. Some proposed low-high values for $\Pi_{h(m)}$ are $x_{EMG} = \{50, 90\}$, $x_{ALP} = \{1, 5\}$, and $x_{NAL} = \{15, 50\}$. Once the low-high combinations are evaluated, improved selections can be made. The values for model hyperparameters $\Pi_{h(m)}$ are taken as percentiles as proposed in Section 3.4.2.

B.1.3 Select parameters Π_p

The parameters $\Pi_p = \{I_{TH}, A_{TH}, x_{NTH}, x_{chn}, N, p, x_{es}, T\}$ must be selected next. As explained in Section 3.6.2, multiple classifiers can be evaluated simultaneously, so more than one value for each of the parameters may be selected. Every possible combination is evaluated.

Parameter T is optimized last. Parameter T essentially removes a block of data from analysis. In order to use as much data as possible for the optimization of all other parameters in Π_p, a low value $T = 60$ [sec] is selected initially. As discussed in Section 3.5.2, parameter T may reduce FPs, but at the same time reduce TN.

B.2 Offline evaluation performance data

Table B.1 contains the selections of $\Pi_{h(m)}$ and Π_p evaluated in a grid throughout numerous offline iterations. The optimal algorithm settings are

Table B.1: Offline parameter set

Parameter	**Value(s)**											
$\Pi_{h(r)}$ = { 20, 1000, 2 }												
x_{EMG}	50	90	100	(NA)								
x_{ALP}	1	5	15	25	35	50	(NA)					
x_{NAL}	15	50	75	95	(NA)							
I_{TH}	0.8	1	1.3	1.4	(1.45)	1.5	1.6	1.7	2	2.5	3	3.8
A_{TH}	0.8	1	1.5	2.5	5	(NA)						
N_{TH}	0.7	0.8	0.9	1	(1.1)	1.2	1.3					
x_{chn}	2	(3)	4	5	6							
N	1	7	8	10	13	14	(15)	16	18	20		
p	0.2	0.3	(0.35)	0.4	0.45	0.5	0.6	1				
x_{es}	(2)	3	4									
T	(60)											
TPR= 0.5999	PPV= 0.5749		FPR= 0.25 /h			F_1 = 0.5871						

encircled. The performance of the optimal combination of settings is given at the bottom of the table.

Next the rejection hyperparameters $\Pi_{h(r)}$ can be optimized. Table B.2 tabulates all selections for $\Pi_{h(r)}$, and optimal settings are encircled. Again the performance of the optimal parameter settings are given.

Table B.2: Offline rejection hyperparameter set

Parameters	**Value(s)**									
x_{mains}	(NA)	140	100	70	50	30	25	20	15	10
x_{max}	(NA)	1500	1250	1000	750	500				
x_{phase}	2.4	2.2	2	1.8	(1.6)	1.5	1.4	1.3	1.2	1.1
$\Pi_{h(m)}$ = { NA, NA, NA }										
Π_p = { 1.45, NA, 1.1, 3, 15, 0.35, 2, 60 }										
TPR=0.5873 PPV=0.5989 FPR=0.2045 /h F_1 =0.5931										

Further optimization of $\Pi_{h(r)}$, $\Pi_{h(m)}$, and Π_p (excluding T) yielded no improvement in algorithm performance, and so these iterations are not detailed here. The optimal $\Pi_{h(r)}$, $\Pi_{h(m)}$, and Π_p (excl. T) have therefore been determined.

The persistence parameter T is evaluated next. It was explained in Section B.1.3 that parameter T should ideally be optimized last to allow for other parameters to be optimized. Table B.3 tabulates the selections for T. The algorithm performance increases with increasing T. Since the maximum threshold for parameter T is set at 180 seconds (see Section 3.5.2), the optimal is selected where $T = 180$ seconds.

It can be said that the offline phase converged with a final classifier. Table B.4 presents the confusion matrix metrics, along with the detection latency. For the provided metrics, counting terms (units) are provided with corresponding durations below in *italics*, if applicable.

Table B.3: Persistence test

Param.	**Value(s)**
$\Pi_{h(r)}$ = { NA, NA, 1.6 }	
$\Pi_{h(m)}$ = { NA, NA, NA }	
Π_p = { 1.45, NA, 1.1, 3, 15, 0.35, 2, ? }	
T	10 20 30 60 80 120 150 180 240 360 480 1080
TPR=0.6929 PPV=0.4999 FPR=0.3297 /h F_1 = 0.5808	
T = 10	TPR = 0.6929 PPV = 0.4728 FPR = 0.4144 F1 = 0.5620
T = 20	TPR = 0.6929 PPV = 0.4728 FPR = 0.4115 F1 = 0.5621
T = 30	TPR = 0.6929 PPV = 0.4748 FPR = 0.4057 F1 = 0.5635
T = 60	TPR = 0.6929 PPV = 0.4931 FPR = 0.3934 F1 = 0.5761
T = 80	TPR = 0.6909 PPV = 0.4957 FPR = 0.3821 F1 = 0.5773
T = 120	TPR = 0.6909 PPV = 0.4959 FPR = 0.3683 F1 = 0.5774
T = 150	TPR = 0.6909 PPV = 0.4968 FPR = 0.3476 F1 = 0.5780
T = 180	TPR = 0.6929 PPV = 0.4999 FPR = 0.3297 F1 = 0.5808
T = 240	TPR = 0.6968 PPV = 0.5069 FPR = 0.3000 F1 = 0.5868
T = 360	TPR = 0.6968 PPV = 0.5108 FPR = 0.2742 F1 = 0.5894
T = 480	TPR = 0.6968 PPV = 0.5149 FPR = 0.2583 F1 = 0.5922
T = 1080	TPR = 0.7057 PPV = 0.5464 FPR = 0.2179 F1 = 0.6159

Table B.4: Offline confusion matrix

Case	TP **[count]** ***[seconds]***	TN ***[seconds]***	FP **[count]** ***[seconds]***	FN **[count]** ***[seconds]***	t_{dl} ***[seconds]***
01	4		0	3	
	720	*145078*	*0*	*189.4278*	*23.25*
02	2		0	1	
	360	*126541.6667*	*0*	*57.3333*	*22*
03	4		0	3	
	720	*135913.7145*	*0*	*172.2855*	*23.25*
04	3		5	1	
	900	*559579.5*	*1260*	*94.5*	*20.3333*
06	1		16	9	
	180	*235052.3*	*4876*	*137.7*	*17*
08	4		1	1	
	720	*70990.4*	*180*	*132.6*	*17.75*
10	4		0	3	
	720	*179172.4287*	*0*	*191.5713*	*18.75*
11	3		4	0	
	540	*123997*	*720*	*0*	*15.6667*
12	11		5	16	
	5724	*66304*	*1980*	*468*	*-11.2727*
13	1		5	11	
	180	*117340*	*1158*	*122*	*28*
14	0		2	8	
	0	*93071.04*	*360*	*168.96*	*NA*
15	10		3	10	
	2160	*139785*	*1620*	*471*	*40.1*
17	3		40	0	
	540	*36266*	*38818*	*0*	*34.3333*
18	4		8	2	
	640	*125731.3334*	*1808*	*105.6667*	*21.75*
19	2		20	1	
	332	*98401.3333*	*8934*	*78.6667*	*24.5*
20	7		0	1	
	1980	*97349.25*	*0*	*36.75*	*24.2857*
21	2		0	2	
	360	*117729.5*	*0*	*99.5*	*35*
22	3		2	0	
	540	*110711*	*360*	*0*	*21*
23	3		0	4	
	900	*94488*	*0*	*222*	*40.6667*

B.3 Online evaluation performance data

Table B.5 tabulates the optimal settings from the offline phase, as well as the performance metrics generated in the online evaluation.

Table B.5: Online test set

$\Pi_{h(r)} = \{140\mu V, 1500\mu V, 1.6\}$	
$\Pi_{h(m)}$ = Not Applicable	
Parameter	**Value(s)**
I_{TH}	1.45
A_{TH}	Not Applicable
N_{TH}	1.1
x_{chn}	3
N	15
p	0.35
x_{es}	2
T	180
TPR=0.5850 PPV=0.4061 FPR=0.3536 /h F_1 = 0.4794	

Table B.6 presents the confusion matrix metrics, along with the detection latency for each case. For the provided metrics, counting terms (units) are

Table B.6: Online confusion matrix

Case	TP **[count]** ***[seconds]***	TN ***[seconds]***	FP **[count]** ***[seconds]***	FN **[count]** ***[seconds]***	t_{dl} ***[seconds]***
05	4		7	1	
	720	*137730.4*	*1848*	*111.6*	*30.75*
07	3		6	0	
	540	*238867*	*1981*	*0*	*19.6667*
09	4		2	0	
	1080	*242898*	*360*	*0*	*42.25*
16	0		27	10	
	0	*48892*	*19424*	*84*	*NA*
24	2		1	14	
	238	*75650.875*	*180*	*573.125*	*25*

provided with corresponding durations below in *italics*, if applicable.

Appendix C

Probability histories

In this Appendix, the probability (P_{ICT}) output histories are given over time for selected cases. In every figure, P_{ICT} is given as a solid blue line, and the onsets of seizures are given by dotted red lines.

C.1 Online evaluation probability history

In this section the probability histories of output during the online evaluation (Section 4.2) is provided. The histories of Cases 05, 07, 09, 16, and 24 are illustrated in Figures C.1, C.2, C.3, C.4, and C.5, respectively.

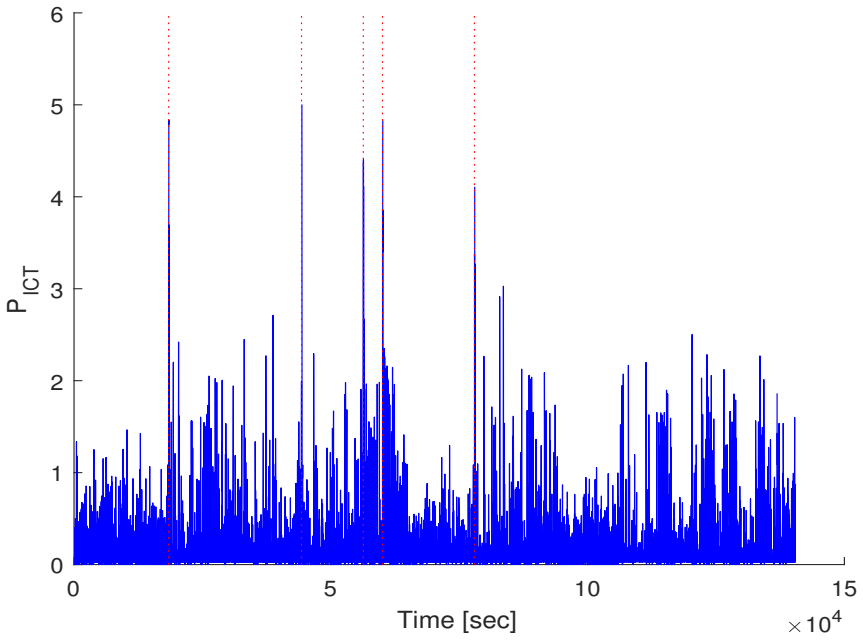


Figure C.1: Probability history of Case 05

In this section it is clear that the global seizure detection threshold I_{TH} should be made relative. Evidently for each case (aside from case 16), a good threshold could be selected for high TPR and PPV. The threshold would however not be a set, global threshold. Some investigation into relative thresholding may lead to promising improvements to the current algorithm.

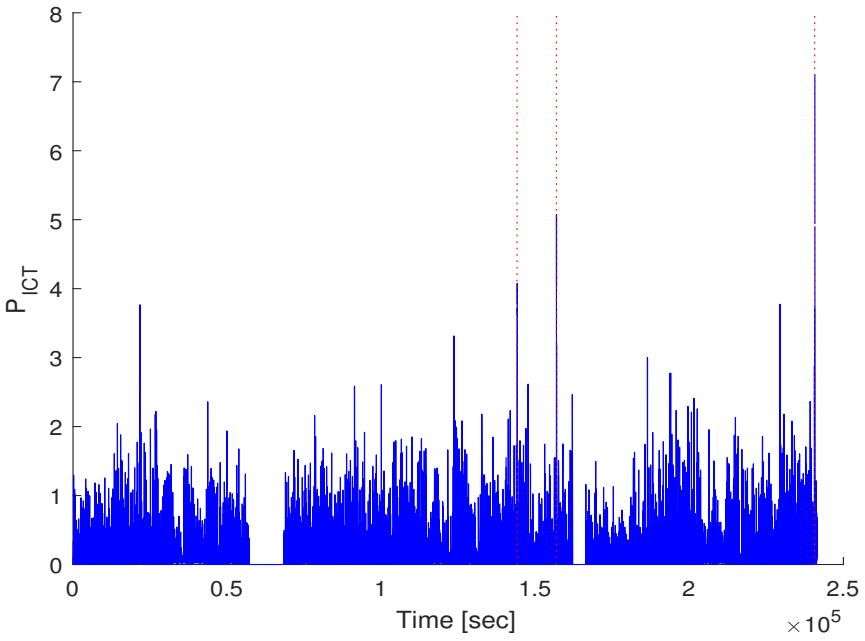


Figure C.2: Probability history of Case 07

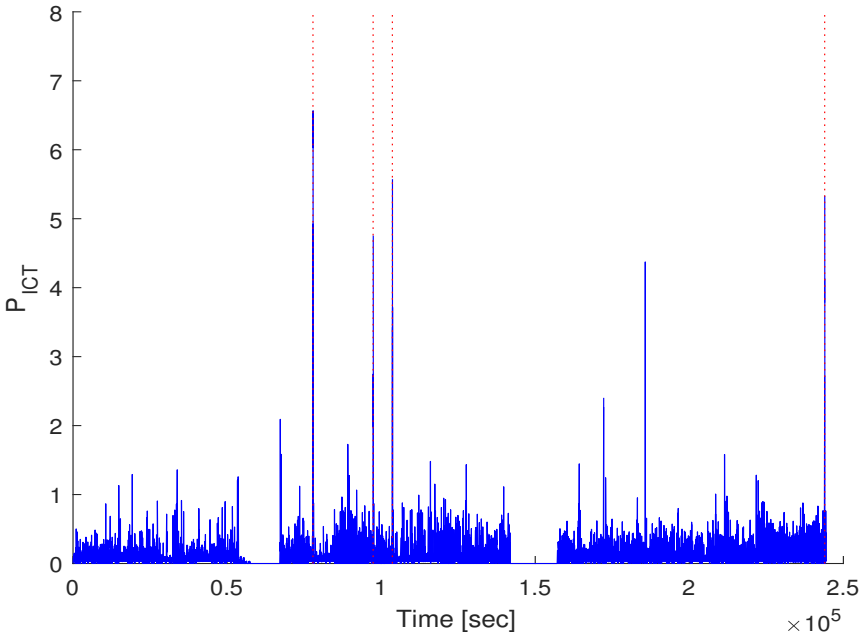


Figure C.3: Probability history of Case 09

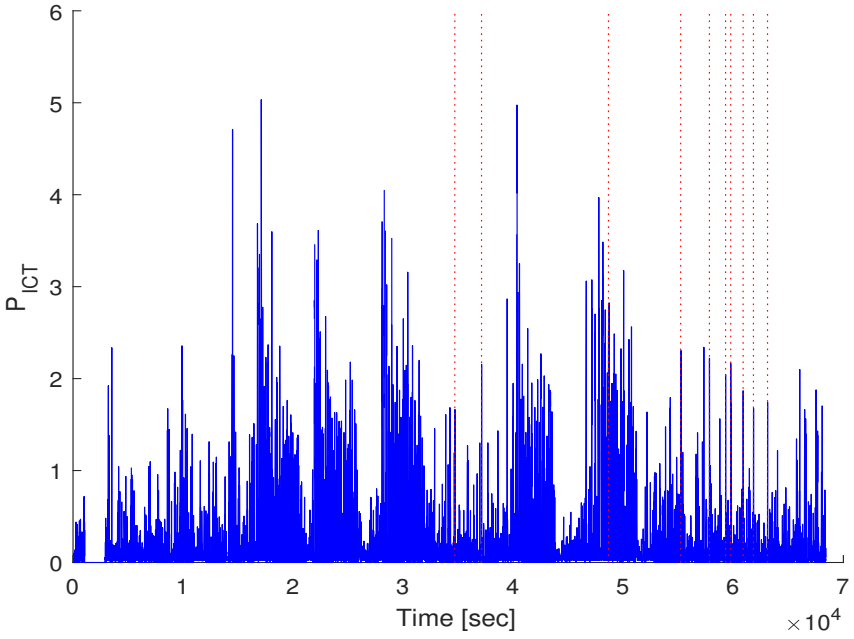


Figure C.4: Probability history of Case 16

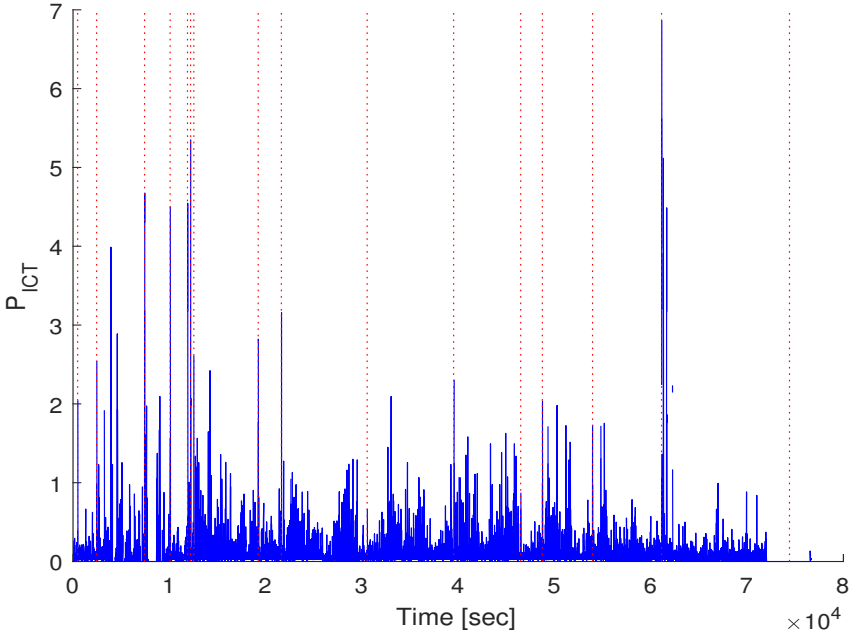


Figure C.5: Probability history of Case 24

C.2 Probability histories with dynamic learning

In order to evaluate the potential of dynamic learning, an addition to the current method is implemented. When processing a given patient in the training set, dynamic learning is implemented. The cases on which the classifier performed poorly in the train-validate phase is evaluated. Cases 06, 13, 14, and 17 are evaluated.

The procedure shall be explained with reference to the Bayes' formula given in Equation 3.5:

$$P(ictal|features) = \frac{P(features|ictal) \cdot P(ictal)}{P(features)}$$

When a given patient is monitored, that patient is referred to as the hold-out case. All other cases in the training set are non-hold-out cases. For the hold-out case, the likelihood function ($P(features|ictal)$) is still derived from all non-hold-out cases, as was done in this study. All other terms of Equation 3.5 are initially derived from non-hold-out cases, however a total of only one hour of inter-ictal data is used. A memory storage variable is created to be filled with the joint-bin counts of new frames. After the hold-out patient has been monitored for 5 minutes, the entire memory storage variable is used to derive a new prior ($P(ictal)$) and normalization constant ($P(features)$), since the terms $P(features|inter-ictal)$, $P(inter-ictal)$ are updated. The posterior ($P(ictal|features)$) is therefore updated. After every new frame imported, the joint-bins represented by the new frame replaces the oldest frame in the memory variable. A new posterior is then created every 5 minutes.

In this procedure all incoming data is assumed to be inter-ictal. This assumption is acceptable, since the ictal data within the last hour, at any time, will be heavily outweighed by the inter-ictal class. Further optimization of this procedure would be required to optimally implement dynamic learning. The purpose for this experiment is however only to evaluate the potential algorithm improvement that dynamic learning could bring.

Figures C.6, C.8, C.10, and C.12 illustrate he probability histories of the selected cases prior to dynamic learning. Figures C.7, C.9, C.11, and C.13 illustrate the histories with the simplistic dynamic learning scheme.

A comparison of the figures provided show that the probability value at actual seizure onset is increased relative to background when implementing dynamic learning. This simplistic implementation would already result in improved results if dynamic thresholding could be implemented successfully.

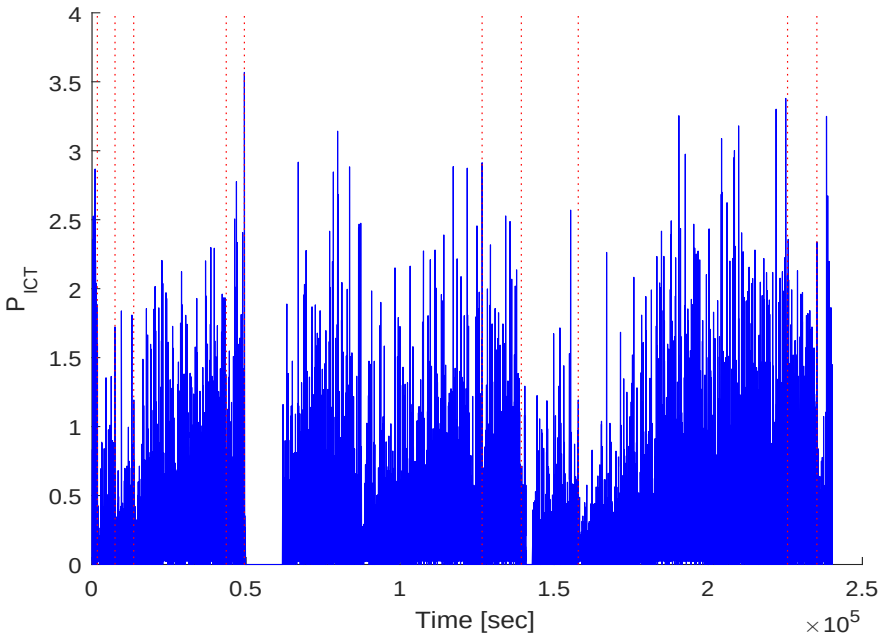


Figure C.6: Probability history of Case 06 (no dynamic learning)

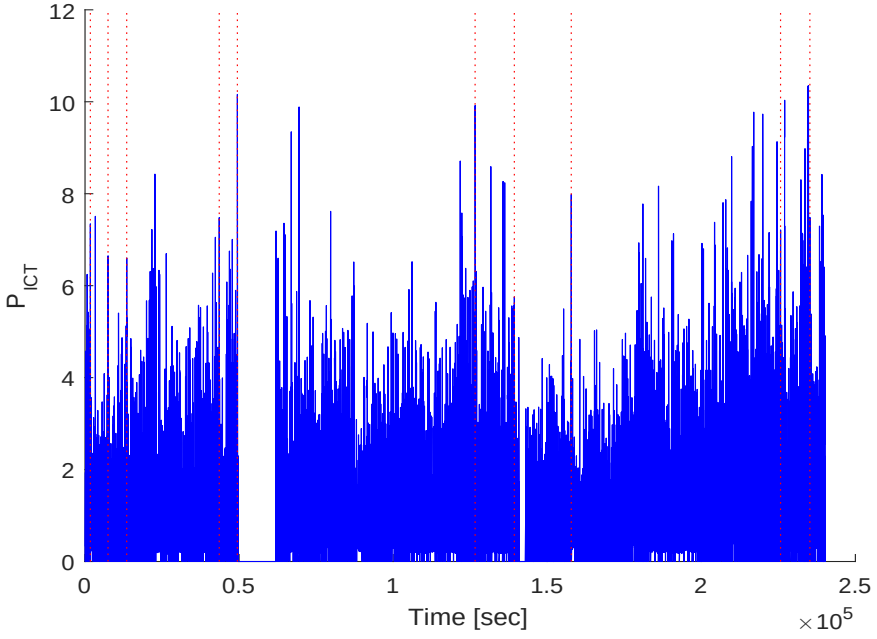


Figure C.7: Probability history of Case 06 (with dynamic learning)

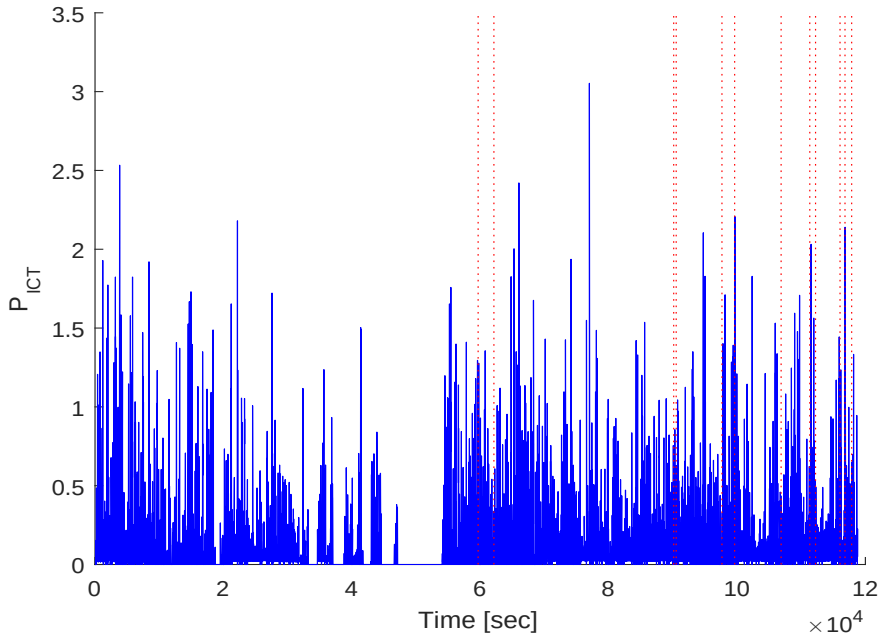


Figure C.8: Probability history of Case 13 (no dynamic learning)

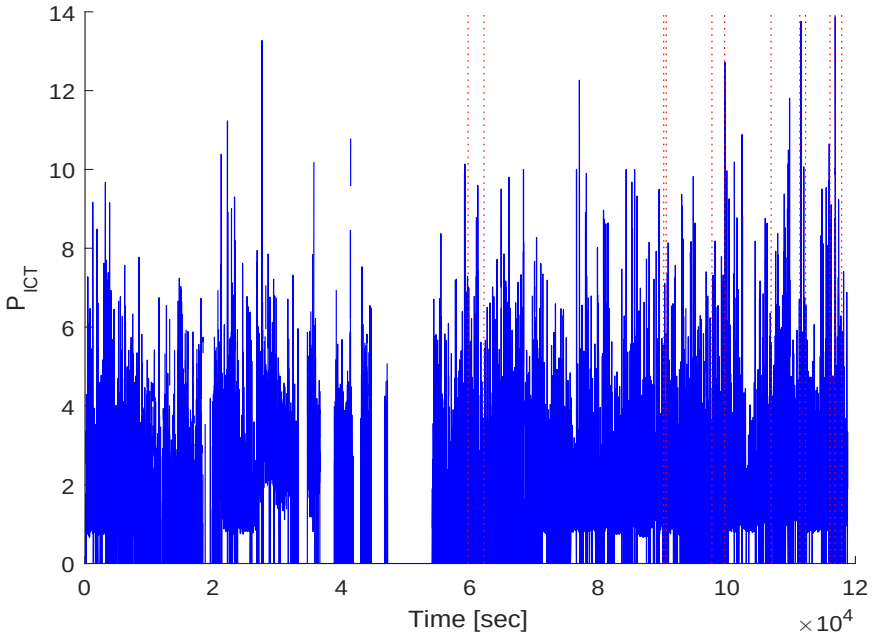


Figure C.9: Probability history of Case 13 (with dynamic learning)

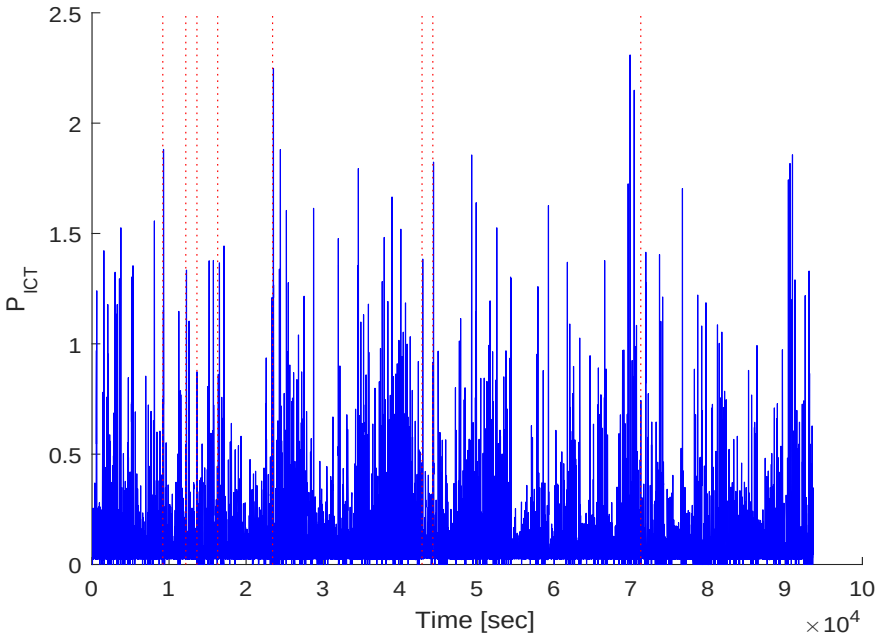


Figure C.10: Probability history of Case 14 (no dynamic learning)

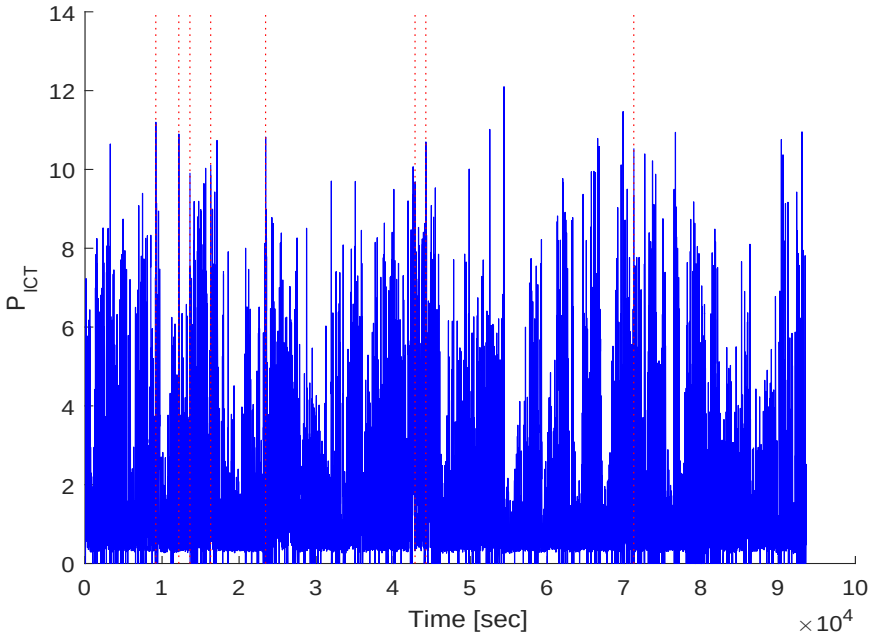


Figure C.11: Probability history of Case 14 (with dynamic learning)

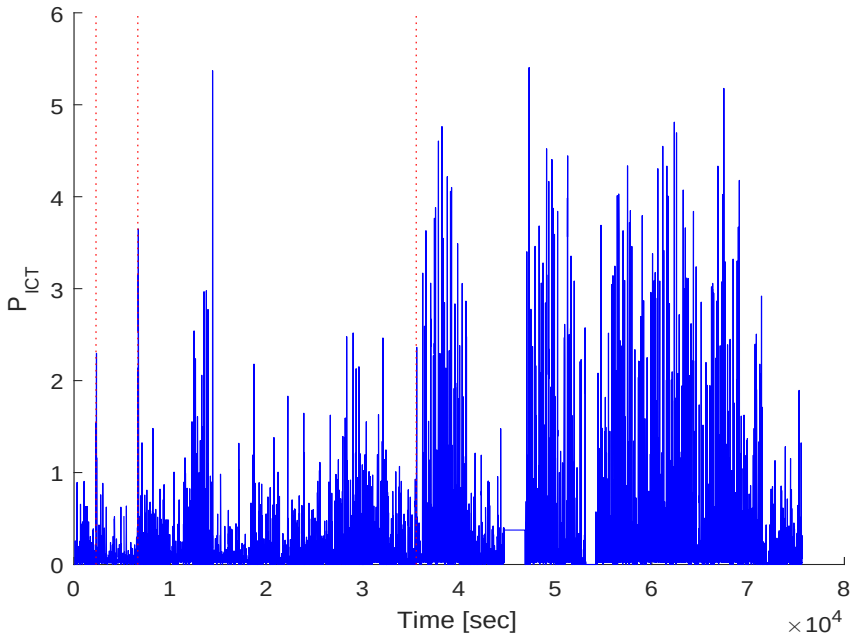


Figure C.12: Probability history of Case 17 (no dynamic learning)

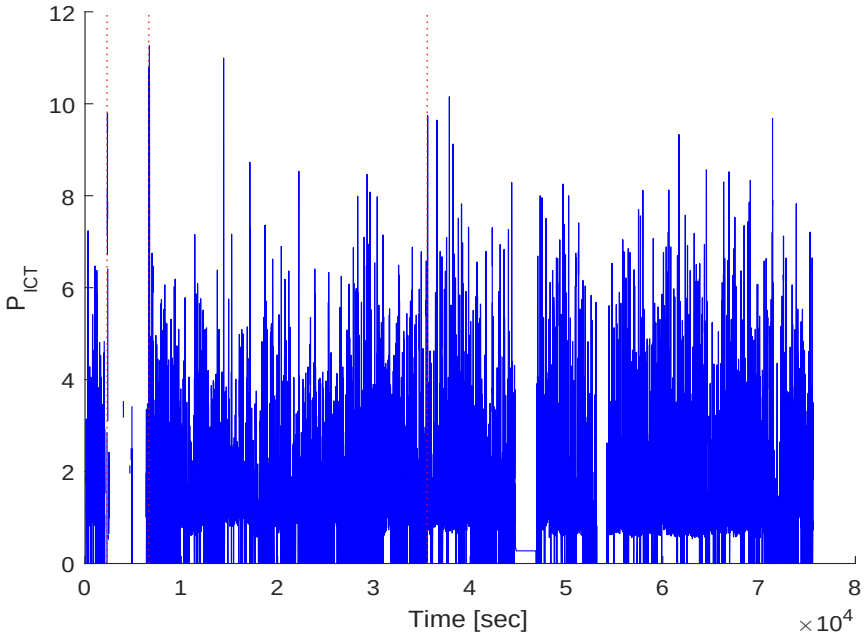


Figure C.13: Probability history of Case 17 (with dynamic learning)

Note that the dynamic learning evaluated here is naïve, simplistic, and unoptimized. Optimized selection of the memory variable size as well as a few other parameters may result in further emphasis on the probabilities at actual seizure onset. This supports the promising improvements that could be achieved with dynamic learning.

According to Saab and Gotman (2005), seizures generally fill the more extreme bins in the histogram (Section 3.4.1). The ictal probability distribution is less clustered than that of inter-ictal probability distributions. Recall that the bin ranges are derived from **all** non-hold-out patients. When the bin ranges are determined it may not encapsulate the seizure ranges for some patients to satisfaction. This leads to inter-patient variability of algorithm performance, and very poor performance on some patients. A novel heuristic is proposed that may overcome there problems. When the mean probability of seizure remains high (above some threshold) then that is indicative that there is not an adequate seizure model. To overcome this, it is proposed that the seizure bin ranges are made more extreme when the mean of the probability history is too high. In doing this, the mean probability will diminish, and seizure events may be highlighted from background. Coupled with dynamic thresholding, this presents a novel heuristic for completely automatic retuning. This novel heuristic is theoretically very promising, and greatly improves the potential for the dynamic learning method proposed.