

Linda Eerikäinen

**Methods for automatic seizure detection
in intensive care: Feature selection and
evaluation**

School of Electrical Engineering

Thesis submitted for examination for the degree of Master of
Science in Technology.

Helsinki 7.4.2014

Thesis supervisor:

Prof. Ari Koskelainen

Thesis advisor:

D.Sc. (Tech.) Mika Särkelä

Author: Linda Eerikäinen

Title: Methods for automatic seizure detection
in intensive care: Feature selection and evaluation

Date: 7.4.2014

Language: English

Number of pages:8+71

Department of Biomedical Engineering and Computational Science

Professorship: Biological Physics and Biomedical Engineering

Code: Tfy-99

Supervisor: Prof. Ari Koskelainen

Advisor: D.Sc. (Tech.) Mika Särkelä

Epileptic seizure is caused by abnormal electrical activity in the brain. When a seizure is nonconvulsive, external indications of seizure, such as muscle contractions, are not visible. Nonconvulsive seizures can be detected only by measuring the electrical signals of the brain with electroencephalogram (EEG).

Nonconvulsive seizures are common in intensive care unit (ICU). Detection of seizures is important, because the delay of diagnosis and duration of seizures have association with mortality and morbidity. For the diagnosis, EEG needs to be reviewed by an experienced reader. The analysis of EEG signals is burdensome and time-taking, and therefore, an automatic detection method for seizures in intensive care would provide a great help.

In this study, seizure markings of two certified EEG readers in EEG records of 50 ICU patients were compared. The agreement between the readers was moderate. Seizure periods agreed by the experts and data from 55 ICU patients without seizures were used to search features from EEG that could distinguish seizure activity from non-seizure activity. 18 features were computed in several time windows from two two-dimensional EEG feature spaces. In addition, spectral features and spike rate were computed from EEG signal. Feature selection was performed with an optimizing method.

Feature combinations of 5, 7, and 10 features were formed. Their performance was compared in an independent data set of EEG records of 40 ICU patients, including patients with and without seizures. 5-feature-model had the best performance among the models.

5-feature-model detected in the independent data set all 11 patients with unequivocal seizures. Median sensitivity over patients was 0.90 and median false positive rate was 0.56 false positives per hour. Results are promising, but further development is needed for reducing the false positive rate.

Keywords: EEG, ICU, epilepsy, seizure, feature selection

Tekijä: Linda Eerikäinen		
Työn nimi: Menetelmiä automaattiselle epileptisten kohtausten tunnistamiselle: Piirteiden valinta ja arviointi		
Päivämäärä: 7.4.2014	Kieli: Englanti	Sivumäärä:8+71
Lääketieteellisen tekniikan ja laskennallisen tieteen laitos		
Professori: Biologinen fysiikka ja lääketieteellinen tekniikka	Koodi: Tfy-99	
Valvoja: Prof. Ari Koskelainen		
Ohjaaja: TkT Mika Särkelä		
<p>Epileptinen kohtaus aiheutuu aivoissa niiden epänormaalista sähköisestä toiminnasta. Kun kohtaus on ei-konvulsiivinen, ulkoisia merkkejä, kuten lihaskouristuksia, ei havaita. Tästä syystä ei-konvulsiiviset kohtaukset voidaan havaita vain mittaamalla aivojen sähköisiä signaaleja aivosähkökäyrällä (elektroenkefalografia, EEG).</p> <p>Ei-konvulsiiviset kohtaukset ovat yleisiä tehohoidossa. Niiden havaitseminen on tärkeää, sillä viivästyneellä diagnoosilla ja kohtauksen kestolla on yhteys kuolleisuuteen ja sairastavuuteen. Diagnosointia varten EEG tarvitsee kokeneen neurofysiologin tulkinnan. Signaalin analysointi on raskasta ja aikaa vievää, ja tästä syystä automaattisesta kohtausten havaitsemisesta olisi tehohoidossa apua.</p> <p>Tässä tutkimuksessa verrattiin kahden neurofysiologin merkintöjä kohtausten ajankohdista 50 tehohoitopotilaalla. Yksimielisyys neurofysiologien välillä oli kohtalainen.</p> <p>Kohtausajanjaksoja, joista asiantuntijat olivat yksimielisiä, sekä dataa 55 tehohoitopotilaalta, joilla ei ollut kohtauksia, käytettiin sellaisten EEG-piirteiden etsimiseen, joilla voitaisiin erottaa kohtaukset jaksoista ilman kohtauksia. 18 piirrettä laskettiin useilla aikaikkunoilla kahdesta kaksiulotteisesta EEG-piirreavaruudesta. Lisäksi EEG:stä laskettiin spektrimuuttujia sekä piikkien määrä minuutissa. Piirteiden valinta suoritettiin optimisointimenetelmällä.</p> <p>Piirreyhdistelmät muodostettiin 5, 7 ja 10 piirteellä, ja niiden suorituskykyä vertailtiin itsenäisellä aineistolla, joka muodostui EEG-mittauksista 40 tehohoitopotilaalla, kohtauksilla ja ilman kohtauksia. 5 piirteen mallilla oli paras suorituskyky. 5 piirteen malli havaitsi itsenäisestä aineistosta kaikki 11 potilasta, joilla oli yksiselitteisiä kohtauksia. Mediaanisensitiivisyys potilaiden yli oli 0.90 ja mediaani väärin havaintojen asteesta 0.56 havaintoa tunnissa. Tulokset ovat lupaavia, mutta lisäkehitystä tarvitaan väärin havaintojen vähentämiseksi.</p>		
Avainsanat: EEG, teho-hoito, epilepsia, epileptinen kohtaus, piirteiden valinta		

Preface

This thesis work was supported by GE Healthcare Finland Oy, and was part of the the SalWe Research Program for Mind and Body of The Finnish Funding Agency for Technology and Innovation (TEKES). In addition to Aalto University School of Electrical Engineering, this thesis work is for the double degree of Master of Science in Technology in the Top Industrial Managers for Europe program completed both in Aalto University and Politecnico di Milano.

I want to thank my team and colleagues at GE for their support and for the valuable learnings of clinical research and product development I have gained during the work. In particular, I want to thank my advisor Mika Särkelä whose enthusiasm towards the subject was contagious and gave a great motivation and inspiration, and whose valuable comments and advice encouraged me during this work. In addition, I want to thank researchers MSc. Antti Tolonen and PhD Mark van Gils from VTT Technical Research Center of Finland for their comments and support.

I want to express my gratitude to professor Bryan Young, University of Western Ontario, Canada, and MD PhD Bin Tu, Columbia University Medical Center, New York, United States, for sharing their expertise and providing the data.

Finally, I want to thank my friends and family for all the support during my studies and this work.

Helsinki, 7.4.2014

Linda M. Eerikäinen

Contents

Abstract	ii
Abstract (in Finnish)	iii
Preface	iv
Contents	v
Symbols and abbreviations	vii
1 Introduction	1
2 Background	3
2.1 Electroencephalography	3
2.2 Electroencephalogram	4
2.2.1 Electroencephalogram in intensive care unit	5
2.2.2 Artifacts	6
2.3 Seizures	7
2.4 Automated seizure detection algorithms	10
3 Material	12
3.1 Development data	12
3.1.1 Reference data	12
3.1.2 Seizure data	12
3.2 Independent data	14
4 Methods	15
4.1 Preprocessing	15
4.1.1 Artifact detection	15
4.2 Feature generation	16
4.2.1 Two-dimensional feature space	16
4.2.2 Principal component analysis	21
4.2.3 Imaginary space	25
4.2.4 Random walk	25
4.2.5 Convex hull	26
4.2.6 Spike detector	27
4.2.7 Spectral features	27
4.2.8 Overview	28
4.3 Evaluation of classifiers	30
4.3.1 Kappa coefficient	30
4.3.2 Correlation coefficient	30
4.3.3 Binary classifier	31
4.4 Feature combination	33
4.5 Feature selection	34
4.5.1 Data selection criteria	34

4.5.2	Sequential floating forward search	34
5	Results	36
5.1	Comparison of two experts	36
5.2	Feature selection	42
5.2.1	Window selection	42
5.2.2	Selection of candidate models	42
5.2.3	Channel fusion	46
5.2.4	Performance in the development data set	46
5.3	Model performance evaluation in an independent data set	51
5.3.1	Any-overlap methods	51
5.3.2	Integral methods	54
5.3.3	Results by patients	56
6	Discussion	62
6.1	Comparison of two experts	62
6.2	Combination of channels	63
6.3	Model comparison and evaluation	63
6.4	Future development	65
7	Conclusions	67
	References	68

Symbols and abbreviations

Symbols

α	EEG band alpha
β	EEG band beta
δ	EEG band delta
θ	EEG band theta
γ	EEG band gamma
κ	Kappa
φ	Phase of a complex number
ψ	Wavelet function
λ	Variance component in principal component analysis
τ	Translation in wavelet transform
Λ	Diagonal matrix
ω	Angular frequency
\vec{a}	The average of the past absolute steps
C	Covariance matrix
d	Distance
e	Eigenvector
\vec{F}	Feature vector
H	Entropy
K	Number of hits in a sector
L	Logit
P	Signal power
p^{angle}	Probability of an angle in a sector
p_e	Expected proportion of agreement
p_o	Observed proportional agreement
p_{step}	Portion of a step from the total path
r	Pearson's correlation coefficient
s	Scale variable in wavelet transform
S	Power spectrum
\hat{S}	Periodogram
t	Time
W	Transformation matrix in principal component analysis
$x(t)$	Continuous signal
$x[t]$	Discrete signal
\vec{z}	Random vector in measurement space

Operators

Abbreviations

ACNS	American Clinical Neurophysiology Society
AUC	Area under curve
cEEG	Continuous electroencephalogram
COM	Center of mass
CWT	Continuous wavelet transform
DWT	Discrete wavelet transform
EEG	Electroencephalogram, electroenphalography
EMG	Electromyogram, electromyography
EMU	Epilepsy monitoring unit
ICU	Intensive care unit
Incr	Phase increment
<i>IF</i>	Instantaneous frequency
<i>LOGPOW</i>	EEG power feature
NaN	Not a number
NCS	Nonconvulsive seizure
NCSE	Nonconvulsive status epilepticus
PCA	Principal component analysis
PED	Periodic epileptiform discharge
PLED	Periodic lateralizing epileptiform discharge
PR	Precision-recall
ROC	Receiver operating characteristics
RWS	Random walker steps
SE	Status epilepticus
std	Standard deviation
SVM	Support vector machine
var	Variance

1 Introduction

Epileptic seizures can be defined as relatively brief disturbances of mental, motor, sensory, or autonomic activity. Seizures can be either convulsive, which means that they consist violent involuntary muscle contractions, or nonconvulsive, in which the muscle contractions are absent. [1] Over 5 minutes of continuous seizure activity or recurrence of two or more seizures with time intervals too short for full recovery are called status epilepticus (SE) [2]. SE requires medication to terminate.

Seizures are common in critically ill patients treated in intensive care unit (ICU) and the majority of the seizures in critically ill are nonconvulsive. Nonconvulsive seizures (NCS) and nonconvulsive status epilepticus (NCSE) can be diagnosed only from electroencephalogram (EEG), which always needs to be reviewed by an expert, and are therefore difficult to detect. [3] [4]

Diagnosis of SE in ICU is often delayed or missed. Detection of the first seizure in all patients who would eventually have a seizure requires more than 48 hours of EEG recording [5]. Median delays from clinical deterioration until diagnosis vary from 48 hours in patients with earlier clinical seizures to 72 hours in patients without history of seizures. [6]

Seizure duration and delay in diagnosis are strongly associated with mortality and morbidity. 10–85% of patients that had nonconvulsive seizures died and 15–30% were disabled depending on the length of the seizure [7]. Due to association with mortality and brain damage, the detection of nonconvulsive seizures has a great importance.

A standard clinical EEG is a measurement of usually 30 minutes. A continuous EEG (cEEG) is used when longterm monitoring is needed. The duration of cEEG is usually 48 hours or more if needed.

Continuous EEG measurement is not currently a standard procedure in ICU. In cEEG it is challenging to obtain high-quality recordings in ICU because critically ill patients are frequently moved to tests or repositioned. Moreover, longterm monitoring generates great amount of raw data for a neurophysiologist to review, causing the analysis to be burdensome. [8]

The indications that usually lead to cEEG measurement for detecting NCSs are altered mental status with or without prior seizures and subtle eye movements. A survey of 330 physicians showed that there is substantial variability in current practice of usage of cEEG and management of NCS and NCSE. The frequency of cEEG review varies greatly and is rarely continuous. Only 18% of EEG is reviewed almost continuously, 17% three or four times per day, 29% twice per day and 21% once per day. Majority of the respondents would start 30 minute EEG when NCS was suspected. [9]

Currently, there are no automatic seizure detection software in the market developed with ICU data, although there are softwares developed with data from epilepsy monitoring units (EMU). An algorithm that detects seizures automatically and produces an alarm every time a seizure is detected would provide help for EEG-reviewer and for ICU personnel, and accelerate the diagnosis. In combination with the algorithm, an EEG-headset that is easy to set and remove, and is suitable for longterm

use, would facilitate the use of cEEG in ICU. This would provide wider use of EEG in ICU and more patients could have better care and faster diagnosis.

In this thesis work the goal is method development for detecting seizures by feature selection and to evaluate the performance of the developed models. The work is focussed, in particular, to detection of seizures with evolution.

The basics of EEG, information about seizures and current seizure detection algorithms are provided in section 2. In section 3 development data and evaluation data are introduced. Section 4 concentrates in methods for feature extraction and feature evaluation for the model development. Results are presented in section 5 and discussion in section 6.

2 Background

2.1 Electroencephalography

Information in the brain is transmitted by electrical signals between neurons. The first measurements of electric brain activity were made in 1924 by German neuropsychiatrist Hans Berger who named the recording electroencephalogram (EEG).

The most widely used electrode system for EEG measurements is internationally standardized 10–20 system. The 21 electrodes of the system are placed on the scalp in positions demonstrated in Figures 1 and 2. The reference points of the system are nasion and inion, i.e. top of the nose and base of the skull in the midline. These reference points are used for measuring in transverse and median planes the perimeters of the skull, which are divided in 10% and 20% intervals for definition of the electrode positions. [10]

Additional intermediate 10% electrode positions to the international 10–20 system are also used. In this system different names are recommended for four electrodes compared to the international 10–20 system. The positions T_3 , T_4 , P_3 , and P_4 are called T_7 , T_8 , P_7 , and P_8 . The nomenclature and locations are standardized by the American Electroencephalographic Society. [10]

In addition, other electrode systems exist for EEG measurements. An electrode montage consisting of eight electrodes outside the hairline have been tested for detection of epileptiform abnormalities using seven channels [11] and for cEEG in ICU using four channels [12].

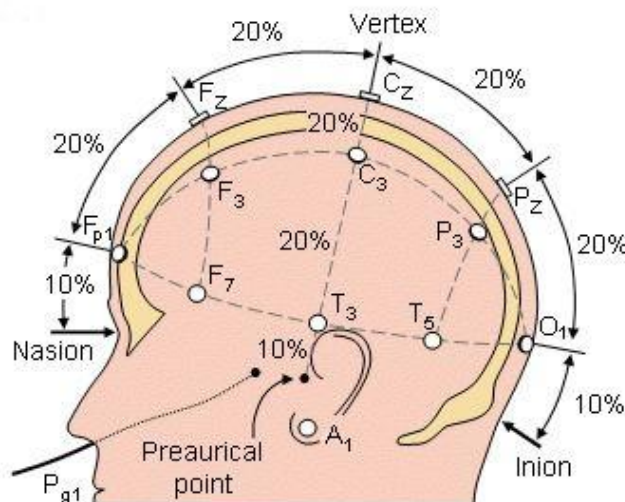


Figure 1: Electrode positions from side in 10–20 EEG montage [10]

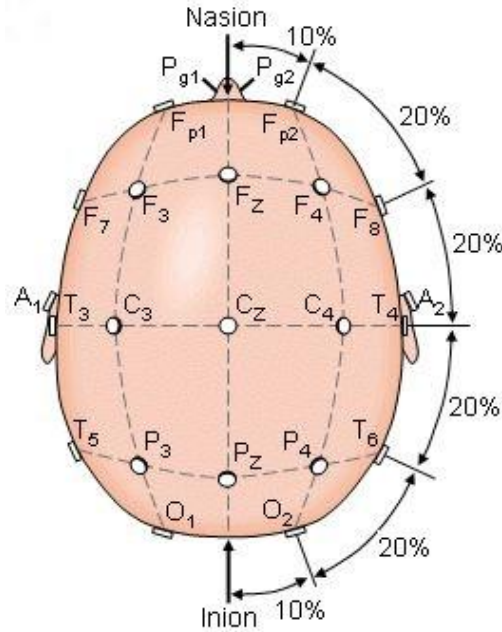


Figure 2: Electrode positions from top in 10–20 EEG montage [10]

Attaching EEG electrodes is time taking and maintaining the contact requires attention. Therefore, there is a variety of commercially available EEG headwear. There are commercial products for diagnostic measurements but also for non-medical use. The forms of EEG headwear vary from caps and nets to rigid headsets.

Algorithm development of this study focuses to a prototype headset designed to be used on ICU patients. The prototype has a setup of 10 referential electrodes. C_z is the reference electrode and ground electrode is located in front of C_z . Other electrodes of the set are T_3 , T_4 , F_3 , F_4 , C_3 , C_4 , P_3 , and P_4 . The setup is designed as a tool for seizure detection. [13]

2.2 Electroencephalogram

Electroencephalogram (EEG) can be measured either from scalp or directly from the surface of the brain. When measured from scalp, the amplitude of the signal is approximately $100 \mu\text{V}$ and from the surface $1\text{--}2 \text{ mV}$.

EEG signal is represented as a graph voltage versus time. Clinicians analyze the signals by searching normal and abnormal patterns. Important features for analysis of the signal are waveform, i.e. shape of the wave, amplitude, i.e. size of the wave in microvolts (μV), and frequency, i.e. number of times a repetitive wave occurs in 1 second.

EEG is divided in four frequency bands: delta (δ) band under 4 Hz, theta (θ) band 4–8 Hz, alpha (α) band 8–13 Hz, beta (β) band 13–20 Hz, and gamma (γ) oscillations 30–70 Hz. Patterns in normal EEG and quantity of activity in different

bands vary with age. When EEG signal contains abnormal patterns, it might be an indicator of a neurological dysfunction. [1]

2.2.1 Electroencephalogram in intensive care unit

In general ICUs more than 10% of patients suffer serious central nervous system complications which associate to their mortality and morbidity [14]. The right classification of the condition is important for the right management of the illness.

EEG provides a sensitive tool for cerebral cortical function assessment of comatose or paralyzed patients. For these patients other assessment tools, such as Glasgow Coma Scale, which is based on eye, verbal, and motor response, or clinical neurological evaluation, cannot be accurately applied or are of limited value. Young et al. [15] have developed an EEG classification system for coma which is presented in Table 1.

Table 1: Coma classification of Young et al.

Category	Subcategory
I Delta/theta > 50% of record (not theta coma)	A. Reactivity B. No reactivity
II Triphasic waves	
III Burst suppression	A. With epileptiform activity B. Without epileptiform activity
IV Alpha/theta/spindle coma (unreactive)	
V Epileptiform activity (not in burst-suppression pattern)	A. Generalized B. Focal or multifocal
VI Suppression	A. < 20 μ V, but > 10 μ V B. \leq 10 μ V

In EEG there are nonspecific patterns of encephalopathy and patterns that suggest specific diagnoses. However, the patterns that are useful in certain diagnoses are rarely distinctively characteristic of a particular disease. [3]

Nonspecific changes in EEG appear during diffuse encephalopathies. In the early stage of encephalopathy, the changes include slowing of the alpha rhythm and excess slowing during wakefulness to theta and then delta rhythm. The following changes are loss of alpha rhythm, more noticeable slowing, loss of normal faster activity, and loss or attenuation of normal sleep transients. In addition, abnormal arousal patterns and frontal intermittent rhythmic delta activity may appear. The worsening of encephalopathy causes changes that include loss of normal state changes and variability, loss of reactivity to external stimuli, burst suppression and in the end

electrocerebral inactivity. Most of these patterns may be produced by a normal brain under strong sedation. [3]

2.2.2 Artifacts

In EEG interpretation it is important to recognize the patterns that are not originating from the brain to avoid misinterpretation. In many cases artifacts can be recognized if there are medium to high amplitude potentials that occur only at one electrode or rhythmical or irregular activity that appears simultaneously in unrelated head regions. [1]

EEG artifacts are divided in two categories based on their origin. The origin is either physiological or non-physiological. The classification between physiological and technical artifacts may overlap, e.g. loose electrode detects movement artifacts and causes artifacts on its own. [16]

Physiological artifacts are blinking and eye-movement, movements in general, muscle artifacts, and electrocardiogram. Artifacts of eye-movement are caused by the potential difference of few millivolts between cornea and retina. When eyes move, the difference in the electric field caused by the movement is picked by electrodes in the vicinity. [16]

Muscle potentials on the scalp cause various forms of artifacts that can be localized or widespread. These artefacts can be reduced by relaxation, change of posture, or, if needed, with filters. Jaw and facial movements are seen in the signal as bursts which are synchronized with the movement and are superimposed to slower waveforms. [16] In Fig. 3 there is an example of face-twitching artifact.

Artifacts from electrocardiogram arise from the electric field associated to cardiac action. This electric field can be picked up to EEG signal, but can also cause movement to the electrodes near arteries due to the pulse pressure wave it produces in the arterial system. [16]

One physiological artefact in the ICU most often recorded in patients with fever is sweat artifact. Sweat causes change in the impedance between the electrode and the skin, and this is seen as slow shifts of the electrical baseline. [17]

The difficulty in artifact removal is that physiological artefacts are signals that can be extremely informative about the state of the patient. In addition, artifact removal methods, such as filtering, may affect to the frequency characteristics of the signal. [16]

Non-physiological artifacts are caused by electrical interference and artifacts arising from electrodes and recording instruments. Mains interference derives from electric and magnetic fields alternating at the frequency of mains supply. In Europe the frequency is 50 Hz and in North America 60 Hz. [16]

Electrode contact and input leads are one source of artifacts. When electrode contact is imperfect, it causes a transient electrochemical change at the electrode-tissue interface. Electrode leads give rise to artifacts when they move or vibrate. Electrode artifacts caused by patient movement are usually widespread and affect many or all channels. [16]

In the ICU the artifacts are often originating from life-support systems, monitor-

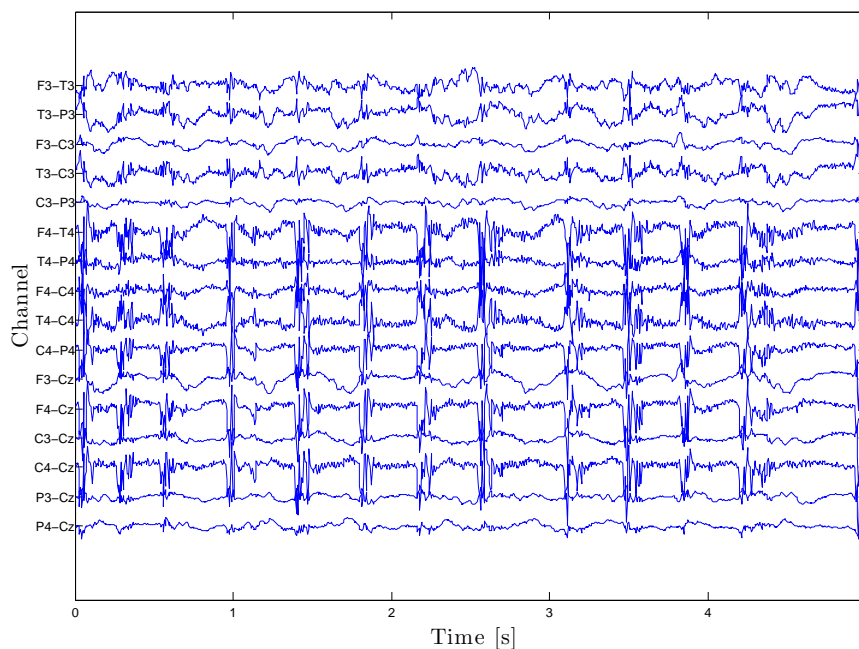


Figure 3: Artifact of face-twitching in EEG signal

ing devices and personnel [3]. Hemoperfusion device induces saw-toothed waveforms that are in the range of 5.5 to 11 Hz. In addition, mechanical ventilator can cause bursts of rhythmic high amplitude slow waves. Personnel, such as nurses, doctors and EEG technologist, and family members of the patient often cause movement artifacts in the environment. [17]

2.3 Seizures

An epileptic seizure is one of the neurological dysfunctions recognizable from EEG. Seizures are caused by abnormal paroxysmal activity in the brain and manifest as relatively brief disturbances of mental, motor, sensory or autonomic activity. Convulsions, which are violent involuntary contractions of muscles, are often present during seizure. [1]

In nonconvulsive seizures muscle contractions are absent. Most NCSs are purely electrographic and even though some subtle signs can be associated with NCSs [8], EEG is needed for diagnosing NCSs.

There is no definitive method for identifying seizures other than visual EEG interpretation. However, visual identification of seizures from EEG varies from one human expert to another. One EEG reader may mark multiple seizures while other reader may mark only one longer seizure or may not mark any seizures. [18]

The patterns in EEG in local epileptiform activity consist of spikes or sharp waves that may be followed by a slow wave. The periods of epileptiform activity during seizure are called ictal and usually persist several seconds. In ictal epileptiform

activity rhythmic, repetitive waveforms begin to appear abruptly. The frequency, topography and form of waveforms vary throughout the seizure. The periods between seizures are typically brief in duration and are called interictal. [1] Post-ictal suppression is a suppression in amplitude of the signal after an ictal phase. In Fig. 4 there is an example of spike activity in an EEG channel recorded from an ICU patient.

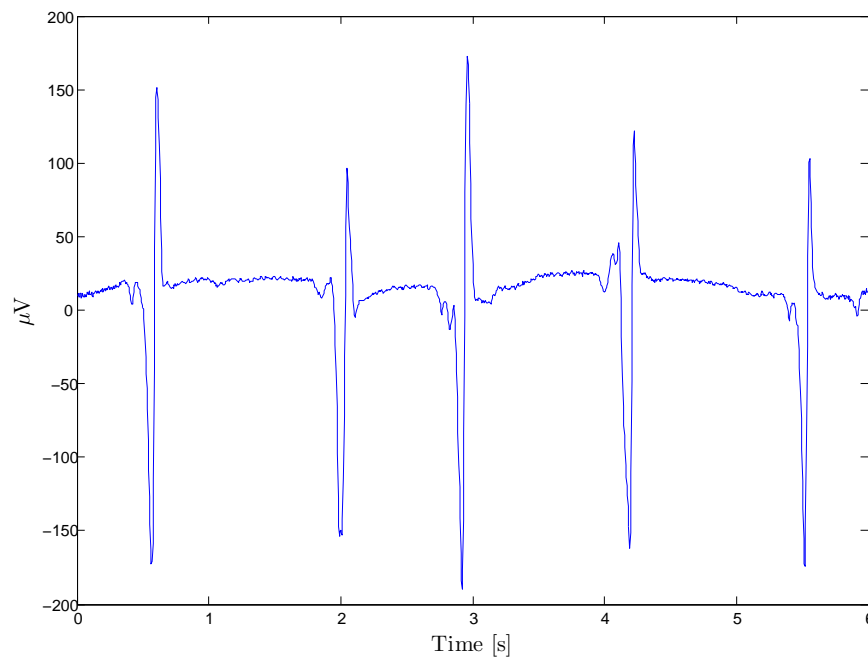


Figure 4: An example of spikes in one channel of EEG

Seizures interpreted from EEG can be divided in two categories according to American Clinical Neurophysiology Society (ACNS) criteria [19]. The categories are unequivocal and equivocal (possible or probable) seizures and the criteria are listed in Table 2. In Fig. 5 there is an example of an unequivocal seizure in an ICU patient.

An EEG pattern that can evolve to electrographic seizure pattern is periodic lateralizing epileptiform discharge (PLED). PLEDs are di- or multiphasic spike or sharp wave complexes that may include a slow wave. The complexes are very brief, usually on fraction of a second. PLED pattern is considered as highly epileptogenic interictal pattern in clinical purposes. In 77% of 170 reported cases seizures were present. [1]

An example of a pattern that is not associated with seizures but is sometimes mistaken for epileptiform activity is triphasic wave pattern. It consist of three phases and every phase has longer duration than preceding one. Triphasic waves are clearly distinguished from background and other slow waves. The second phase has usually the greatest amplitude and the polarity of the second phase is positive. The duration of the whole wave complex is approximately from 0.25 to 0.5 s. [1]

Seizures are common in ICU. Different studies have shown that majority of the

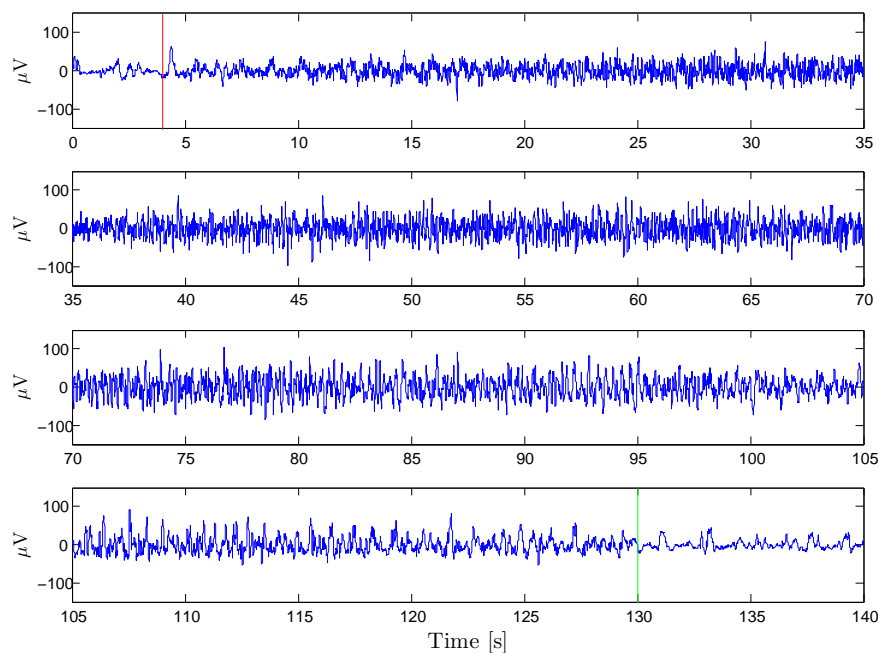


Figure 5: An example of unequivocal seizure. Red line is the start of the seizure and green line is the end.

seizures in critically ill are nonconvulsive [3] [4]. 6-48% of patients in ICU may have NCSs depending on the clinical condition of the patient [5], [8], [20, 21, 22, 23].

Nonconvulsive seizures appear in variety of conditions in critically ill patients. Conditions in which nonconvulsive seizures are recognised in critically ill are central nervous system infection (26% had NCSs) [5], brain tumor (23%) [5], subarachnoid hemorrhage (18%) [5], ischemic stroke (6%) [22], intracerebral hemorrhage

Table 2: Criteria for unequivocal and equivocal seizures

Unequivocal electrographic seizures:

- Generalized spike-wave discharges at 3/s or faster, or
- Clearly evolving discharges of any type that reach a frequency $>4/s$, whether focal or generalized
- All with a clear onset and offset in relation to the background EEG

Equivocal electrographic seizures:

- Evolving periodic epileptiform discharges (PEDs) that are less than 4/s, whether focal or generalized
 - Electrodecremental episodes with a clear clinical correlate of seizure activity
 - High frequency burst activity
-

(28%) [22], and severe traumatic brain injury (22%) [24]. 33% of seizures detected in patients during cEEG monitoring were epilepsy related [5].

Nonconvulsive seizures are associated with increased mortality and morbidity. The two factors affecting to mortality are seizure duration and delay of diagnosis. In one study, patients that had nonconvulsive seizures of duration less than 10h 10% died and 30% were disabled, seizures from 10 to 20h 33% died and 50% were disabled, and seizures that lasted more than 20h 85% died and the rest 15% were disabled. [7]

2.4 Automated seizure detection algorithms

The variety of algorithms that have been developed for seizure detection from EEG is very wide. Different seizure detectors have been developed for adults [25, 26, 27, 28] and neonates [29, 30, 31]. Detectors use either scalp EEG or intracerebral EEG [32]. The algorithms are mainly for general use, but can also be patient-specific [27] [33], i.e. the algorithm is first trained with data from the patient.

In this work the goal is to develop methods that could be used in a commercial algorithm. Therefore, the focus is on seizure detectors that are commercially available.

The first seizure detection method is by Gotman [25], [26], and it has been integrated to several commercial devices. The algorithm is distributed by Natus Medical Incorporated. The method is based on recognition of spikes and sharp waves by breaking down the EEG signals into half-waves. From the half-waves were computed for every epoch an average amplitude, an average duration and a squared coefficient of variation which measures the regularity of half-wave duration. For the three parameters a set criteria had to be met for seizure detection.

The Reveal algorithm of Wilson et al. [27] is distributed by Persyst Development Corporation. Reveal algorithm is based upon Matching Pursuit algorithm that converts EEG into sum of overlapping 'atoms'. These atoms are localized in time and frequency. The classification of seizure components is made by using small neural networks. According to the manufacturer, in version 2010.02.17 of the software, the neural network algorithms have been trained with over 800 long-term EEG recordings from adults and children including more than 33,000 spikes and 670 seizures. The output of the final rule is perception value that varies from 0 to 1 and expresses the degree of certainty of the detection. This value, however, is not visible to the user.

The latest commercially available seizure detection system is IdentEvent algorithm by Kelly et al. [28] distributed by Optima Neuroscience. The algorithm translates EEG signals into three EEG descriptors and examines their spatiotemporal dynamics. The three descriptors are pattern-match regularity statistic, local maximum frequency and amplitude variation. The criteria set to these descriptors are compared to thresholds for defining whether the activity is from seizure.

IdentEvent and Reveal algorithms were compared with the same data set of 47 patients. When the perception scores of Reveal was 0.5, Reveal algorithm reached slightly better sensitivity but had approximately six times more false detections per

24h than IdentEvent. When perception scores were raised to 0.8 and 0.9, sensitivity of Reveal was less than sensitivity of IdentEvent, but false detection rates still remained higher respect to IdentEvent. In this particular patient set the sensitivity of IdentEvent was approximately the same or better than sensitivity of Reveal and false detection rates were significantly smaller. [28]

In the original material of development of Reveal patients from ICU were not included to the training data [27] as well as IdentEvent was evaluated with EMU data [28]. Therefore, functioning with ICU patients is not guaranteed.

In this work the focus is to develop methods for seizure detection in ICU where the needs are different than in EMU. Patients referred to an EMU have previous diagnosis of epilepsy. In EMU the purpose is to establish the diagnosis and find the best therapy, including medical and surgical interventions. Patients undergo long-term video-EEG monitoring and seizure and spike detection software is used for marking the potential seizures and interictal discharges. In addition, trained EEG technologist screens the EEG. [34]

In ICU, patients are critically ill and monitored for their vital functions. Patients are connected to several monitoring devices, which require attention of the nurses and produce alarms. Experienced EEG readers are not present in ICU like in EMU. Continuous EEG measurement is started for seizure detection only when there are indications for need [9]. In addition, EEG in ICU patients is heterogeneous as described in section 2.2.1.

3 Material

3.1 Development data

3.1.1 Reference data

The reference data set for the feature selection in this study consisted of continuous EEG recordings from 55 patients in ICU. Patients for the development were randomly selected from a larger data set, which was the data set in the study by Tanner et al. [35]. Patients with seizures or with other epileptiform activity were excluded from the selection. The recordings were obtained by using full 10–20 montage and recorded with a standard EEG device manufactured by XL-Tek. The total duration of EEG recordings was 35 d 16 h and 3 min and the median duration of recordings was 17 h 22 min.

All EEG recordings were analysed by an experienced neurologist Dr. G. B. Young. The recordings were classified using the system in Table 1 [15]. The reference data included recordings that were classified in categories I-IV and VI, except in class IIIA. List of patients and their coma classification is presented in Table 3.

For feature selection, single derivation per patient from the EEG data was selected. The derivation was selected randomly for the patients in the reference data set.

3.1.2 Seizure data

The data set including epileptic seizures was collected from 50 patients in ICU. The full 10-20 montage and XL-Tek device were used for the seizure data set. The duration of the EEG recordings of all the 50 patients was in total 79 d 17 h and 5 min and the median duration was 38 h 15 min. The derivation with most evident seizure activity was chosen for the feature selection to emphasize the seizure characteristics in the signal. This was done visually by the author.

All recordings were reviewed this time by two EEG readers, Drs. B. Tu and G. B. Young. The seizures marked by the experts were labelled either unequivocal or equivocal according to the criteria explained in section 2.3.

From the 50 patients, 24 patients were selected for feature selection. The criteria for patient selection was that patient had unequivocal seizures marked by both experts.

Table 3: Classification of reference patients in development set according to coma classification of Young et al.

patient ID	Classification	patient ID	Classification
case001	I	case089	I
case003	I	case090	I
case006	I	case093	VI
case007	I	case095	IIIB
case009	VI	case097	I
case011	I	case100	IV
case014	I	case107	II
case015	VI	case110	IIIB
case018	VI	case111	I
case020	I	case114	VI
case021	VI	case118	II
case024	I	case124	VI
case026	II	case125	I
case031	I	case129	I
case034	I	case141	II
case040	II	case142	I
case041	I	case145	VI
case042	VI	case146	IIIB
case044	I	case155	VI
case047	II	case161	II
case054	IV	case165	II
case056	I	case169	VI
case057	I	case171	IIIB
case061	I	case172	I
case067	IIIB	case174	IIIB
case069	I	case175	VI
case071	IIIB	case180	II
case074	IIIB		

3.2 Independent data

Independent data sets were selected from ICU patients with and without seizures for the evaluation of the methods. Data sets included 20 ICU patients with seizures marked by an expert and 20 ICU patients without seizures. Patients without seizures were classified with the coma classification of Young et al. and the proportion of patients from every class were the same than in the development set. Patients without seizures and their classification are listed in Table 4.

Table 4: Classification of reference patients in evaluation set according to coma classification of Young et al.

patient ID	Classification	patient ID	Classification
case002	VI	case092	I
case010	II	case105	I
case017	VI	case120	I
case022	I	case121	IV
case028	I	case131	IIIB
case039	VI	case140	I
case045	I	case151	VI
case053	II	case162	IIIB
case065	II	case167	IIIB
case087	I	case182	I

4 Methods

4.1 Preprocessing

In this study, the signals from a subset of 9 electrodes in full 10-20 montage were used. The 9 electrodes are the electrodes of the headset prototype designed for longterm EEG monitoring in ICU introduced in chapter 2.1 and are C_3 , C_4 , F_3 , F_4 , P_3 , P_4 , T_3 , and T_4 , and C_z as a reference. EEG signal from every channel was first downsampled to 200 Hz and then 16 derivations were formed. The derivations are listed in Table 5.

Table 5: Derivations for one-channel EEG feature calculation

$F_3 - T_3$	$F_4 - T_4$
$T_3 - P_3$	$T_4 - P_4$
$F_3 - C_3$	$F_4 - C_4$
$T_3 - C_3$	$T_4 - C_4$
$C_3 - P_3$	$C_4 - P_4$
$F_3 - C_z$	$F_4 - C_z$
$C_3 - C_z$	$C_4 - C_z$
$P_3 - C_z$	$P_4 - C_z$

Frequency bands of 50 Hz, 60 Hz, and 100 Hz were removed with notch filters. Filtering of bands of 50 Hz and 60 Hz were done to remove mains interference in frequencies used in Europe and in North America. In addition, the signal was high-pass filtered with a cut-off frequency of 0.5 Hz.

4.1.1 Artifact detection

EEG requires artifact removal before the signal can be used for the feature calculations. Artifacts that are removed are electromyogram (EMG) and signals with very high amplitude.

The EMG artifact detector was developed to remove artifacts caused by muscle contractions. The development for the artifact detection was made outside this thesis work. The EEG was filtered with two finite impulse response (FIR) filters. The first filter was a passband filter for the EEG band (2-15 Hz) and the second filter a high pass filter for the EMG band and attenuates frequencies below 35 Hz. From the signals in EEG and EMG bands, sliding standard deviation within 1 second window were calculated. The periods, in which the ratio of standard deviations $\text{std(EMG)}/\text{std(EEG)}$ and standard deviation of EMG std(EMG) were above predefined thresholds, were signed as EMG artifact. If there were gaps smaller than 1 second between periods of artifact, the periods were united and the gap was also marked as artifact. In Fig. 6 there is an example of EMG artifact detection. The sample frequency of detections is 1 Hz.

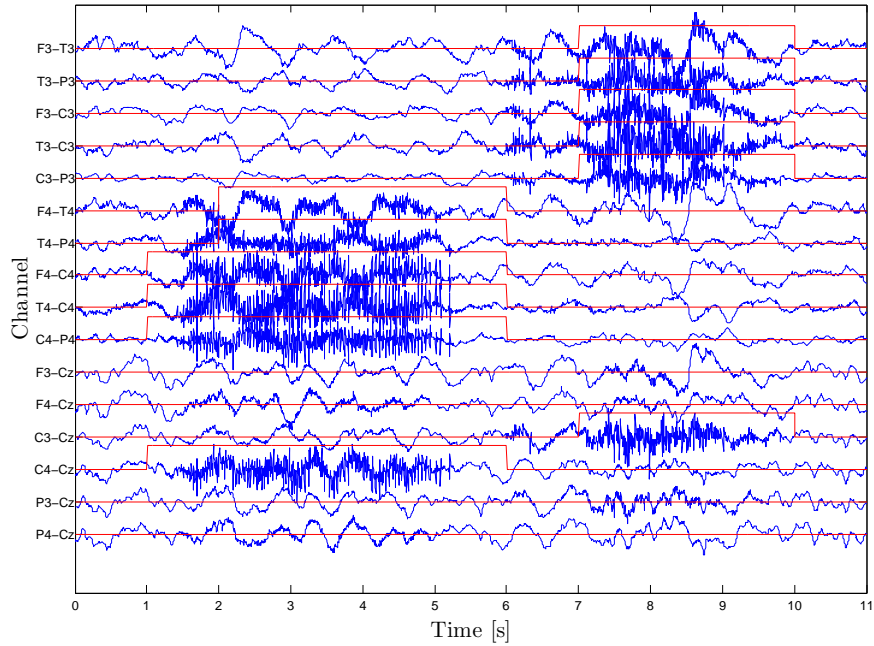


Figure 6: An example of EMG artifact detection. Red line has value 1, when artifact is detected, and 0 when no artifact is detected. EMG artifact detection signal has a sample frequency of 1 Hz

Very high amplitudes in the signal indicate poor electrode contact. When an electrode produces a very high voltage compared to voltages produced by other electrodes the signal is considered being an artifact. For high voltage detection EEG signal was divided in five-second long epochs that start by every second. For every epoch the voltage from an electrode to all other electrodes was calculated. The number of times the voltage to another electrode exceeds a predefined threshold were counted. In an epoch this is done for all the electrodes. The electrode that has the highest count of voltages to other electrodes above the threshold was considered producing artifactual signal. Gaps that are shorter than two minutes between artifact periods are marked as artifact as well. The sample frequency of detections is 1 Hz as in EMG detector.

4.2 Feature generation

4.2.1 Two-dimensional feature space

In this thesis work, method development was started by analyzing possible parameters in a two-dimensional EEG feature space presented earlier in a Master's thesis work by A. Tanner [36]. The features of the two-dimensional space are instantaneous frequency (IF) of the signal and base-10 logarithm of signal power ($LOGPOW$). The objective of this thesis was to find parameters that could distinguish seizure

activity from EEG background activity in this feature space and in EEG signal.

Signal power is computed from the stationary wavelet transform. Wavelet analysis is useful when good localization in time and frequency is desired. The following introduction to wavelet transform is based on book Bioelectrical Signal Processing in Cardiac and Neurological Applications by Sörnmo and Laguna [37].

Wavelets are basis functions that have two parameters: translation in time and scaling in time. With scaling and translation it is possible to analyze the presence of both global waveforms and fine structures in the signal using wavelet analysis.

Scaling and translating a mother wavelet $\psi(t)$ with parameters s and τ defines a family of wavelets $\psi_{s,\tau}(t)$. When $s < 0$ the wavelet is contracted and when $s > 0$ the wavelet is expanded.

Continuous wavelet transform (CWT) of a continuous signal $x(t)$ is defined as

$$w(s, \tau) = \int_{-\infty}^{\infty} x(t) \frac{1}{\sqrt{s}} \psi\left(\frac{t - \tau}{s}\right) dt. \quad (1)$$

CWT defines the convolution between signal $x(t)$ and a filter whose impulse response is $\psi(-t/s)/\sqrt{s}$ and can therefore be interpreted as a linear filtering operation of $x(t)$.

For computational purposes a discrete form of wavelet transform is useful. The scaling and translation parameters are discretized typically using dyadic sampling, $s = 2^{-j}$ and $\tau = k2^{-j}$, where j and k are integers, and the discretized wavelet function is then

$$\psi_{j,k}(t) = 2^{j/2} \psi(2^j t - k). \quad (2)$$

From equations 1 and 2 we obtain discrete wavelet transform (DWT)

$$w_{j,k} = \int_{-\infty}^{\infty} x(t) \psi_{j,k}(t) dt. \quad (3)$$

For computing the features *IF* and *LOGPOW*, first stationary wavelet transform was applied. The mother wavelet used was Daubechies-5. Signal power and instantaneous frequency were calculated from wavelet approximations roughly corresponding to 0.5-16Hz. [36]

The feature *LOGPOW* is base-10 logarithm of signal power. The average of signal power over N samples is computed as

$$P[t] = \frac{1}{N} \sum_{k=t-N+1}^t x[k]^2. \quad (4)$$

For instantaneous frequency computation, Hilbert transform is applied. The introduction to Hilbert transform is from Johansson's work [38], where the signal processing application of the transform is kept in mind.

The definition of Hilbert transform $\hat{f}(t)$ of a function $f(t)$ is

$$\hat{f}(t) = \frac{1}{\pi} PV \int_{-\infty}^{\infty} \frac{f(\tau)}{t - \tau} d\tau, \quad (5)$$

where PV is the Cauchy principal value. An analytic signal can be created from a real signal with Hilbert transform. In the time domain the signal is

$$z(t) = f(t) + i\widehat{f}(t) = A(t)e^{i\varphi(t)}, \quad (6)$$

where $A(t)$ is an instantaneous amplitude and $\varphi(t)$ an instantaneous phase. In the polar notion

$$A(t) = \sqrt{f^2(t) + \widehat{f}^2(t)} \quad (7)$$

and

$$\varphi(t) = \arctan\left(\frac{\widehat{f}(t)}{f(t)}\right). \quad (8)$$

Instantaneous (angular) frequency is then

$$\omega(t) = \frac{d\varphi(t)}{dt} \quad (9)$$

In this work, in addition to the two-dimensional $IF - LOGPOW$ space presented previously in Tanner's work, a new two-dimensional space, which consists of means of $IF[t]$ and $LOGPOW[t]$ computed in 60s window, i.e. \overline{IF}_{60s} and \overline{LOGPOW}_{60s} , is introduced. Let the space be called $\overline{IF - LOGPOW}_{60s}$ space. The sample frequency of IF and $LOGPOW$ is 1 Hz.

A seizure from a patient in the development data is presented in Fig. 7. The figure above is the seizure in $IF - LOGPOW$ -space and the figure below is the same seizure in $\overline{IF - LOGPOW}_{60s}$ -space. The figures show, that a seizure with evolution, i.e. a seizure with increase first in frequency and then in amplitude, can be seen as a loop in the feature spaces. In the $\overline{IF - LOGPOW}_{60s}$ -space the loop is much smoother, due to the averaging of the samples.

Features computed from the two-dimensional EEG feature spaces aim to detect the loop-like behaviour in the spaces. In the beginning of the seizure, when frequency starts to increase, the distance between samples is expected to grow in IF direction.

Let $\vec{F}[t]$ be a vector that has $IF[t]$ and $LOGPOW[t]$ as its components. The difference over m samples is defined as

$$\Delta\vec{F}_m[t] = \vec{F}[t] - \vec{F}[t - m]. \quad (10)$$

From difference we can define the distance between two points

$$d_m = \|\Delta\vec{F}_m[t]\|_2. \quad (11)$$

For the distance between start and end points of a window of N samples $m = N$. Path length (PL) of steps over m samples in a window N is defined as

$$PL_N[t] = \sum_{n=t-N+m}^{n=t} \|\Delta\vec{F}_m[n]\|_2. \quad (12)$$

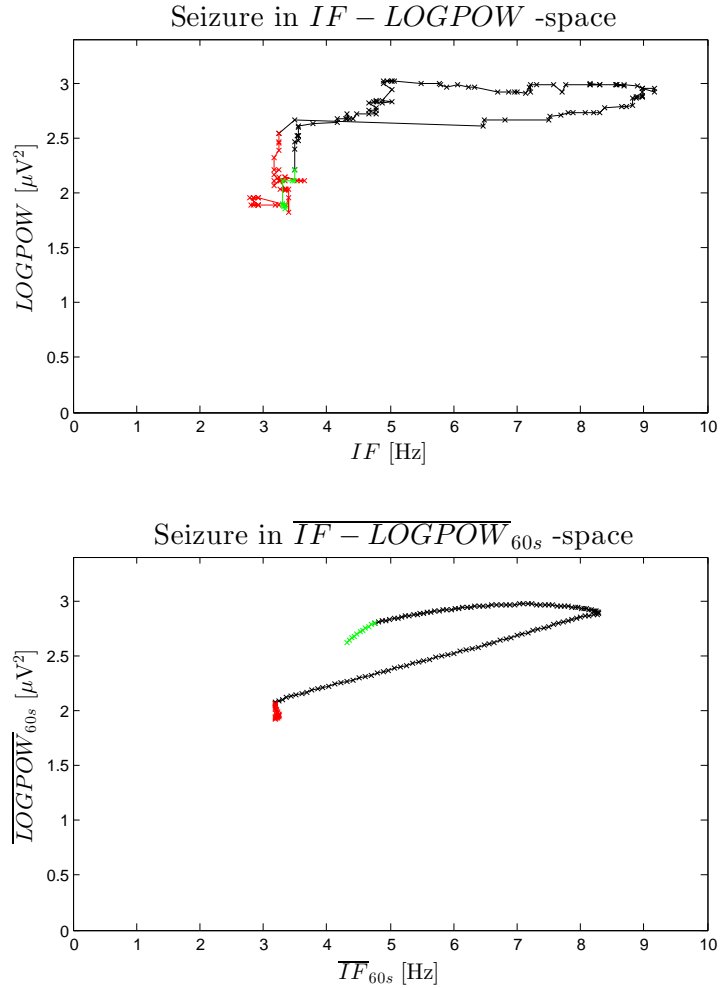


Figure 7: An example seizure in $IF - LOGPOW$ -space and in $\overline{IF} - \overline{LOGPOW}_{60s}$ -space. Red points are the preceding 60s and green points the 10s after seizure ended.

Features $\Delta\vec{F}$, PL , and d were introduced in Tanner's work.

The following features in this section were developed in this work, excluding angles. The standard deviation of steps over m samples in a time window of N samples is computed by

$$std[t] = \left(\frac{1}{N-m} \sum_{n=t-N+m+1}^{n=t} \left(\|\Delta\vec{F}_m[n]\|_2 - \overline{\|\Delta\vec{F}_m\|_2} \right)^2 \right)^{\frac{1}{2}}. \quad (13)$$

Moving to seizure zone from background zone in the feature space was expected to have a certain direction. This was investigated by analyzing the angles between

IF -axis and a vector from one data point to another. The angles are computed from differences over m samples between data points in $LOGPOW$ direction and IF direction as follows

$$\theta_m[t] = \arctan \left(\frac{LOGPOW[t] - LOGPOW[t - m]}{IF[t] - IF[t - m]} \right) \quad (14)$$

The 360 degrees range was divided into 12 equal sectors starting from 0 degrees. The sample distance m was determined to 5 samples when calculating the angles to be divided in the sectors. The mode was calculated in a window N and number of hits k in the mode i . The number of hits k_i were used for computing the probability of the angle θ_m in a certain sector

$$p_n^{angle} = \frac{k_i}{N}, \quad \begin{cases} n = t - N + m, \dots, t - 1, t \\ i = 1, 2, \dots, 12 \end{cases} \quad (15)$$

The portion of the step from the total path length is computed by

$$p_n^{step} = \frac{\|\Delta \vec{F}_m[n]\|_2}{PL_N[t]}, \quad n = t - N + m, \dots, t - 1, t. \quad (16)$$

p_n^{angle} and p_n^{step} are needed when calculating entropy for angles and steps. Define normalized entropy as

$$H_N[t] = \frac{-\sum_{n=t-N+m}^{n=t} p_n \ln p_n}{\ln N}. \quad (17)$$

Entropy $H_N[t]$ was calculated both for angles and for steps, i.e. distances between samples that are 5 samples from each other.

In addition, the angle θ_N between IF -axis and the vector from the first sample in a window N to the last sample in the window was computed by equation 14 setting $m = N$.

The angle between the first sample and the last sample itself does not give sufficient information to distinguish the direction of evolution from background EEG. As we see in Fig. 8, there are proportionally more angles in sector 0–45° in seizure data than in reference data. The number of angles that fell in sector 0–45° in a window N are computed by equation 18. N remains the same for the calculation of the angle between the first and last sample and for the number of hits in the sector, i.e. if the angle is calculated between samples 60s apart, the number of hits in the sector is calculated in 60s window.

$$K_{\theta_N} = \sum_{n=t-N+1}^{n=t} k[n], \quad k[t] = \begin{cases} 1, & \text{when } 0^\circ < \theta_N[t] < 45^\circ \\ 0, & \text{otherwise} \end{cases} \quad (18)$$

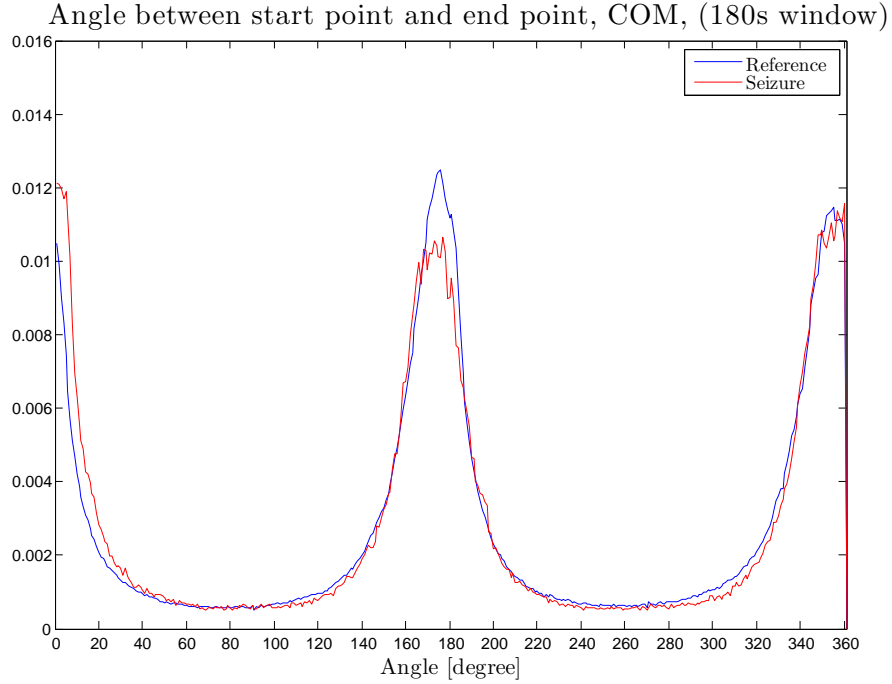


Figure 8: Histogram of angles between first and last sample in a 180s window in $\overline{IF - LOGPOW}_{60s}$ space

4.2.2 Principal component analysis

In this work the principal component analysis (PCA) was applied to the two-dimensional feature spaces. In PCA a set of measurement data is transformed into a lower dimensional feature space. The new variables are linear combinations of the original variables and are called principal components. Introduction to PCA here is mainly based on book Classification, Parameter Estimation and State Estimation by van der Heijden et al. [39]

Let \vec{z} be a vector in our measurement space. Without loss of generality, we can assume that \bar{z} , which is the expectation of the random vector \vec{z} , is equal to zero. The transformation from M -dimensional measurement space to D -dimensional feature space happens by means of operation

$$\vec{y} = W_D \vec{z}. \quad (19)$$

W_D is selected such that a minimum mean square error is yielded when an unbiased linear MMSE estimate \hat{z}_{MMSE} for \vec{z} is based on \vec{y} :

$$W_D = \underset{W}{\operatorname{argmin}} \left\{ \mathbb{E} \left[\|\hat{z}_{MMSE}(\vec{y}) - \vec{z}\|^2 \right] \right\} \quad \text{with} \quad \vec{y} = W \vec{z}. \quad (20)$$

For uniqueness, the information carried in the individual elements of \vec{y} must add up individually and the elements of y must be uncorrelated. Hence, the covariance

matrix $C_{\vec{y}}$ of \vec{y} is a diagonal matrix. Let Λ be the diagonal matrix and $C_{\vec{z}}$ the covariance matrix of \vec{z} , then

$$C_{\vec{y}} = W_D C_{\vec{z}} W_D^T = \Lambda_D. \quad (21)$$

In our case, the measurement space is two-dimensional and the transformation is made to two-dimensional space, hence $D=M$. From that follows, that $C_{\vec{z}} W_M^T = W_M^T \Lambda_M$. Λ_M is diagonal matrix, hence its diagonal elements are the corresponding eigenvalues of eigenvectors of $C_{\vec{z}}$. The eigenvectors of $C_{\vec{z}}$ are columns of W_D^T . The corresponding eigenvalues are the variance components of \vec{y} along the eigenvectors [40].

An additional requirement for the solution to be unique is that columns of W_D^T has to have unit length. This is fulfilled when $W_D W_D^T = I$, I being the $M \times M$ unit matrix.

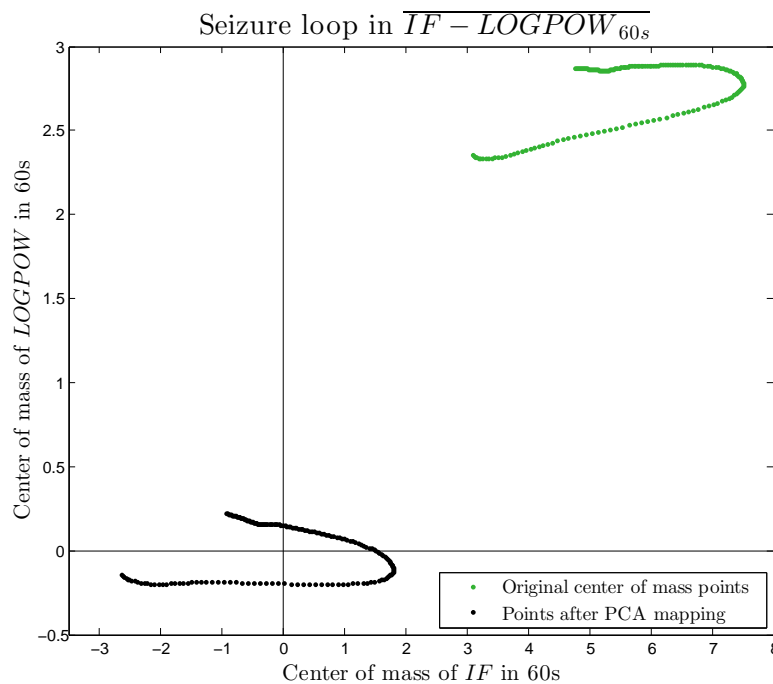


Figure 9: Green points demonstrate centers of mass calculated in a 60 second window. The data is taken from the beginning of a seizure from a case in development data set in derivation $C_3 - P_3$. Black points are transformed points in PCA space that have mean in origo.

In Fig. 9 green data points represent the original points in $\overline{IF - LOGPOW}_{60s}$ space. Points are 180 second-by-second values from the beginning of a seizure in channel $C_3 - P_3$ in a case from development data set. Black points are the values transformed with PCA, centered in origo, and aligned with principal axes.

PCA maps the values to another two-dimensional space defined by two principal components that are orthogonal. Variance is largest in the direction of the first

component. The direction of the greatest variance is in our interest.

The directions of the principal components are given as coefficients in directions of IF - and $LOGPOW$ -axis. Let the coefficients be noted as C_{IF} and C_{LOGPOW} in respective directions. The angle between the first PCA component and IF -axis is computed as

$$\theta^{PCA}[t] = \arctan \left(\frac{C_{LOGPOW}^{first}[t]}{C_{IF}^{first}[t]} \right) \quad (22)$$

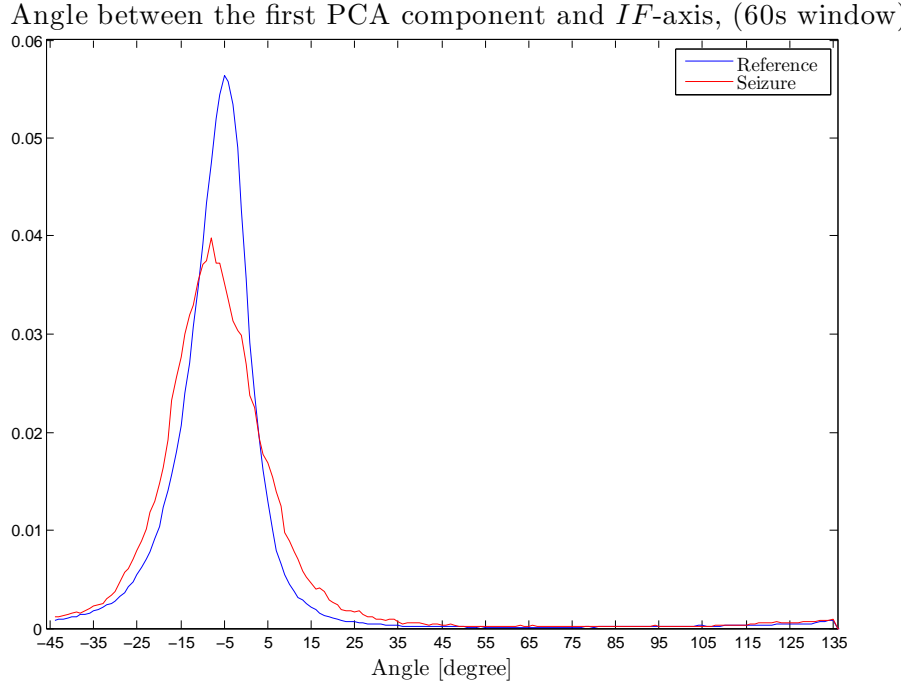


Figure 10: Histogram of angles between first PCA component and IF -axis in $\overline{IF - LOGPOW}_{60s}$ space, when PCA is computed in 60s window

In the previous subsection the number of angles that fell in sector $0-45^\circ$ in a window N was computed. In the same manner, the number of values of θ^{PCA} that fall in a certain sector can be computed. N is the duration of the window in which PCA was computed and the window, in which we search angle values that fall in the sector. Fig. 10 shows that approximately in sector $-15-5^\circ$ the values of reference data are proportionally more than in seizure data and in sector $10-45^\circ$ the values are proportionally less than in seizure data. Therefore, the sectors are chosen to be $-15-5^\circ$ and $10-45^\circ$ and the computation is defined as

$$K_{\theta^{PCA}} = \sum_{n=t-N+1}^{n=t} k[n], \quad k[t] = \begin{cases} 1, & \text{when } -15^\circ < \theta^{PCA}[t] < 5^\circ \\ 0, & \text{otherwise} \end{cases} \quad (23)$$

$$K_{\theta_2^{PCA}} = \sum_{n=t-N+1}^{n=t} k[n], \quad k[t] = \begin{cases} 1, & \text{when } 10^\circ < \theta^{PCA}[t] < 45^\circ \\ 0, & \text{otherwise} \end{cases} \quad (24)$$

Let λ_i be the variance component of \vec{y} corresponding to eigenvector \vec{e}_i . The component that explains the greatest part of the variance is λ_1 , which is the variance in direction of the first principal component. The proportion of the total variance which is explained by the component λ_1 can be computed as

$$\text{var}_\% = \frac{\lambda_1}{\sum_{i=1}^{i=M} \lambda_i}. \quad (25)$$

4.2.3 Imaginary space

As seen in the previous chapter, with PCA it is possible to transform the data points to PCA space and place the mean in origo. The principal components are aligned with principal axes. In this work we choose the direction of the first component as real axis and the direction of the second component as imaginary axis, i.e., we are now in the complex plane and our data points can be treated as complex numbers.

Let φ be the phase of the complex number $x_{PCA} + iy_{PCA}$. Phase shift in a window N is

$$\Delta\varphi_N[t] = \varphi[t] - \varphi[t - N] \quad (26)$$

We are also interested in how phase evolves in a time window N . Phase increment describes how many times the phase increases in that window.

Phase increment was computed as

$$\text{Incr}_{\varphi_N}[t] = \sum_{n=t-N+1}^{n=t} \text{incr}[n], \quad \text{incr}[t] = \begin{cases} 1, & \text{when } \Delta\varphi_N[t] > 0 \\ 0, & \text{otherwise} \end{cases} \quad (27)$$

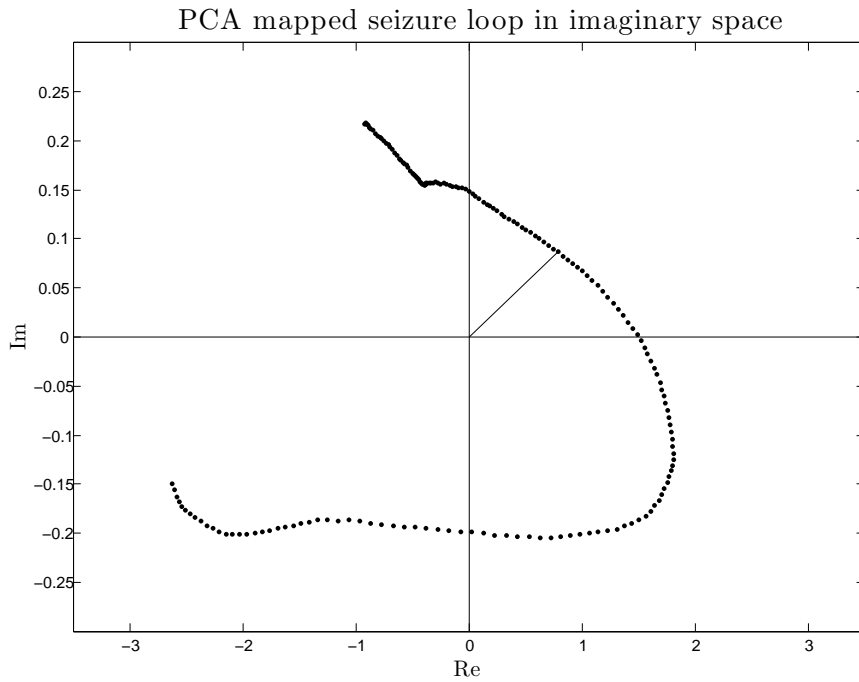


Figure 11: Points in PCA space transformed to complex space

4.2.4 Random walk

In the random walk algorithm, which was developed as a part of A. Tanner's Master's thesis work [36], the idea is that the feature tracing steps follow a path of a random

walker. The steps are expected to be drawn randomly from an autoregressive process. When a seizure starts, the steps are no longer drawn randomly and the process becomes organized. The random walk algorithm aims to quantify this change from random to organized.

The average step, when we have a history of N seconds, is

$$\vec{a}[t] = \frac{\sum_{n=t-N}^t |\Delta \vec{F}_m[t]|}{N} \quad (28)$$

A step of the random walker is the difference between the latest point and the average value in a block size b N seconds before that point.

$$\vec{F}_N[t] = \frac{1}{b} \sum_{k=1}^b \Delta \vec{F}_{N-k}[t] \quad (29)$$

RWS is an index that describes the amount of steps taken by the random walker in both x and y directions. Subscripts 1 and 2 referring to the components of the vector, i.e. $x[t]$ and $y[t]$, RWS is defined

$$RWS = \left\| \left(\frac{\vec{F}_{N,1}[t]}{a_1}, \frac{\vec{F}_{N,2}[t]}{a_2} \right) \right\|_2 \quad (30)$$

When the step is not part of the background activity, the value of RWS is large, whereas a very small value indicates that step does not stand out from the background activity.

4.2.5 Convex hull

Convex hull is an algorithm developed by A. Tanner [36] which works in the two dimensional space described in chapter 4.2.1. The idea of the detector is to distinguish changes in the feature vectors from the background data that indicate evolution of an seizure.

The algorithm uses the concept of convex hull to establish the boundaries of the background activity. Let $X = \{\vec{x}_1, \vec{x}_2, \dots, \vec{x}_k | \vec{x}_{1..k} \in \mathbb{R}^n, k \in \mathbb{N}\}$. The convex hull of set X can be defined

$$\text{conv}(X) = \left\{ \sum_{i=1}^k a_i \vec{x}_i | \vec{x}_i \in X, a_i \in \mathbb{R}^+, \sum_{i=1}^k a_i = 1, i \in \mathbb{N} \right\}. \quad (31)$$

In the algorithm of Tanner *quickhull* algorithm is used for the calculation of the hull.

When the feature vectors breach the hull, an elementary detection is made. After this, if the elementary detection has a certain minimum duration t_{min} and the breach has happened to the right direction, i.e. in a valid sector, a number of tests are made.

The tests are used to give a score to the breach. Breach scores are calculated for all channels and these scores are then combined and filtered. That gives us the final output of the evolution detector. The evolution detector output is compared to a predefined threshold to validate, if the patient has a seizure or not.

4.2.6 Spike detector

Spike detector was developed outside this thesis work. Detector uses EEG derivations where the electrodes 9 electrodes mentioned in the section 4.1 were referenced to their average. Bandstop filters with stop bands at 48-52 Hz and 58-62 Hz were used to remove 50 Hz and 60 Hz. The signal was then high-pass filtered with a cut-off frequency of 1 Hz. After computing the derivations and filtering, the following steps are performed for every channel separately.

At first, the spike detector searches extrema in the signal. The local maxima and local minima, i.e. all the samples that have either highest or lowest voltage, are searched in a neighborhood of predefined duration centered in the sample. From consecutive local maxima only the one with the highest voltage and from consecutive local minima the one with the lowest voltage are left and others are removed.

The pairs of consecutive local maximum and local minimum describe a halfwave and the voltage difference of maximum and minimum describes the height of the wave. A wave is formed by two consecutive halfwaves.

A set of features is extracted from the wave and from halfwaves that form the wave. The first step excludes all the waves with feature values that exceed set limits and assumes that those waves must not be spikes.

The second step uses a statistical model to classify the waves. The model assumes features to be normally distributed for spikes and for non-spikes. In assumption some features are correlated and covariance between these features is assumed to be zero. For the rest of the features covariances are estimated from the training data. For a feature vector of a wave, likelihood is computed for both classes using the normal distributions.

The third step is to compute a posterior probability for the spike by applying Bayes's theorem. A wave is classified as a spike if both spike probability and spike likelihood are above thresholds.

Spike detections are combined from different channels into one detection signal. This is made by marking only one detection, when the detections in different channels occur in a time window. Spike rate is computed from the one detection signal, and is the number of spikes in a one minute time window.

4.2.7 Spectral features

The power spectrum of a stationary signal $x(n)$ is defined by [37]

$$S_x(e^{j\omega}) = \sum_{k=-\infty}^{\infty} r_x(k)e^{j\omega k}, \quad (32)$$

where $r_x(k)$ is the correlation function that has to be estimated from $x(n)$ before power spectrum can be computed. $r_x(k)$ can be estimated with an time average estimator

$$\hat{r}_x(k) = \frac{1}{N} \sum_{n=0}^{N-1-k} x(n+k)x(n), \quad k = 0, \dots, N-1, \quad (33)$$

Inserting the correlation function estimate to eq. 32, an estimate of power spectrum and can be obtained by periodogram as

$$\hat{S}_x(e^{j\omega}) = \sum_{k=-N+1}^{N-1} \hat{r}_x(k)e^{j\omega k}. \quad (34)$$

Spectral features for feature selection were computed from preprocessed EEG data in bands that are listed in Table 6. Periodogram was performed from sliding window of 1024 EEG samples, thus frequency resolution of approximately 0.2 Hz was obtained. 1024 samples in EEG with sampling frequency of 200 Hz give 5.12 seconds of signal. Sliding window was moving second-by-second and periodogram was averaged over 5 sliding windows that passed the current sample.

In addition to absolute band powers, band powers were normalized with 0.2–100 Hz band power and 0.2–30 Hz band power. 0.2–30 Hz will be from this on called EEG band.

Table 6: Spectral bands in band power computation

Band	Frequency range [Hz]
Delta (δ)	0.2–3.4
Theta (θ)	3.6–8.0
Alpha (α)	8.2–13.0
Lower beta (β_1)	13.2–20.0
Higher beta (β_2)	20.2–30.0
Total beta (β)	13.2–30.0
Gamma (γ)	30.2–70.0
EMG	70.2–100.0
Total	0.2–100.0

4.2.8 Overview

The majority of features for the method development were calculated in the two-dimensional spaces defined in section 4.2.1: $IF - LOGPOW$ space and $\overline{IF - LOGPOW}_{60s}$ space. IF and $LOGPOW$ values were extracted from 16 derivations introduced in section 4.1. Feature values used for the method development are from one derivation per patient as explained in sections 3.1.1 and 3.1.2.

Data contaminated with artifacts was removed with methods presented in chapter 4.1.1 before the feature computation. If less than 10% of data in the window was removed due to artifacts, feature was calculated in the window. Otherwise feature calculation for the current sample was not performed.

In $IF - LOGPOW$ space the time windows for feature calculations were fixed at 30s and 60s. In $\overline{IF - LOGPOW}_{60s}$ space the time windows were 60s and 180s. Table 7 summarizes the features calculated from the selected derivation in these

windows. For example, feature center of mass IF was calculated in windows 30s and 60s in $IF - LOGPOW$ space and center of mass \overline{IF}_{60s} in windows 60s and 180s in $\overline{IF} - \overline{LOGPOW}_{60s}$ space.

Table 7: EEG features calculated from one derivation in $IF - LOGPOW$ space in 30s and 60s windows and in $\overline{IF} - \overline{LOGPOW}_{60s}$ space in 60s and 180s windows

	Feature
Distance	Center of mass IF/\overline{IF}_{60s}
	Center of mass $LOGPOW/\overline{LOGPOW}_{60s}$
	Path length (eq. 12)
	Standard deviation of step lengths (eq. 13)
	Step entropy (eq. 16 and 17)
	Distance from start point to end point (eq. 11)
Angle	Evolution angles in sector 0–45° (eq. 18)
	Number of hits in mode
	Angle entropy (eq. 15 and 17)
PCA	Angle of first PCA component (eq. 22)
	Number of PCA angles in sector -15–5° (eq. 23)
	Number of PCA angles in sector 10–45° (eq. 24)
	Variance of the first PCA component aligned with corresponding eigenvector
	Explained variance (eq. 25)
	Phase shift (eq. 26)
	Phase increment (eq. 27)

Normalized steps of random walker were as well calculated from one derivation. In random walker, the length of history for computing average step was set to 180s. Block sizes were chosen to 10s, when window was 30s, and 20s when window was 60s. Normalized steps were computed with these parameters in both $IF - LOGPOW$ and $\overline{IF} - \overline{LOGPOW}_{60s}$ spaces.

In addition to the features listed in Table 7 and random walker steps, outputs of spike detector and evolution detector were taken into the feature selection. Spike detector gives only one output combined from all the derivations and convex hull detector was modified to give an output for every channel.

The output of spike detector is spikes/second. For feature selection the output was summed over 60s. The output of convex hull evolution detector was computed with 60s and 180s history data.

In addition to the spike rate and the features from two-dimensional spaces all the absolute and normalized values of band powers were included to the feature selection.

4.3 Evaluation of classifiers

4.3.1 Kappa coefficient

A measure of agreement, kappa (κ), can be used to compare the ability of two raters to classify data into one of several different groups. In computation of kappa, observed frequencies, i.e. the proportion of exact agreements, and expected frequencies, i.e. the proportion of agreements by chance, are taken into account. [41]

In seizure detection there are two groups: seizure and no seizure. The benefit of kappa in analysis of agreement in seizure classification is that the observed and expected frequencies include agreement in both groups. This way, the skewness in classification distribution is taken into account. In seizure classification the amount of non-seizure data is usually much larger than the amount of seizure data.

When n is the number of observations and g , the number of categories the observed proportional agreement is

$$p_o = \sum_{i=1}^g \frac{f_{ii}}{n}, \quad (35)$$

where f_{ii} is the number of agreements for category i . When r_i is the number of samples in category i according to rater 1 and c_i number of samples in category i according to rater 2, the expected proportion of agreements by chance is

$$p_e = \sum_{i=1}^g \frac{r_i c_i}{n^2} \quad (36)$$

Kappa coefficient is

$$\kappa = \frac{p_o - p_e}{1 - p_e}. \quad (37)$$

In Table 8 it is listed how to interpret the value of kappa. [41]

Value of κ	Strength of agreement
< 0.2	Poor
0.21–0.40	Fair
0.41–0.60	Moderate
0.61–0.80	Good
0.81–1.00	Very good

4.3.2 Correlation coefficient

The degree of association can be measured with correlation coefficient, also called simply correlation. Pearson's correlation coefficient is a quantity r that measures so

called straight-line association between values of the two variables and can take any value in a range from -1 to +1. Pearson's correlation coefficient is computed as

$$r(X, Y) = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}} \quad (38)$$

where x_i and y_i are the i th samples of X and Y [41].

4.3.3 Binary classifier

Receiver operating characteristic (ROC) curve is a useful tool for evaluating the performance of a binary classifier. The trade-off between sensitivity and specificity, i.e. the rate of true positives identified as such, and the rate of true negatives identified as such, can be evaluated from the curve. ROC curve is a plot of sensitivity against 1-specificity. Sensitivity and specificity are defined as

$$\text{Sensitivity} = \frac{\text{true positives}}{\text{true positives} + \text{false negatives}} \quad (39)$$

$$\text{Specificity} = \frac{\text{true negatives}}{\text{true negatives} + \text{false positives}} \quad (40)$$

In our case true positives are the samples correctly identified as samples from a seizure period and true negatives are the samples correctly identified as samples from a non-seizure period. False positives are the samples from a non-seizure period but identified as a seizure period, whereas false negatives are the samples from a seizure period identified as a non-seizure period.

The area under the ROC curve (AUC) is a value that estimates the probability that a randomly chosen sample from one population has a greater value than a sample from the other population. [42]

The ROC curves may give an over-optimistic view of the performance of an algorithm, if there is a large skew in the class distribution [43], which is often the case in seizure detection. Usually the total duration of data without seizures is much longer than the total duration of seizure periods. An alternative to sensitivity-specificity curves when there is a large skew in the class distribution are precision-recall curves [43], which are often used in information retrieval [44].

Recall is the same as sensitivity, and defined in eq. 39. Precision tells the proportion of the positives that are truly positive. In seizure detection, that is the proportion of the correct detections from all the detections. Precision can be computed as [43]

$$\text{Precision} = \frac{\text{true positives}}{\text{true positives} + \text{false positives}}. \quad (41)$$

Any-overlap and overlap-integral methods can be used for sensitivity and specificity computation when seizure markings of two experts or detections of an algorithm and markings of an expert are compared. In the any-overlap method the metrics used for the comparison are any-overlap sensitivity and false positive rate

(FPR). In the overlap-integral method the metrics are overlap-integral sensitivity and overlap-integral specificity. [18]

In the any-overlap sensitivity, if the markings overlap at any part, they are counted as a match. Sensitivity of reader X with respect to reader Y is given by the number of overlapped markings divided by the number of seizure markings made by reader Y. [18] FPR is the number of events marked by reader X that do not overlap with markings of reader Y divided by the record duration in hours.

In the overlap-integral method seizures are not described as discrete events. EEG recording is described as a seizure density function which varies between 0 and 1 each second over the duration of the record. During seizure the value of the density function is 1 and otherwise it is 0. Sensitivity and specificity are computed using the seizure density functions of the two readers. [18]

4.4 Feature combination

One way of combining different variables is the multiple linear regression model. In the multiple linear regression model, a combination of the explanatory variables, which are in this case the selected features, express the dependent variable.

The outcome variable of interest in many studies is the presence or absence of some condition. In our case, information of presence or absence of seizures is desired. For such data, ordinary multiple linear regression cannot be used. Instead, we can use multiple linear logistic regression for forming a prognostic index, which is in our case a seizure probability index. The difference between multiple linear regression and multiple linear logistic regression is, that the linear regression model predicts the value of the dependent variable, whereas the linear logistic regression model predicts the transformation of the dependent variable. [41]

The transformation is called the logit transformation, $\text{logit}(p_{sz})$. p_{sz} is the seizure probability and $1 - p_{sz}$ the probability that the patient is not having a seizure. Their ratio $p_{sz}/(1 - p_{sz})$ is called the odds. Thus,

$$\text{logit}(p_{sz}) = \log_e \left(\frac{p_{sz}}{1 - p_{sz}} \right). \quad (42)$$

Let b be the coefficient of feature x . Define L as logit of the probability p_{sz} that in moment t patient is having a seizure, then

$$L[t] = \log \left(\frac{p_{sz}[t]}{1 - p_{sz}[t]} \right) = b_0 + b_1 x_1[t] + b_2 x_2[t] + \dots + b_k x_k[t], \quad (43)$$

where k is the number of the features. We can resolve seizure probability from the logit function as

$$p_{sz}[t] = \frac{e^{L[t]}}{1 + e^{L[t]}}. \quad (44)$$

4.5 Feature selection

4.5.1 Data selection criteria

The feature values were computed for every second, which produces a large amount of data. In addition, the features are computed in time windows up to 30s. As a consequence, the change in features second by second is not large. Therefore, preprocessing of the data set was necessary before the feature selection.

Reference and seizure data sets were selected differently. For the seizure data set, it was important to include data only from seizure periods. Criteria that the seizure data set had to fulfill is listed in the Table 9.

Table 9: Criteria for seizure data set

a)	Data of all patients have the same weight
b)	Data is from patients that have unequivocal seizures marked by both experts
c)	Data is from a seizure period that both experts have marked as a seizure

The reference data set was reduced by dividing data to 30 second periods and calculating maximum for each period. This was done for every patient and for every feature. The values of a certain feature from all the patients formed a feature vector for that feature. For example, the feature vector of spike rate included the 30-second maximas of spike rates from all the reference patients.

For the seizure data, the first 10 minutes of the seizure were taken into account if the seizure duration exceeded 10 minutes. Then, the data were divided to 30-second periods and maxima of periods were calculated. If seizure duration was less than 10 minutes, the seizure period was extended by 10 seconds to include post-ictal suppression. All the values that were NaN (not a number) due to artifacts were removed. These steps were done for every patient and for every feature. The feature vectors were formed by first removing the invalid values and then replicating the remained values in order to have an equal amount of values per patient.

4.5.2 Sequential floating forward search

The formed feature vectors, described in the previous section, were used for classification. Sequential backward and forward selection methods and floating backward and forward selection methods are techniques that measure classification capabilities of feature vectors. In this work the selected method is the sequential floating forward selection (SFFS). Introduction here to the method is presented in the book Pattern Recognition by Theodoridis and Koutroumbas [45].

In forward methods, the selection procedure starts by computing criterion value for each of the features and feature with the best criterion value is chosen. Next, a feature, that in combination with the first feature produces the best criterion value, is searched, and the two features form a subset.

In the following step, a feature, which in combination with the existing subset of features produces best criterion value, is added. In floating forward search, the feature, that has least effect to the criterion value when it is removed from the formed subset, is searched. If this is the same feature which was just added, it will be included to the subset and selection procedure continues to the selection of next feature to include.

If the feature that has least effect to the criterion value is not the one which was just added, and the criterion value after its removal from the subset is poorer than the criterion value before the newest feature was added to the subset, backward search is terminated. The selection procedure continues by searching the next feature to add to the subset.

In cases, when the criterion value improves after removal of the least significant feature but having the newest feature in the subset, the feature that adds least effect to the criterion value is removed. The newest feature is then added to the subset. From this set, the least significant feature is again searched. It is checked whether the removal of the least significant feature improves the criterion value and in that case the feature is removed and backward search continues as before. If there is no improvement, backward search is stopped and feature selection continues by searching the best feature to add to the combination. If in backward search subset is reduced to two features, backward search is not performed further.

Floating forward search continues until the wanted number of features for the subset is reached.

5 Results

5.1 Comparison of two experts

In this study, the seizure data set presented in section 3.1.2 was analyzed by two experts, who marked the seizure start points and end points, and the seizures as unequivocal or equivocal. Unequivocal seizures that overlapped in the marking of experts were found in 24 patients. The agreed unequivocal seizure periods of EEG were used for feature selection. The number of seizure patients, the number of seizures, and median and total durations of unequivocal seizures are listed in Table 10.

Table 10: Seizure durations of unequivocal seizures reviewed by two experts

	N of seizure patients	N of seizures	Median (range) [mm:ss] ([hh:mm:ss])	Total [hh:mm:ss]
Reader 1	32	817	01:16 (00:00:08-09:00:18)	42:06:11
Reader 2	33	1578	00:48 (00:00:07-18:10:11)	72:30:28
Agreed	24	698	01:16 (00:07-11:52:00)	20:29:55

For the evaluation of the rate of agreement, the metrics any-overlap sensitivity, overlap-integral sensitivity, FPR, and kappa were computed for 50 patients of the seizure data set. The metrics were computed separately for unequivocal seizures and for all seizures, i.e. the type of the seizure was not considered. Median and mean values of the metrics are listed in Table 11. Median kappa was 0.44 for unequivocal seizures, which is in the range of moderate agreement, and for all seizures median kappa was 0.38, which means poor agreement.

Sensitivities for unequivocal seizures were slightly better for reader 1 referenced to reader 2 than for reader 2 referenced to reader 1. Median any-overlap sensitivities were 0.79 and 0.71, respectively, and median overlap-integral sensitivities were 0.70 and 0.51. In all seizures the difference between median sensitivities, when reference reader changes, is more notable. Median sensitivities varied in any-overlap from 0.79 to 0.65 and in overlap-integral from 0.70 to 0.52, when reference reader changed from 2 to 1.

Results for unequivocal seizures presenting all patients are listed in Table 12 and for all seizures in Table 13. Tables show how the number of patients is distributed according to kappa and FPR/any-overlap sensitivity/overlap-integral sensitivity.

For unequivocal seizures the agreement according to kappa was 0.8–1 for 16 patients. This means, that for 16 patients there was a very good agreement on seizure markings. In this patient group, there were 13 patients that had FPR zero when reader 2 was referenced to reader 1 and 11 patients that had FPR zero when reader 1 was referenced to reader 2. In any-overlap sensitivity 7 patients had sensitivity between 0.8 and 1 disregarding whether seizures were marked by reader 1 or 2. In

overlap-integral sensitivity the number of patients with sensitivity from 0.8 to 1 was 7 for reader 1 and 6 for reader 2.

Any-overlap and overlap-integral sensitivities could not be calculated when reader which is the reference had not made any seizure markings. These cases are in the tables in section NaN. When kappa is below 0.2 and sensitivity is NaN it means that the reader has marked seizures when the reference reader has not. When kappa score is 0.8–1.0 and sensitivities are NaN, both of the readers agree that the patient does not any have seizures.

When looking unequivocal seizures with kappa between 0.8–1, 9 patients had any-overlap or overlap-integral sensitivity in category NaN, i.e. did not have seizures according to both readers, disregarding who was the reader. When agreement according to kappa was very good, the sensitivity was either high or readers agreed that patient did not have any seizures.

For all seizures the agreement according to kappa was very good for 8 patients. This is half of the number of patients with very good agreement when only unequivocal seizures were considered. 5 of the patients with very good agreement for all seizures had FPR zero when reader 2 was referenced to reader 1 and only one patient had FPR zero when reader 1 was referenced to reader 2. The any-overlap and overlap-integral sensitivities were from 0.8–1 for 7 patients when the reference was reader 2 and for 6 patients when the reference reader was 1.

Total seizure durations of every patient against readers are presented in Fig. 12 for unequivocal seizures. There are seven patients for which the difference in seizure marking durations is more than an hour. There is one case in which reader 1 has marked over 10 hours of seizures and reader 2 has marked none. There are, on the other hand, cases in which reader 2 has marked approximately 8.5 hours and 33 hours of seizures and reader 1 has marked none. Correlation coefficient for unequivocal seizure durations was 0.024 which indicates that there is no association.

Table 11: Median and mean kappa, FPR, and sensitivities for unequivocal and all seizures reviewed by two experts (reader 2 against reader 1/reader 1 against reader 2)

Unequivocal seizures				
	Kappa	FPR [h ⁻¹]	Any-overlap sensitivity	Overlap-integral sensitivity
Median	0.44	0/0.02	0.71/0.79	0.51/0.70
Mean	0.45	2/0.07	0.57/0.59	0.51/0.57

All seizures				
	Kappa	FPR [h ⁻¹]	Any-overlap sensitivity	Overlap-integral sensitivity
Median	0.38	0.04/0.10	0.65/0.79	0.52/0.70
Mean	0.40	0.23/1.81	0.79/0.70	0.81/0.67

Table 12: Inter-rater agreement results (reader 2 against reader 1/reader 1 against reader 2) for unequivocal seizures

		Kappa					
		all	≤ 0.2	0.2-0.4	0.4-0.6	0.6-0.8	0.8-1.0
FPR [h ⁻¹]	all	50	21	3	6	4	16
	>1	4/0	2/0	1/0	0/0	1/0	0/0
	0.20-1.00	6/5	6/0	0/0	0/3	0/0	0/2
	0.12-0.20	4/3	1/2	1/1	1/0	0/0	1/0
	0.04-0.12	3/12	1/7	0/0	0/1	0/3	2/1
	0.00-0.04	5/6	2/2	0/2	0/0	3/0	0/2
	0	28/24	9/10	1/0	5/2	0/1	13/11
Any- overlap sensi- tivity	NaN	18/17	9/8	0/0	0/0	0/0	9/9
	≤ 0.2	8/11	8/11	0/0	0/0	0/0	0/0
	0.2-0.4	5/0	2/0	1/0	2/0	0/0	0/0
	0.4-0.6	2/3	0/1	0/1	1/0	1/1	0/0
	0.6-0.8	3/3	0/0	0/1	1/1	2/1	0/0
	0.8-1.0	14/16	2/1	2/1	2/5	1/2	7/7
Overlap integral sensi- tivity	NaN	18/17	9/8	0/0	0/0	0/0	9/9
	≤ 0.2	10/10	9/10	1/0	0/0	0/0	0/0
	0.2-0.4	6/1	2/1	0/0	4/0	0/0	0/0
	0.4-0.6	0/5	0/0	0/3	0/0	0/2	0/0
	0.6-0.8	4/2	0/1	0/0	0/0	3/1	1/0
	0.8-1.0	12/15	1/1	2/0	2/6	1/1	6/7

In Fig. 13 are the total seizure durations for all seizures. When all seizures are considered the correlation of seizure duration markings is better. Correlation coefficient for all seizures was 0.41. There were 8 patients in which the difference in seizure durations were more than two hours. The most problematic patients are the ones when other reader marks several hours of seizures, and other reader does not mark any or just some minutes of seizures.

Table 13: Inter-rater agreement results (reader 2 against reader 1/reader 1 against reader 2) for all seizures

		Kappa					
		all	≤ 0.2	0.2-0.4	0.4-0.6	0.6-0.8	0.8-1.0
FPR [h ⁻¹]	all	50	16	10	11	5	8
	>1	3/7	0/2	2/2	1/1	0/0	0/2
	0.20-1.00	8/10	4/1	2/3	0/3	1/1	1/2
	0.12-0.20	8/6	3/4	1/2	1/0	2/0	1/0
	0.04-0.12	3/10	1/4	3/0	0/4	1/0	1/2
	0.00-0.04	5/10	1/5	0/2	4/1	0/1	0/1
	0	20/7	7/0	2/1	5/2	1/3	5/1
Any- overlap sensi- tivity	NaN	0/5	0/5	0/0	0/0	0/0	0/0
	≤ 0.2	10/8	9/7	1/1	0/0	0/0	0/0
	0.2-0.4	9/1	2/0	1/0	6/0	0/0	0/0
	0.4-0.6	5/3	1/1	3/1	1/1	0/0	0/0
	0.6-0.8	9/11	3/0	2/4	1/3	1/3	2/1
	0.8-1.0	17/22	1/2	3/4	3/7	4/2	6/7
Overlap integral sensi- tivity	NaN	0/5	0/5	0/0	0/0	0/0	0/0
	≤ 0.2	11/8	11/7	0/1	0/0	0/0	0/0
	0.2-0.4	10/3	2/2	4/1	4/0	0/0	0/0
	0.4-0.6	6/4	1/0	2/0	3/2	0/2	0/0
	0.6-0.8	7/6	0/0	3/3	0/1	2/1	2/1
	0.8-1.0	16/24	2/2	1/5	4/8	3/2	6/7

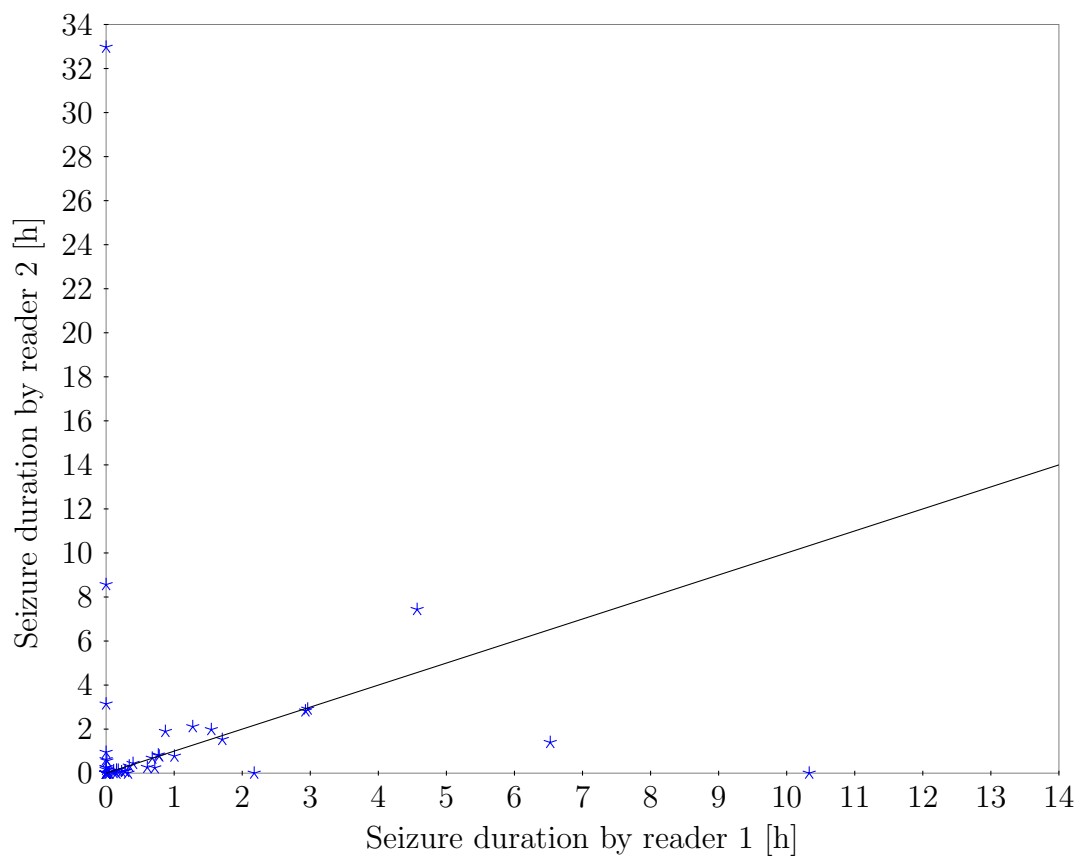


Figure 12: Correlation of unequivocal seizure durations marked by two experts. Black line is the correlation coefficient 1

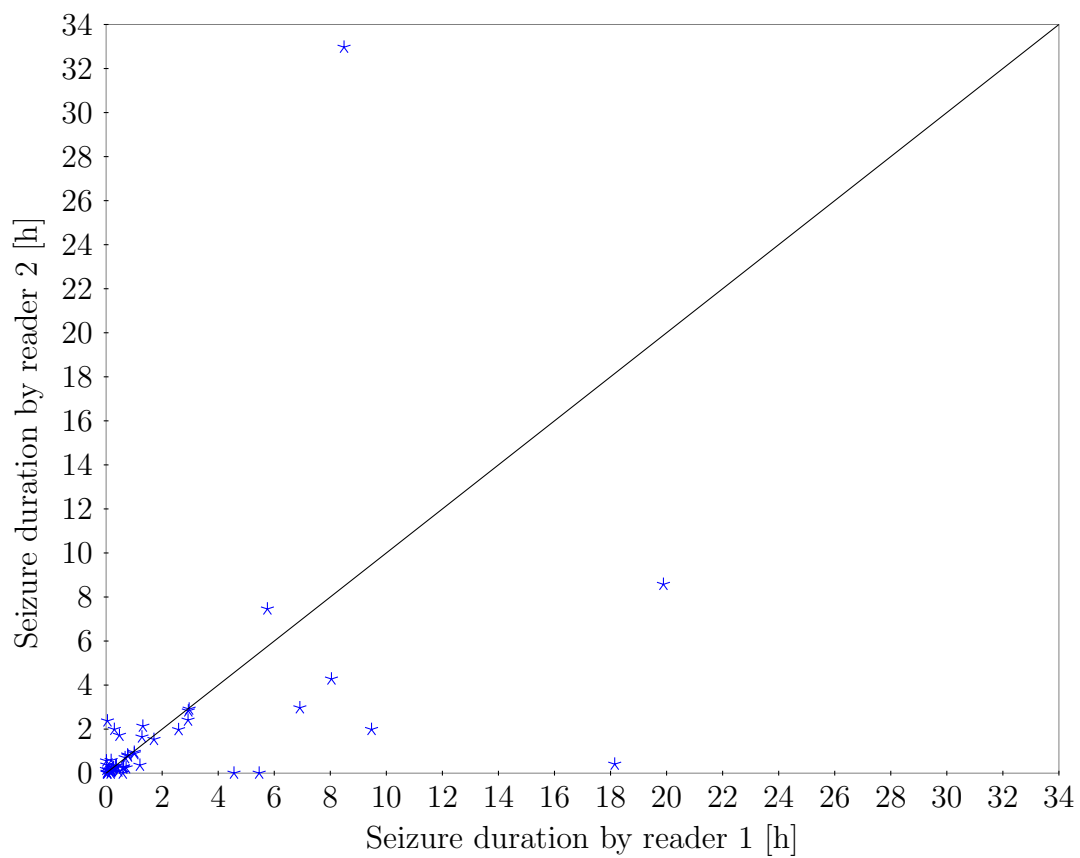


Figure 13: Correlation of all seizure durations marked by two experts. Black line is the correlation coefficient 1

5.2 Feature selection

5.2.1 Window selection

Feature selection was performed to find a subset of features that reaches the best performance in classifying seizure periods and non-seizure periods. The first step in feature selection was to evaluate whether it is important to select features from different time windows, and if not, select the optimal time window for feature computation. Data set for the window selection was a subset of the development data set and formed as explained in section 4.5.1. Feature selection was made with all the calculated features, so that all the window dependent features listed in Table 7 were calculated in all time windows and spaces. In addition, selection was made with reduced feature sets by selecting window dependent features only from one window at a time. Feature selection was computed with sequential floating forward search and the features were combined with multiple linear logistic regression model. The criterion value was AUC produced by the linear model. AUC values, when selection was made up to 8, are presented in Table 14.

Table 14: Comparison of AUC for combination of features calculated in different windows and feature spaces

Space	Window	AUC
all	all	0.9888
$IF - LOGPOW$	30s	0.9735
$IF - LOGPOW$	60s	0.9780
$\overline{IF - LOGPOW}_{60s}$	60s	0.9718
$\overline{IF - LOGPOW}_{60s}$	180s	0.9687

The classification performance based on AUC of feature selection was nearly as good for 60s-features in the $IF - LOGPOW$ -space as for all features. The difference of AUC in different windows is not statistically significant, but by selecting only one time window for feature computation, complexity of the algorithm is reduced. The algorithm is wanted to react rapidly when a seizure starts and therefore a short time window is favourable. In addition, spike rate is computed only from 60s-window. These motives lead to selecting the features from only 60s-window in $IF - LOGPOW$ -space for further selection.

5.2.2 Selection of candidate models

The feature selection for automatic detection methods was performed after the features in 60s-window in $IF - LOGPOW$ -space were selected from the total data set for further feature selection. In the first stage of feature selection, ROC and AUC were calculated for individual features. Features for the feature selection are listed in Table 15 ranked by their AUC value. The best AUC was obtained with spike rate and was 0.880.

Table 15: AUC values of independent features

Feature	AUC
Spike rate	0.880
Band power of alpha band normalized by total band	0.840
Absolute band power of alpha band	0.829
Random walker steps	0.821
Absolute band power of lower beta band	0.819
Band power of alpha band normalized by EEG band	0.816
Band power of total beta band normalized by total band	0.812
Band power of delta band normalized by EEG band	0.805
Evolution angle in sector 0–45°	0.805
Absolute band power of lower beta band	0.804
Center of mass IF	0.800
Absolute band power of higher beta band	0.783
Absolute band power of theta band	0.782
Band power of theta band normalized by total band	0.777
Band power of total beta band normalized by total band	0.774
Band power of lower beta band normalized by EEG band	0.774
Variance of the first PCA component	0.774
Band power of delta band normalized by total band	0.772
Band power of theta band normalized by EEG band	0.764
Distance between start point and end point	0.757
Number of hits in mode	0.745
Band power of total beta band normalized by EEG band	0.739
Explained variance	0.719
Band power of higher beta band normalized by total band	0.718
Angle entropy	0.710
PCA angle	0.703
Center of mass $LOGPOW$	0.700
Band power of higher beta band normalized by EEG band	0.695
Absolute band power 70.2–100 Hz	0.677
Absolute band power of gamma band	0.676
Absolute band power of the total band	0.659
Phase shift	0.655
Convex hull	0.653
Phase increment	0.652
PCA angle values in sector -15–5°	0.641
Band power of 70.2-100 Hz normalized by total band	0.623
Standard deviation of step lengths	0.614
Absolute band power of delta band	0.591
Path length	0.571
Step entropy	0.568
PCA angle values in sector 10–45°	0.567
Band power of gamma band normalized by total band	0.506

The next stage of feature selection was the combination of different features. Different features were combined, like in window selection phase, with multiple linear logistic regression model, to produce a seizure probability. SFFS was performed for the prepared data set up to 15 rounds and the criterion value was AUC. In Fig. 14 the top figure represents the evolution of AUC when the number of features increases. In the bottom figure is the increase of AUC when the number of features is increased by one.

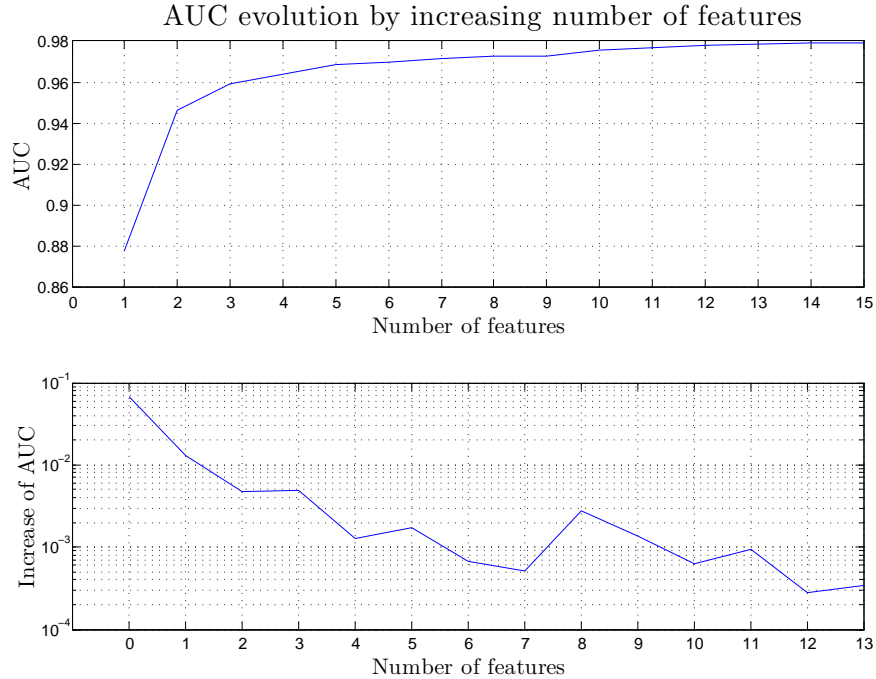


Figure 14: Evolution of AUC when number of features is increased

The number of features selected for testing for the algorithm were 5, 7, and 10. In the bottom figure it is shown that after 5, 7, and 10 features the increase in AUC decreases more significantly than with other number of features. These numbers of features are the local maxima of the curve. The selected 5, 7, and 10 features are listed in Table 16.

When 5 features were selected, three of the features were among the best nine features when features were evaluated individually. In 7 and 10 selected features, five of the features were in the group of the best nine features. Four of the five features were the same, but the last of the five features in 7-feature-model was the band power of alpha band normalized by total band, and in 10 features the band power of lower beta band normalized by total band.

Table 16: Selected features for combinations of 5, 7, and 10 features

5 features

-
1. Spike rate
 2. Evolution angle in sector 0–45°
 3. Random walker steps
 4. Band power of delta band normalized by EEG band
 5. Center of mass *LOGPOW*
-

7 features

-
1. Spike rate
 2. Evolution angle in sector 0–45°
 3. Random walker steps
 4. Band power of delta band normalized by EEG band
 5. Center of mass *LOGPOW*
 6. Path length
 7. Band power of alpha band normalized by total band
-

10 features

-
1. Spike rate
 2. Evolution angle in sector 0–45°
 3. Random walker steps
 4. Band power of delta band normalized by EEG band
 5. Center of mass *LOGPOW*
 6. Path length
 7. Explained variance of the first PCA component
 8. Band power of lower beta band normalized by total band
 9. Band power of higher beta band normalized by EEG band
 10. Band power of 70.2–100Hz band normalized by total band
-

5.2.3 Channel fusion

The linear model produces a seizure probability index for every channel. For a more robust performance, final seizure probability index is a mean of greatest seizure probabilities from multiple channels. For detection making seizure probability index is compared to a threshold.

The number of channels was selected by investigating changes in any-overlap sensitivity and false positive rate in 7-feature-candidate, since based on AUC, specificity and false positive rate, it was considered as the best candidate. Median and mean any-overlap sensitivities and false positive rates of 7 features changing the number of channels from 1 to 6 and to all channels are presented in Fig. 15. The best trade-off between any-overlap sensitivity and false positive rate was achieved with four channels. There were no change in median sensitivities between three and four channels, but with four channels median false positive rate reached zero.

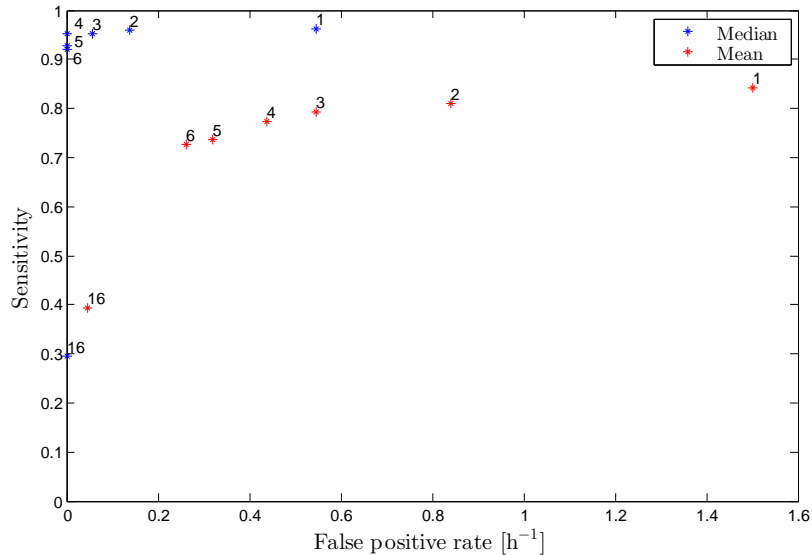


Figure 15: Comparison of number of channels when seizure probability is averaged over channels

5.2.4 Performance in the development data set

After selecting from the features three possible candidates sets, the seizure probability calculation with the candidates was tested with the development data set. With the development data it was possible to produce a continuous seizure probability output and compare it to the seizure markings of the EEG readers.

Any-overlap sensitivity was computed with the seizure data, and overlap-integral specificity and false positive rate were computed with the reference data with com-

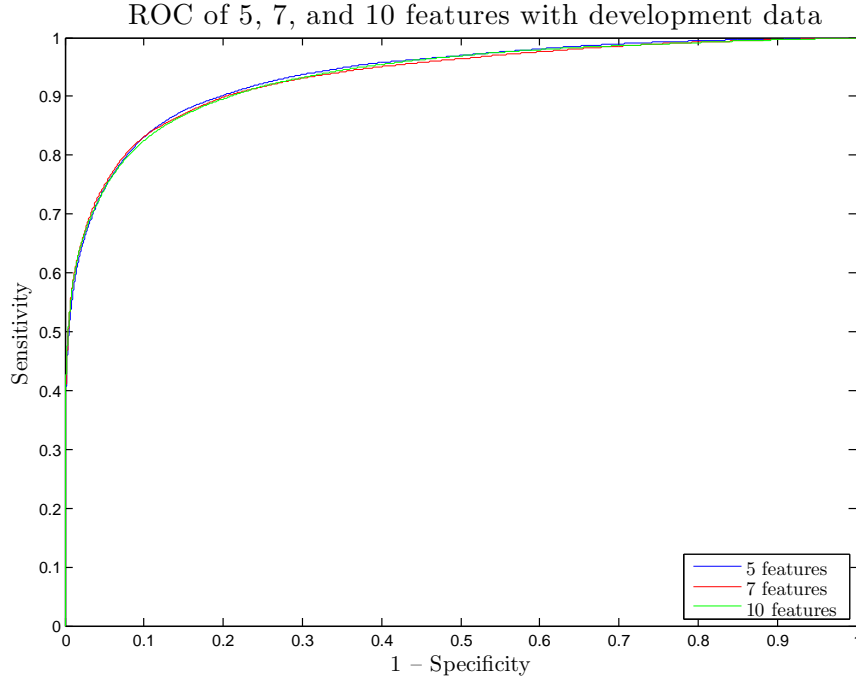


Figure 16: ROC of 5, 7 and 10 features in development set

binations of 5, 7, and 10 features. The threshold was set to 0.75. Seizure probability was computed over four channels. In addition, AUC was computed for the three feature combinations.

Median and mean values of the metrics over the patients in the development set are presented in Table 17. 10 features reached the best sensitivity, but 7 features had the best specificity and lowest false positive rate.

Table 17: Comparison of sensitivity, specificity and false positive rate in models of 5, 7, and 10 features

	Any-overlap sensitivity			Overlap-integral specificity			False positive rate [h^{-1}]		
	5	7	10	5	7	10	5	7	10
Median	0.940	0.949	0.975	0.9996	1	0.9997	0.052	0	0.040
Mean	0.789	0.795	0.815	0.9943	0.9949	0.9868	0.476	0.438	0.970

ROC curves of all the feature combination candidates calculated with the development data are presented in figure 16. According to ROC curves there is no significant difference between the three different models.

In Fig. 17 are presented the values of any-overlap sensitivity for each patient calculated with models of 5, 7, and 10 features. Sensitivities ranged from 0.057 to 1. There were 6 patients for which there was a significant change in sensitivity when

the number of features was changed. The sensitivity increased for the majority of these patients when the number of features was increased.

The overlap-integral specificities for all patients without seizures are shown in Fig. 18. Specificity ranged from 0.6568 to 1. In only one patient there was a significant change when the number of features was changed. In this case, the specificity decreases when number of features increase.

The FPR variability in patients is shown in Fig. 19. There are 21 patients that had 0 FPR with all feature combinations and 40 patients that had FPR below 0.5. There were very great changes in FPR in three patients when the number of features change. With 10 features FPR arises very high in these patients. In addition, there were moderate changes in FPR in 9 patients.

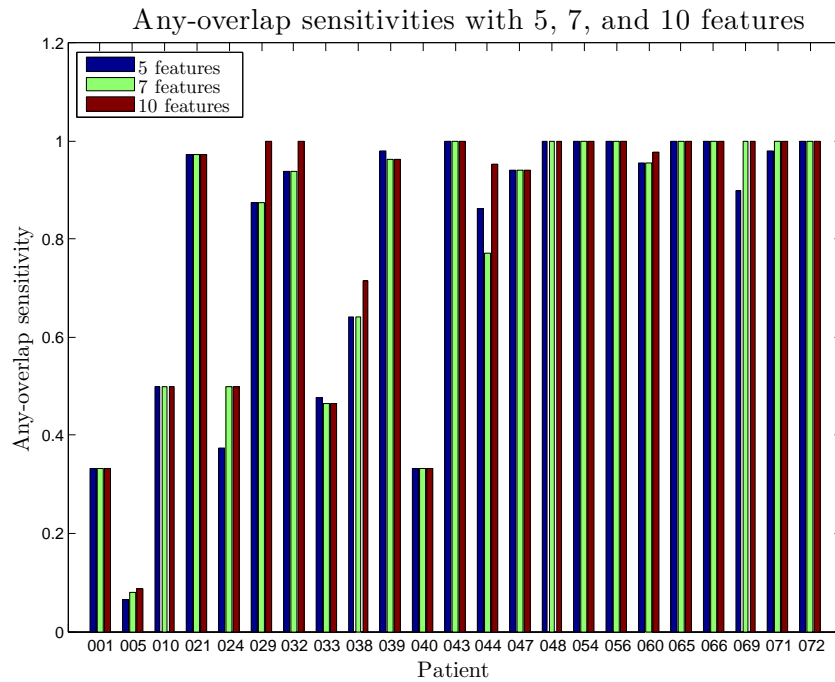


Figure 17: Any-overlap sensitivities for patients in seizure development set with models of 5, 7, and 10 features

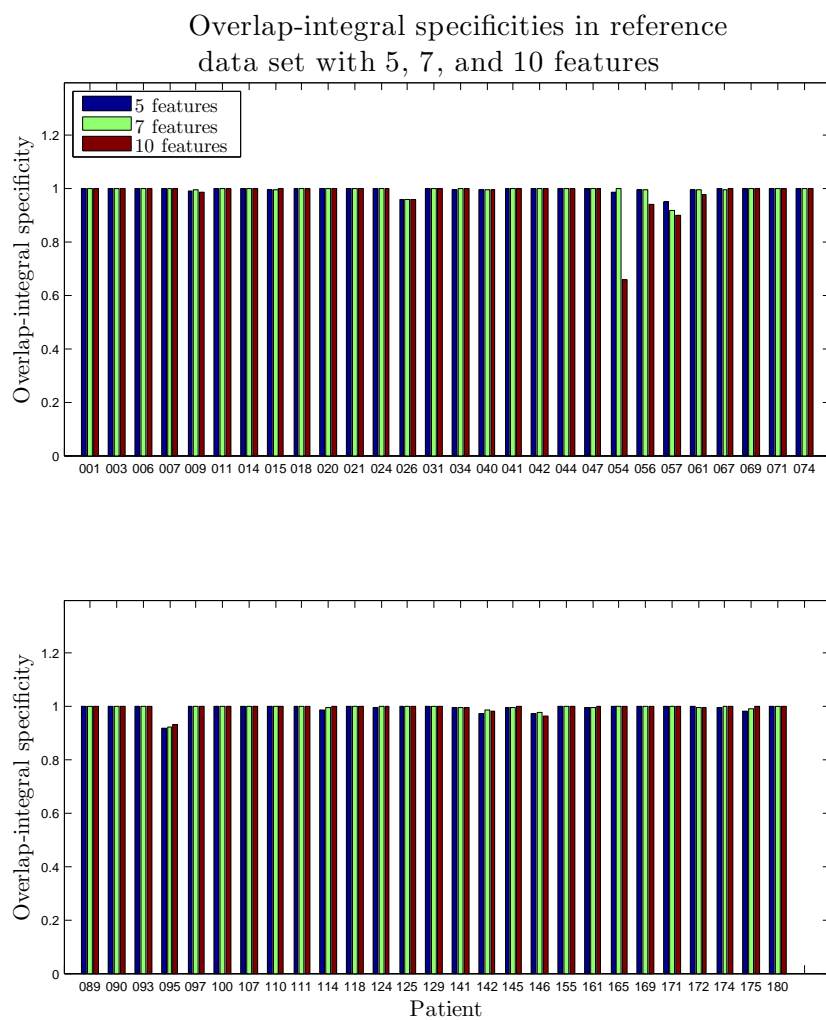


Figure 18: Overlap-integral specificities for patients in reference development set with models of 5, 7, and 10 features

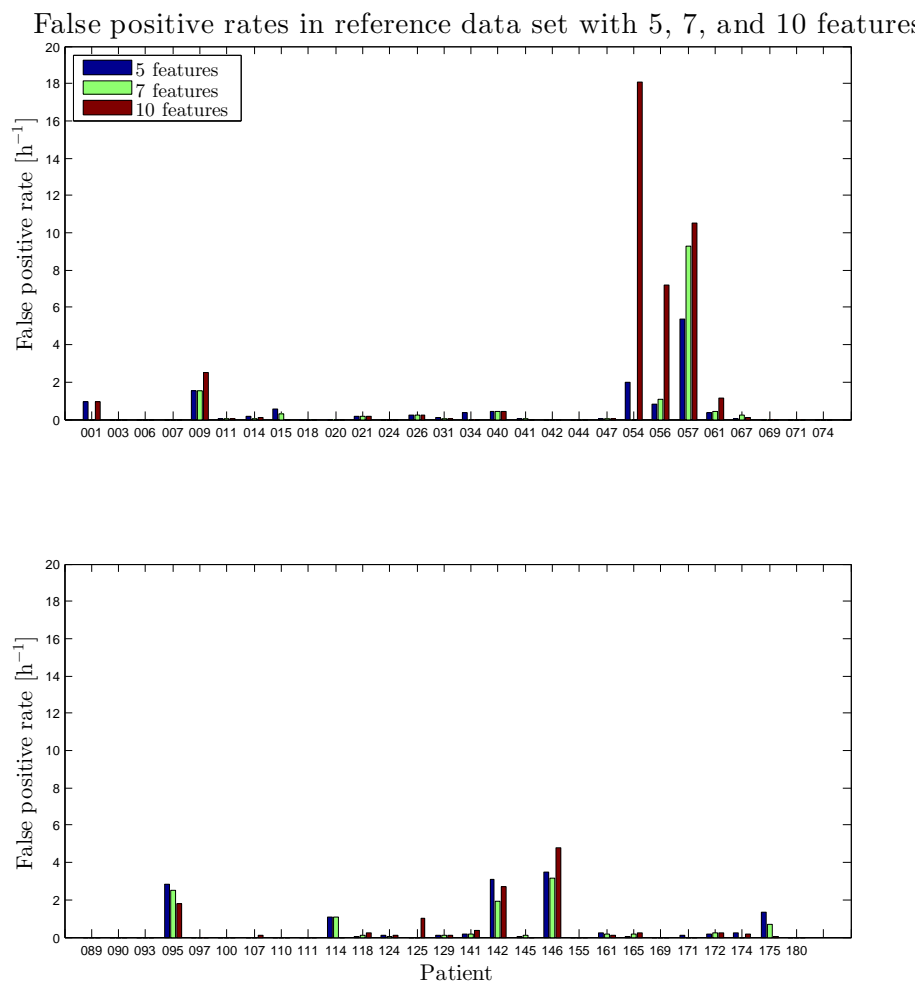


Figure 19: False positive rates for patients in reference development set with models of 5, 7, and 10 features

5.3 Model performance evaluation in an independent data set

Performances of the three model candidates were evaluated with a set of 20 ICU patients that had either equivocal, unequivocal or both kinds of seizures marked by reader 1 of the experts that annotated the development data. In addition, there were a set of 20 ICU patients without seizures. The data set of seizure patients was for evaluating true positive and false negative detections, whereas patients without seizures were for evaluation of false positive and true negative detections.

AUC values of all the three feature models are listed in Table 18. On the first line are the AUCs for data set used in the feature selection. Second line is for AUCs calculated from the development set and last AUC values are from the evaluation data. In the feature selection data the greatest AUC is with 10 features. With the development data and evaluation data 5-feature-model has the highest AUC. However, the difference between AUC in different models is not statistically significant. In further performance evaluation any-overlap and overlap-integral methods were used.

Table 18: Comparison of AUC for feature selection, development and evaluation data with models of 5, 7, and 10 features

	Model		
	5	7	10
AUC in feature selection data	0.9687	0.9717	0.9757
AUC in development data	0.9365	0.9336	0.9341
AUC in evaluation data	0.9009	0.8935	0.8935

5.3.1 Any-overlap methods

Any-overlap methods are useful for evaluating how well the models work as detectors. When there is an overlap with the seizure detection and the seizure marking by an expert, the detection is made correctly.

In Fig. 20 are presented mean any-overlap sensitivity over patients against mean FPR over patients of models of 5, 7, and 10 features. In addition, the mean any-overlap sensitivities and the mean FPRs of algorithm of Persyst Development Corporation, and of inter-reader comparison are marked as points. Inter-reader results are from the original development data of 50 ICU patients. In the inter-reader comparison the false positives are the false positives in the same data set of seizure patients.

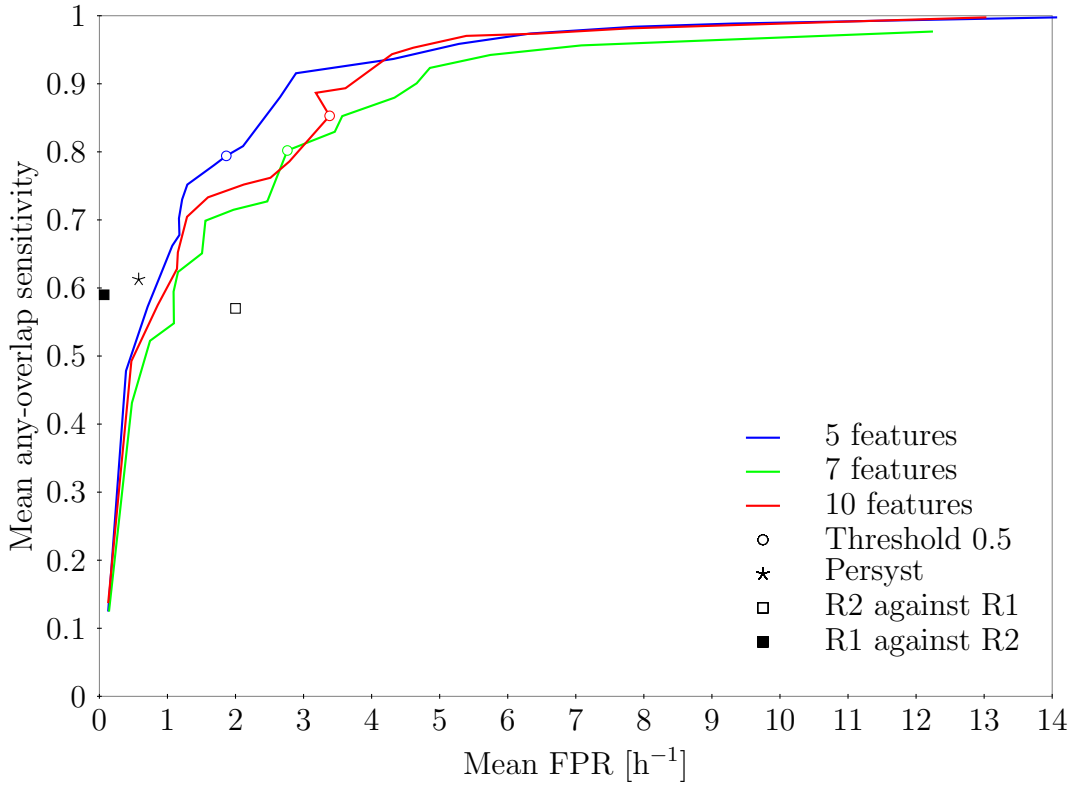


Figure 20: Mean any-overlap sensitivity–mean FPR of models of 5 (blue), 7 (green), and 10 (red) features

According to the curves 5-feature-model has the best performance among the three developed models. At the FPR level of 0.44, Persyst has sensitivity 0.61 that is slightly over the models. However, the sensitivity of 5-feature-model can be increased to 0.79, before step increase in FPR occurs. Reader 1 against reader 2 has the same sensitivity with lower FPR than any of the models or the algorithm of Persyst, but reader 2 against reader has poorer sensitivity with same FPR with respect to the models. The models perform as means of the two experts.

Precision-recall curves of the three models for any-ovelap recall and precision when all data was pooled together is presented in Fig. 21. Precision was computed considering false positives of the data without seizures, thus values for every patient could not be computed.

Also in precision-recall comparison 5 features appear the best of the three models. Precision with threshold of 0.5 is 0.50 which means that half of the detections are made correctly. Sensitivity with the same threshold is 0.82. Precision of detector of Persyst is 0.63 when sensitivity is 0.58. Inter-reader points are also marked in the figure. Reader 2 against reader 1 is approximately on the curve with 5 features. When the reference reader changes, recall has the same value as precision and precision the same value as recall. Also here we have to bear in mind as before that the data for inter-reader comparison is different than for detector evaluation and the false positives are the false positives seizure patients.

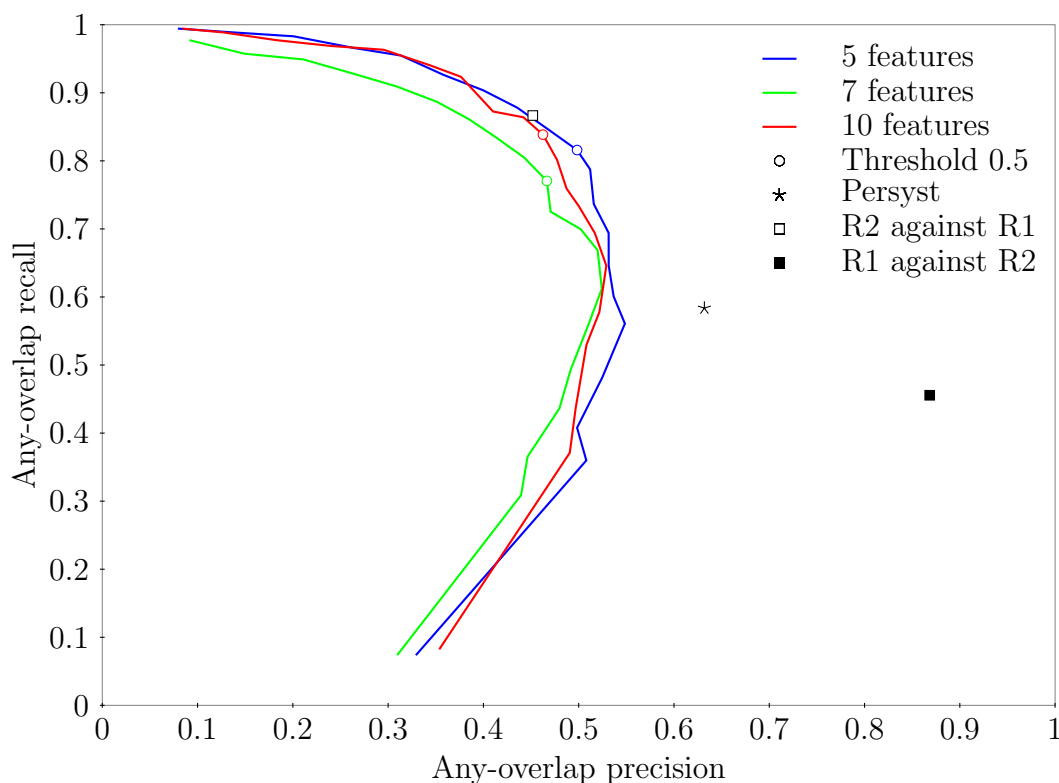


Figure 21: Any-overlap recall–any-overlap precision of models of 5 (blue), 7 (green), and 10 (red) features

As the model of 5 features gave the best sensitivity-FPR and precision-recall comparison, the model of five features was taken into closer analysis. Fig. 22 demonstrates the differences in any-overlap sensitivity – FPR curves of 5 features when the metrics are counted as means or medians over all patients, or when all patient data is pooled together to gain one value for the patient set. When all data is pooled together, every seizure has the same weight, and when mean is considered sensitivity and FPR of every patient has the same weight. The median sensitivities are better when compared to the mean sensitivities with the same FPR. This indicates that there are a few patients with low sensitivity that has decreasing effect to the mean sensitivity.

Threshold 0.5 has been marked to figure. Median any-overlap sensitivity is 0.90 and median FPR is 0.56/h when threshold is 0.5.

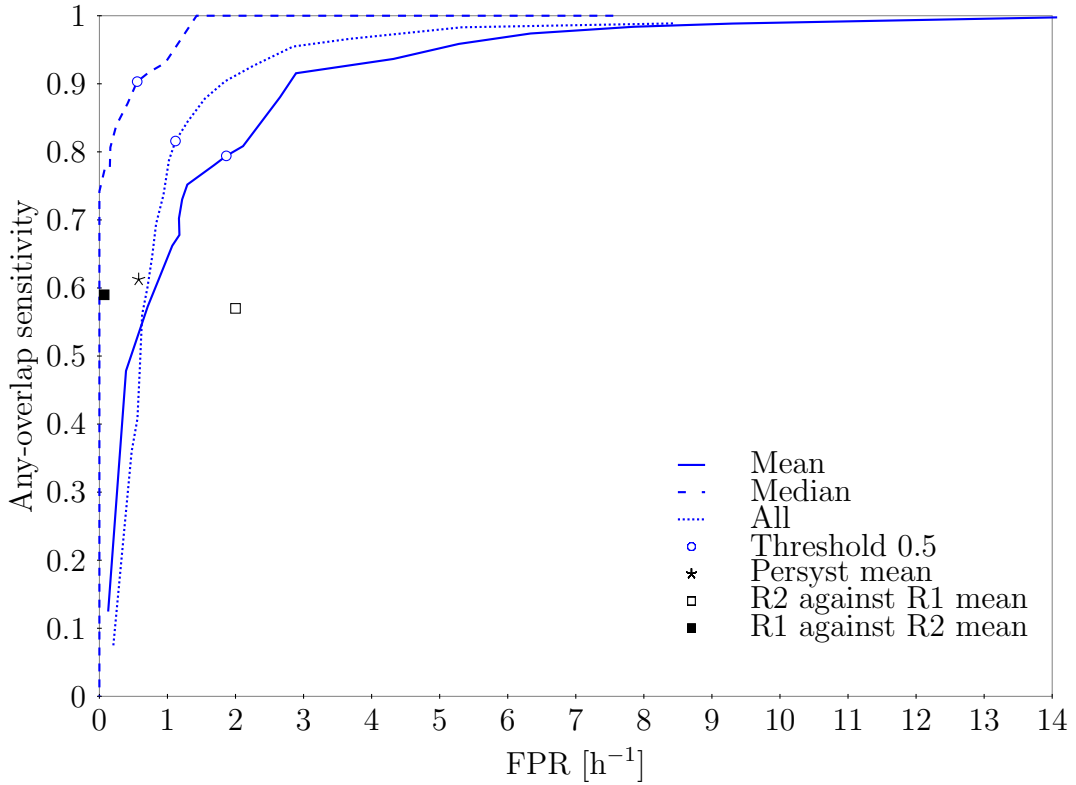


Figure 22: Mean and median any-overlap sensitivities and FPRs and any-overlap sensitivity and FPR of 5–feature–model when all patient data is pooled together

5.3.2 Integral methods

From a clinical point of view it is important also to evaluate how well durations of the seizures are estimated by the model, because seizure duration has association to patient outcome. Integral methods are useful for this evaluation.

In Fig. 23 are presented the durations of detections for every patient of the developed model with 5 features with threshold 0.5 and algorithm of Persyst against durations of unequivocal seizure markings of the expert. The correlation coefficient for the developed model was -0.24 and for the algorithm of Persyst -0.18, i.e. there is no correlation. Black line in the figure represents correlation coefficient 1. There are patients for which the expert has marked less than half an hour of seizures or no seizures, but detectors detect more than 4 hours of seizures. For these patients, the estimation of total seizure duration is very inaccurate.

In Fig. 24 are the detection durations of the detectors against durations of all seizure markings of the expert and correlation coefficient 1 marked as a black line. For all seizures correlation the coefficients are 0.14 for both the developed model and for the algorithm of Persyst. Correlations according to the correlation coefficients can be considered the same for both algorithms. In this figure, there is a group of problematic patients that have more than 2 hours of seizures marked by the expert,

but detectors detect under 2 hours of seizures. Large part of seizure period remains undetected for these patients.

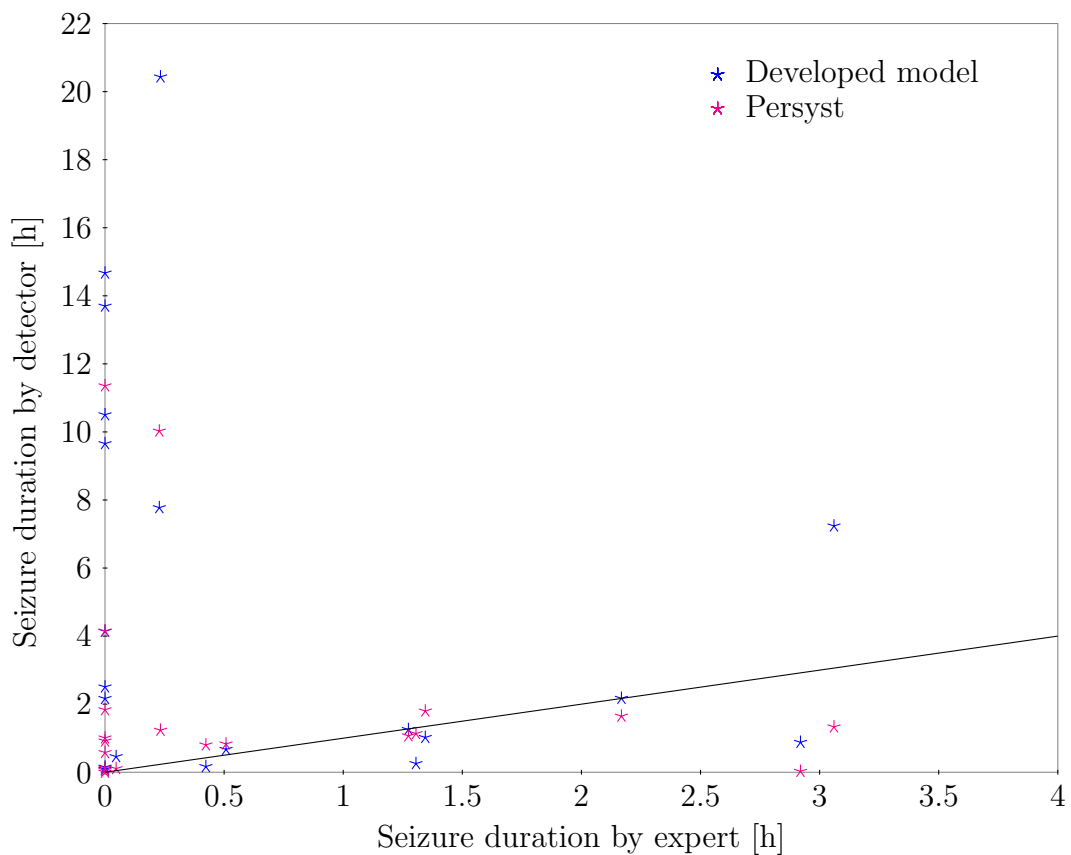


Figure 23: Correlation of unequivocal seizure durations of two detectors and durations of seizures marked by human expert. Black line is the correlation coefficient 1.

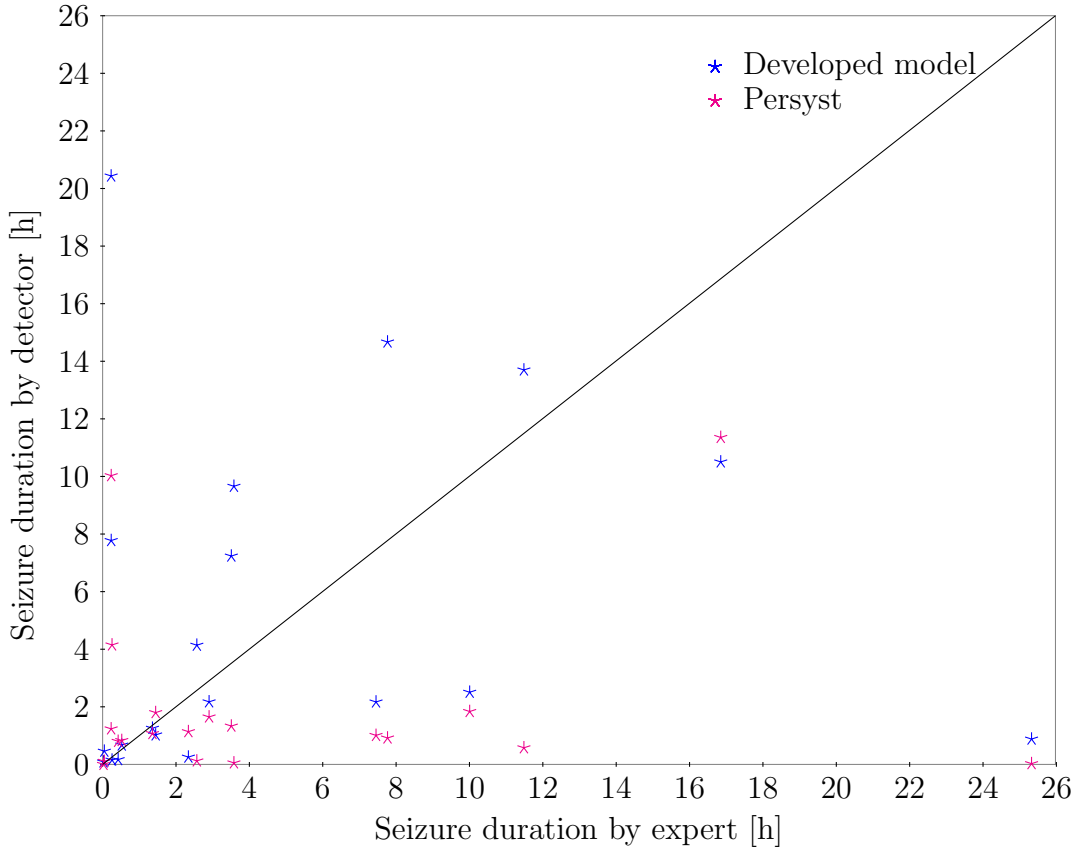


Figure 24: Correlation of all seizure durations of two detectors and durations of seizures marked by human expert. Black line is the correlation coefficient 1.

5.3.3 Results by patients

The results for the 5-feature-model with a threshold of 0.5 are listed in Table 19 for unequivocal seizures, in Table 20 for all seizures, and in Table 21 for reference set without seizures. For unequivocal seizures the median any-overlap sensitivity was 0.903 and mean 0.794. For all seizures the median any-overlap sensitivity was 0.852 and mean 0.770. These are lower than for unequivocal seizures.

The histogram of any-overlap sensitivities of 5, 7, and 10 features for the patients in the independent set with unequivocal seizures are presented in Fig. 25. There were nine patients which did not have any unequivocal seizures according to the EEG reviewer and are not seen in the figure. In seizure patients, there is a significant variability in sensitivity in four patients when the number of features changes. For rest of the patients, the sensitivity remains the same with all the feature combinations or the variability is really small.

The overlap-integral specificity of models with 5, 7, and 10 features for all patients in the reference set are presented in Fig. 26. There is only one patient (case022) for which increasing the number of features causes significant changes in specificity. In this case, specificity starts to decrease when the number of features increases.

False positive rates are presented in Fig. 27. There are three patients for which

FPR remains zero for all combinations of features. For three patients FPR increases significantly and for one patient it decreases significantly when the number of features increases. In addition, there are three patients that have moderate changes in their FPR.

Table 19: Results for unequivocal seizures in evaluation set with model of 5 features and threshold 0.5

case ID	Any-overlap sensitivity	Overlap-integral sensitivity	Overlap-integral specificity	Number of seizures	Number of detections
case077	0.500	0.194	0.9984	12	16
case079	-	-	0.9966	0	20
case082	-	-	0.9967	0	6
case084	1	0.823	0.6743	38	238
case086	0.917	0.727	0.9937	12	25
case089	-	-	0.4408	0	54
case090	1	0.603	0.9809	38	20
case093	0.770	0.385	0.9780	61	65
case095	1	0.851	0.9915	1	52
case096	0.500	0.199	0.9926	2	66
case099	0.263	0.064	0.9926	38	39
case104	-	-	0.9353	0	306
case107	-	-	0.3464	0	119
case112	1	0.708	0.8828	7	281
case115	-	-	0.7212	0	295
case118	0.903	0.811	0.4299	31	305
case122	-	-	0.9435	0	39
case126	0.881	0.506	0.9763	101	193
case129	-	-	0.3812	0	206
case136	-	-	0.9008	0	214
Mean	0.794	0.534	0.8276	17	128
Median	0.903	0.603	0.9599	1.5	66
Total	-	-	-	341	2559

Table 20: Results for all seizures in evaluation set with model of 5 features and threshold 0.5

case ID	Any-overlap sensitivity	Overlap-integral sensitivity	Overlap-integral specificity	Number of seizures	Number of detections
case077	0.500	0.194	0.9984	12	16
case079	0.333	0.040	0.9968	3	20
case082	1	0.702	0.9974	1	6
case084	1	0.823	0.6744	39	238
case086	0.923	0.725	0.9939	13	25
case089	1	0.592	0.7278	1	54
case090	0.955	0.591	0.9822	44	20
case093	0.725	0.367	0.9898	69	65
case095	1	0.851	0.9915	1	52
case096	0.684	0.034	0.9991	19	66
case099	0.145	0.036	0.9899	69	39
case104	0.792	0.120	0.9547	96	306
case107	0.800	0.775	0.4937	10	119
case112	0.714	0.653	0.8846	28	281
case115	1	0.875	0.7899	2	295
case118	0.903	0.811	0.4299	31	305
case122	0.154	0.092	0.9521	13	39
case126	0.771	0.408	0.9778	131	193
case129	1	0.936	0.5356	1	206
case136	1	0.815	0.9477	1	214
Mean	0.770	0.522	0.8648	29	128
Median	0.852	0.623	0.9662	13	66
Total	-	-	-	584	2559

Table 21: Results for reference evaluation set with model of 5 features and threshold 0.5

case ID	False positive rate [h ⁻¹]	Overlap-integral specificity	Number of detections	Duration of detections	Duration of record
case002	0	1	0	0:00:00	3:21:42
case010	0.565	0.9962	12	0:04:54	21:15:20
case017	0	1	0	0:00:00	0:23:55
case022	2.293	0.9917	1	0:00:13	0:26:10
case028	0.263	0.9976	5	0:02:47	19:01:53
case039	0	1	0	0:00:00	0:53:56
case045	4.641	0.9941	3	0:01:30	0:38:47
case053	4.997	0.7713	5	0:14:20	1:00:02
case065	1.839	0.9901	6	0:01:56	3:15:46
case087	0.148	0.9950	3	0:06:02	20:14:46
case092	7.521	0.5203	155	9:54:58	20:36:31
case105	1.969	0.9929	1	0:00:13	0:30:28
case120	0.219	0.9986	5	0:01:52	22:49:23
case121	1.457	0.9839	32	0:21:13	21:58:04
case131	0.117	0.9985	3	0:02:20	25:43:54
case140	9.008	0.9499	3	0:02:07	0:19:59
case151	0.564	0.9952	11	0:05:51	20:08:19
case162	0.928	0.9894	26	0:17:51	28:00:40
case167	0.246	0.9963	6	0:05:29	24:24:39
case182	0.535	0.9925	13	0:10:53	24:18:10
Mean	1.865	0.9577	15	0:34:43	12:58:07
Median	0.555	0.9946	5	0:02:33	19:35:06
Total	-	-	290	11:34:29	259:22:24

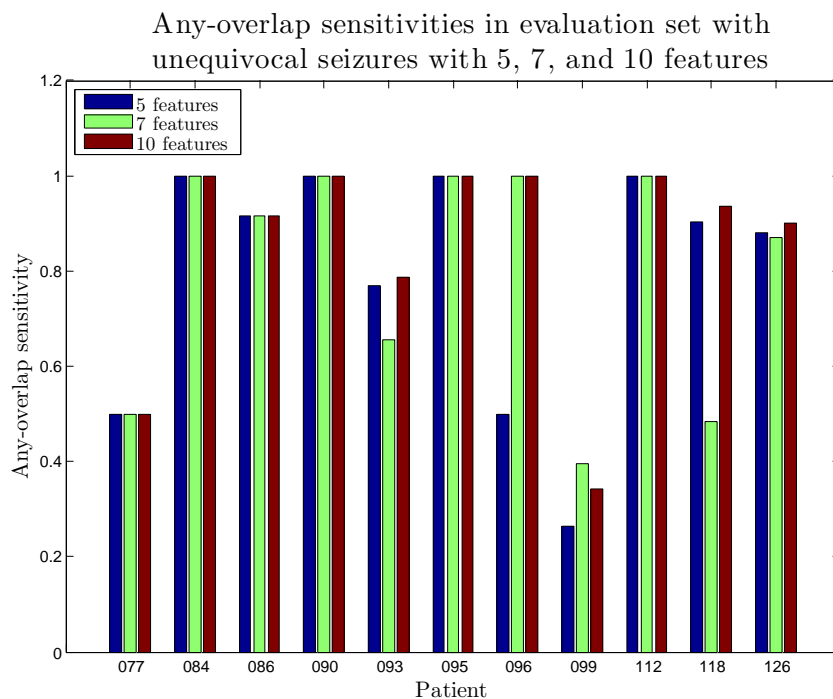


Figure 25: Any-overlap sensitivity for unequivocal seizures in evaluation set with models of 5, 7, and 10 features

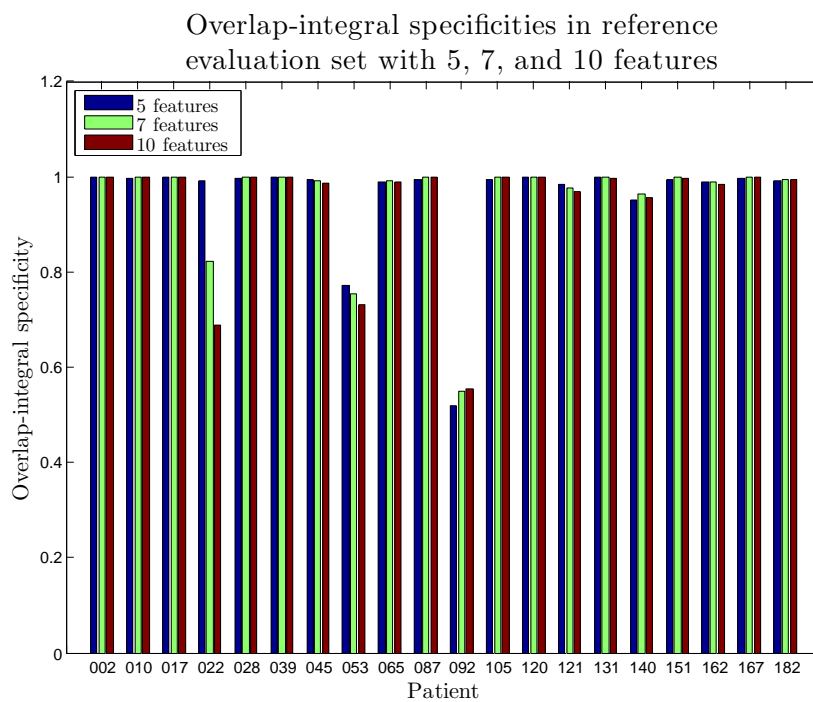


Figure 26: Overlap-integral specificity in reference evaluation set with models of 5, 7, and 10 features

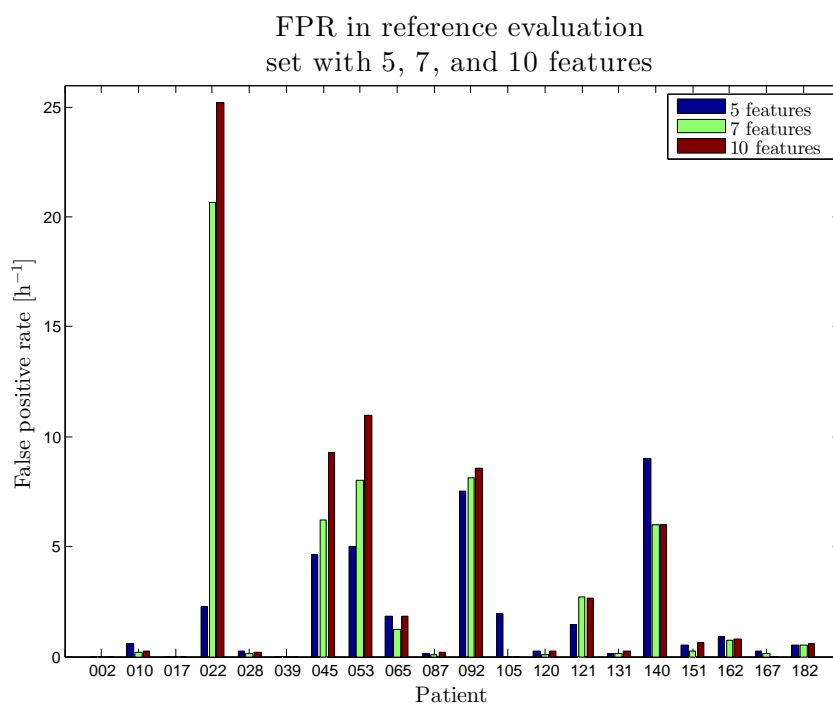


Figure 27: False positive rate in reference evaluation set with models of 5, 7, and 10 features

6 Discussion

The purpose of this study was to find features from EEG to distinguish seizures with evolution from non-seizure periods in EEG. A data set of EEG records with seizures were analyzed by two experienced EEG reviewers. Seizure markings of the two experts were compared and the agreement between experts was moderate. A data set of agreed seizures and a reference data set without seizures were used for development of features for classification of seizure periods and non-seizure periods. A part of the features were calculated in two different feature spaces with different time windows and a part from the original EEG signal. The selection of features was performed with optimization methods. The developed feature model gives promising results for seizure detection, but still needs some further development.

6.1 Comparison of two experts

The development data used for the model development was reviewed by two EEG readers. As the comparison of two experts demonstrates there is only moderate agreement in seizure markings in EEG of ICU patients. The agreement according to kappa was very good for 16 patients when only unequivocal seizures were considered. It is a little less than for one third of the patients. For all seizures very good agreement was found only in 8 patients, which is half of the number of the patients compared to the group of unequivocal seizures. This indicates that objective seizure reference is very difficult to assess, especially for equivocal seizures.

When the agreement between the experts was very good according to kappa, sensitivity of readers was good. However, when the sensitivity of the readers was good, the agreement according to kappa could be fair or poor. This is due to that sensitivity metrics do not take into account false positive markings with respect to the other reader. For this reason, sensitivity as one figure is not sufficient for measuring agreement. The sensitivities against both readers should be computed for interpretation of the level of agreement. For example, if reader X marks many seizures and reader Y only a few seizures, and both readers agree on the few seizures marked by reader Y, the sensitivity of reader X against reader Y will be high as well as FPR, but the sensitivity of reader Y against reader X will be low with low FPR. Instead, kappa, which is only one figure, weighs both seizure periods and non-seizure periods, and thus the skewness of the classification distribution is taken into account. This is important, because usually the duration of non-seizure periods is much longer than the duration of seizure periods.

There is a definition of characteristics of EEG signal during seizure, but every EEG viewer has knowledge gained by experience and a personal way of reviewing the signal. The level of experience affects to the markings of the reviewer. This might cause differences in the EEG review.

In records, where there were a strong disagreement, the quality of the record may have an effect. If there is a great amount of artifacts, one reader may decide that the quality of the record is not sufficient for the review while the other may mark seizures. In addition, there are EEG patterns that divide expert opinions whether

they are included in seizure activity or not.

6.2 Combination of channels

The models for seizure detection were formed by combining different features to produce one index. In feature selection the spike rate was the first feature in all of the models. The spike rate is a feature which is formed from information from all of the channels whereas rest of the features are calculated only from one channel. When seizure probabilities are combined from several channels, the spike rate is counted several times in channel fusion. This forms a bias to the index averaged over channels.

The combination of information from different channels was performed by averaging the seizure probability index over four channels that had the largest values. When seizure is very local and seen only on few channels, four channels may not be the optimal number of channels for averaging. In these cases, the seizure probability index might become too low for detection due to the averaging. Combining the channels should be more adaptive to whether the seizure is local or generalized. The information from different number of channels could be combined differently in cases, depending on how spread the seizure is.

6.3 Model comparison and evaluation

The performance of the developed models with 5, 7, and 10 features did not differ much when based on the development data. With the independent data set the 5-feature model appeared the best. In the development phase, it is possible to have very optimistic performance figures by adding more features and making the model more complex. It is important not to overtrain the model. In the evaluation with the independent data, the performance figures showed that with 7 and 10 features there was already some possible overfitting.

The intra-patient variability in performance of the models was mainly small when the number of features was increased. There were proportionally more patients in the evaluation set than in the development set with a significant increase or decrease in sensitivity or FPR when the number of features was increased. If there were differences between the models in FPR in a patient, in the majority of cases FPR. The differences in sensitivity in two patients in the evaluation set showed significantly lower sensitivity with 7 features than with other models. The only feature that differed in the 7-feature model with respect to other models was the normalized alpha band power. Poorer performance with greater number of features indicates that there was some overfitting.

The differences in specificity between the models in each patient were very little. In both development and evaluation sets there was only one patient for whom the difference was significant. EEG records are in average very long and the length of EEG periods correctly detected as non-seizure periods are very long with respect to the periods falsely detected as seizures. The specificity is a quite robust metric for

small differences between the models in the number of false detections, if the record duration is long and durations of the false detections are short.

The inter-patient variability in performance of the models was larger than the intra-patient variability. Great inter-patient variability but small intra-patient variability indicate that the seizures between patients appear very different, but in one patient the seizures appear similar to each other. One particular feature set that works in one patient, may not work with other patients.

The main difficulty in developing a general detection algorithm is the variability of seizures between patients, and therefore a great deal of data is required. The original development data were 50 patients with seizures of which 32 or 33 patients with unequivocal seizures depending on the reviewer. When only the patients with unequivocal seizures agreed by both experts were selected, the number of patients was reduced to 24 patients with total 698 seizures and total seizure duration of 20.5 hours. Even after the reduction, the amount of seizures is large and very near to the amount of data used in the development of the commercial Reveal algorithm.

Absolute performance evaluation of the models is a difficult task, because there is a lack of absolute reference for seizures. In this study, model evaluation with the independent data set was referenced to seizure markings of one EEG reader. As noticed in the EEG reader comparison, there is a great disagreement between human experts. The numbers describing the performance of the models may have been different when compared to the seizure markings of other EEG readers.

In this study, several performance metrics were used for evaluating the models in order to give a thorough view of their performances. Different metrics emphasize different aspects of the performance. The difficulty is to combine the information given by all the different metrics and to define, which are the requirements for a good detector that fulfills the needs in clinical use.

Due to the subjectivity of seizure markings of the experts, the performance of the models should be compared to the level of agreement of the experts. In this study a direct comparison is not possible. The agreement analysis of the experts has been made with the data used for the method development and the evaluation of the model has been done with a different data set. Even though a direct comparison between the metrics is not reliable, the level of agreement of the experts gives a reference value to which reflect the performance of the model.

When the mean any-overlap sensitivity and the mean FPR results were compared to the median any-overlap sensitivity and the median FPR over patients, median values appeared much better. This indicates that the sensitivities and FPRs are not normally distributed. There are a few problematic patients that caused the mean value to be lower than the median.

The performance of the model was better for unequivocal seizures than for all seizures. This is not surprising, because the model was developed with a data set of only unequivocal seizures. From a clinical point of view, it is important to detect equivocal seizures as well. As seen in the comparison of experts, the agreement between experts is poorer for equivocal seizures. Defining an objective reference for equivocal seizures is even more difficult than for unequivocal seizures.

In mean and median any-overlap sensitivity with a threshold of 0.5 the perfor-

mance was good, but the false positive rate given by the model was too high. One source for false positive detections is EMG. EMG removal was used in the artifact detection, but better methods for EMG detection are needed. The risk in EMG detection is that if it is made too sensitive, it may remove data in which EMG is not dominant and seizures may be missed.

The precision of the model was not good enough yet. Approximately half of the detections were incorrect and this requires improvement. By reducing false detections the precision will improve also.

Correlation of seizure durations between automatic detections and seizure markings by the expert was poor for all seizures according to the correlation coefficient. For unequivocal seizures there was no correlation. Both the detector developed in this work and the detector of Persyst had similar correlation coefficients. When considering unequivocal seizures, there were patients that had much longer seizure durations according to the detectors than according to the expert. For these patients many of the detections are false positive. For all seizures, there was, however, a group of patients that had much shorter duration of seizures according to detectors compared to markings of the expert. In these cases, a great part of the seizure period was not detected by the detectors. For unequivocal seizures the most problematic patients for the detectors were those that produced excessive detections and for all seizures the most problematic cases were those that missed detections. As mentioned earlier, the detector developed in this study was developed with data set of unequivocal seizures, which explains the lower sensitivity for equivocal seizures and missed detections in all seizures. Based on the correlation analysis, seizure duration estimation is not satisfactory yet.

Compared to the commercial algorithm of Persyst, the 5-feature-model had higher FPR and poorer precision with the sensitivity of around 0.6 than Persyst. However, with a lower threshold the model had a sensitivity of approximately 0.8. Sensitivity is an important factor and it should be sufficiently high for seizure patients and individual seizures. Sensitivity of around 0.8 is at same level as the median sensitivity of the experts based on the expert comparison. The model developed in this study was developed and evaluated with ICU data whereas the algorithm of Persyst was developed with EMU data. This may explain, why the sensitivity of the algorithm of Persyst is not higher than 0.6.

6.4 Future development

In this study, the agreement in seizure markings between two experts was studied. The agreement between those two experts appeared to be moderate. To form a clearer and more reliable idea of the degree of agreement between human experts reviewing ICU patient EEG, comparison between a larger group of readers is needed. A better understanding of the degree of agreement between the experts would give also a better reference to the performance of automatic detectors.

One of the greatest challenges in seizure detector development is the data used in the development work and as a reference for the performance evaluation. For future development and performance evaluation a data set reviewed by more than

two human experts would give more objective view of the performance.

In the future development the reduction of the number of the false positive detections is of great importance. Combining a well functioning artifact detector with the model would reduce the false positive detections.

In the model developed in this work, the spike rate was included in the seizure probability index of each channel. As the spike rate is a feature combined from several channels, in future development inclusion of the spike rate to the detector could be handled differently.

In order to improve the sensitivity of the algorithm, new, and more adaptive ways of combining information from different channels should be developed. Seizure activity can be local or general and channel combination should take this into account. In addition, a time limit for seizure detection could be optimized.

7 Conclusions

In this thesis work the objective was to find methods for automatic seizure detection from EEG of ICU patients. A data selection for development database was made by comparing seizure markings of two experts. The agreement between the two experts was moderate for unequivocal seizures and poorer for all seizures.

For the development of detection methods only seizures agreed by experts were selected. From the selected data several features were computed in different spaces and time windows. The most suitable time window and three models with different feature combinations were selected with optimization methods. The models were compared by evaluating their performance with statistical methods. The best model was a model of 5 features.

Performance of the model of 5 features gave promising results. The model was able to detect all seizure patients. However, the detector should produce less false positive detections in order the rate to be acceptable in clinical use. Further development is still needed to reduce false positive detections. With a better artifact detector combined to the detector, the amount of false positive detection should be most likely reduced substantially.

References

- [1] B. J. Fisch, *Fisch & Spehlmann's EEG Primer*. Elsevier, third ed., 1999.
- [2] G. M. Brophy, R. Bell, J. Claassen, B. Alldredge, T. P. Bleck, T. Glauser, S. M. Laroche, J. J. Riviello, L. Shutter, M. R. Sperling, D. M. Treiman, and P. M. Vespa, "Guidelines for the evaluation and management of status epilepticus," *Neurocritical Care*, vol. 17, pp. 3–23, Aug. 2012.
- [3] L. J. Hirsch and R. P. Brenner, *Atlas of EEG in Critical Care*. Wiley-Blackwell, 2010.
- [4] J. Jirsch and L. J. Hirsch, "Nonconvulsive seizures: developing a rational approach to the diagnosis and management in the critically ill population," *Clinical Neurophysiology : Official Journal of the International Federation of Clinical Neurophysiology*, vol. 118, pp. 1660–70, Aug. 2007.
- [5] J. Claassen, S. a. Mayer, R. G. Kowalski, R. G. Emerson, and L. J. Hirsch, "Detection of electrographic seizures with continuous EEG monitoring in critically ill patients," *Neurology*, vol. 62, pp. 1743–8, May 2004.
- [6] F. W. Drislane, M. R. Lopez, A. S. Blum, and D. L. Schomer, "Detection and treatment of refractory status epilepticus in the intensive care unit," *Journal of Clinical Neurophysiology : Official Publication of the American Electroencephalographic Society*, vol. 25, pp. 181–6, Aug. 2008.
- [7] G. B. Young, K. G. Jordan, and G. S. Doig, "An assessment of nonconvulsive seizures in the intensive care unit using continuous EEG monitoring: an investigation of variables associated with mortality," *Neurology*, vol. 47, pp. 83–9, July 1996.
- [8] D. Friedman, J. Claassen, and L. J. Hirsch, "Continuous electroencephalogram monitoring in the intensive care unit," *Anesthesia and Analgesia*, vol. 109, pp. 506–23, Aug. 2009.
- [9] N. S. Abend, D. J. Dlugos, C. D. Hahn, L. J. Hirsch, and S. T. Herman, "Use of EEG monitoring and management of non-convulsive seizures in critically ill patients: a survey of neurologists," *Neurocritical Care*, vol. 12, pp. 382–9, June 2010.
- [10] J. Malmivuo and R. Plonsey, *Bioelectromagnetism: Principles and Applications of Bioelectric and Biomagnetic Fields*. Oxford U.P., 1995.
- [11] S. L. Bridgers and J. S. Ebersole, "EEG outside the hairline: Detection of epileptiform abnormalities," *Neurology*, vol. 38, no. 1, pp. 146–149, 1988.
- [12] G. B. Young, M. D. Sharpe, M. Savard, E. Al Thenayan, L. Norton, and C. Davies-Schinkel, "Seizure detection with a commercially available bedside EEG monitor and the subhairline montage," *Neurocritical Care*, vol. 11, pp. 411–6, Dec. 2009.

- [13] E. Ikonen, “No Secret Science,” Master’s thesis, Aalto University, School of Arts, Design and Architecture, 2012.
- [14] T. P. Bleck, M. C. Smith, S. J. Pierre-Louis, J. J. Jares, J. Murray, and H. C. A., “Neurologic complications of critical medical illnesses,” *Critical Care Medicine*, pp. 98–103, 1993.
- [15] G. B. Young, R. S. McLachlan, J. H. Kreeft, and J. D. Demelo, “An electroencephalographic classification for coma,” *The Canadian Journal of Neurological Sciences. Le Journal Canadien des Sciences Neurologiques*, vol. 24, pp. 320–5, Nov. 1997.
- [16] C. Binnie, R. Cooper, F. Mauguière, J. Osselton, P. Prior, and B. Tedman, *Clinical Neurophysiology : EEG, Paediatric Neurophysiology, Special Techniques and Applications*, vol. 2. Elsevier Science, 2003.
- [17] D. M. White and A. C. Van Cott, “EEG artifacts in the intensive care unit setting,” *American Journal of Electroneurodiagnostic Technology*, vol. 50, pp. 8–25, Mar. 2010.
- [18] S. B. Wilson, M. L. Scheuer, C. Plummer, G. B. Young, and S. Pacia, “Seizure detection : correlation of human experts,” *Clinical Neurophysiology*, vol. 114, pp. 2156–2164, 2003.
- [19] L. J. Hirsch, S. M. LaRoche, N. Gaspard, E. Gerard, A. Svoronos, S. T. Herman, R. Mani, H. Arif, N. Jette, Y. Minazad, J. F. Kerrigan, P. Vespa, S. Hantus, J. , G. B. Young, E. So, P. W. Kaplan, M. R. Nuwer, N. B. Fountain, and F. W. Drislane, “American Clinical Neurophysiology Society’s Standardized Critical Care EEG Terminology: 2012 version.” Published online, <https://www.acns.org/pdf/guidelines/Guideline-14.pdf>, 2012.
- [20] J. D. Pandian, G. D. Cascino, E. L. So, E. Manno, and J. R. Fulgham, “Digital Video-Electroencephalographic Monitoring in the Neurological-Neurosurgical Intensive Care Unit,” *Arch Neurol*, pp. 1090–1094, 2004.
- [21] N. Jette, J. Claassen, R. Emerson, and L. J. Hirsch, “Frequency and Predictors of Nonconvulsive Seizures During Continuous Electroencephalographic Monitoring in Critically Ill Children,” *Arch Neurol*, pp. 1750–1755, 2006.
- [22] P. M. Vespa, K. O’Phelan, M. Shah, J. Mirabelli, S. Starkman, C. Kidwell, J. Saver, M. R. Nuwer, J. G. Frazee, D. a. McArthur, and N. a. Martin, “Acute seizures after intracerebral hemorrhage: A factor in progressive midline shift and outcome,” *Neurology*, vol. 60, pp. 1441–1446, May 2003.
- [23] R. J. DeLorenzo, E. J. Waterhouse, A. R. Towne, J. G. Boggs, D. Ko, G. A. DeLorenzo, A. Brown, and L. Garnett, “Persistent nonconvulsive status epilepticus after the control of convulsive status epilepticus,” *Epilepsia*, vol. 39, pp. 833–40, Aug. 1998.

- [24] P. M. Vespa, V. Nenov, and M. R. Nuwer, "Continuous EEG Monitoring in the Intensive Care Unit: Early Findings and Clinical Efficacy," *Journal of Clinical Neurophysiology*, vol. 16, no. 1, pp. 1–13, 1999.
- [25] J. Gotman, "Automatic recognition of epileptic seizures in the EEG," *Electroencephalography and Clinical Neurophysiology*, vol. 54, pp. 530–540, 1982.
- [26] J. Gotman, "Automatic seizure detection : improvements and evaluation," *Electroencephalography and Clinical Neurophysiology*, vol. 76, pp. 317–324, 1990.
- [27] S. B. Wilson, M. L. Scheuer, R. G. Emerson, and A. J. Gabor, "Seizure detection : evaluation of the Reveal algorithm," *Clinical Neurophysiology*, vol. 115, pp. 2280–2291, 2004.
- [28] K. M. Kelly, D. S. Shiau, R. T. Kern, J. H. Chien, M. C. K. Yang, K. A. Yandora, J. P. Valeriano, J. J. Halford, and J. C. Sackellares, "Assessment of a scalp EEG-based automated seizure detection system," *Clinical Neurophysiology*, vol. 121, no. 11, pp. 1832–1843, 2010.
- [29] M. A. Navakatikyan, P. B. Colditz, C. J. Burke, T. E. Inder, J. Richmond, and C. E. Williams, "Seizure detection algorithm for neonates based on wave-sequence analysis," *Clinical Neurophysiology*, vol. 117, pp. 1190–203, June 2006.
- [30] B. R. Greene, S. Faul, W. P. Marnane, G. Lightbody, I. Korotchikova, and G. B. Boylan, "A comparison of quantitative EEG features for neonatal seizure detection," *Clinical Neurophysiology : Official Journal of the International Federation of Clinical Neurophysiology*, vol. 119, pp. 1248–61, June 2008.
- [31] A. Temko, E. Thomas, W. Marnane, G. Lightbody, and G. Boylan, "EEG-based neonatal seizure detection with Support Vector Machines," *Clinical Neurophysiology : Official Journal of the International Federation of Clinical Neurophysiology*, vol. 122, pp. 464–473, Aug. 2010.
- [32] Y. Khan and J. Gotman, "Wavelet based automatic seizure detection in intracerebral electroencephalogram," *Clinical Neurophysiology*, vol. 114, pp. 898–908, May 2003.
- [33] H. Qu and J. Gotman, "A Patient-Specific Algorithm for the Detection of Seizure Onset in Long-Term EEG Monitoring : Possible Use as a Warning Device," *IEEE Transactions on Biomedical Engineering*, vol. 44, no. 2, pp. 115–122, 1997.
- [34] J. L. Smolowitz, S. C. Hopkins, T. Perrine, K. E. Eck, L. J. Hirsch, and M. O'Neil Mundinger, "Diagnostic utility of an epilepsy monitoring unit," *American Journal of Medical Quality : The Official Journal of the American College of Medical Quality*, vol. 22, no. 2, pp. 117–22, 2007.

- [35] A. E. J. Tanner, M. O. K. Särkelä, J. Virtanen, H. E. Viertiö-Oja, M. D. Sharpe, L. Norton, C. Davies-Schinkel, and G. B. Young, “Application of Subhairline EEG Montage in Intensive Care Unit: Comparison with Full Montage,” *In press, Journal of Clinical Neurophysiology*, 2014.
- [36] A. Tanner, “Automatic seizure detection using a two-dimensional EEG feature space,” Master’s thesis, Aalto University, 2011.
- [37] L. Sörnmo and P. Laguna, *Biomedical Engineering : Bioelectrical Signal Processing in Cardiac and Neurological Applications*. Academic Press, 2005.
- [38] M. Johansson, “The Hilbert transform,” Master’s thesis, Växjö University, 1999.
- [39] F. van der Heijden, R. Duin, D. de Ridder, and T. D.M.J., *Classification, Parameter Estimation and State Estimation - An Engineering Approach using MATLAB*. Wiley, 2004.
- [40] R. C. Gonzales and R. E. Woods, *Digital Image Processing*. Addison Wesley, 1993.
- [41] D. G. Altman, *Practical Statistics for Medical Research*. Chapman and Hall/CRC, 1991.
- [42] M. Bland, *An Introduction to Medical Statistics*. Oxford Press, third ed., 2000.
- [43] J. Davis and M. Goadrich, “The relationship between Precision-Recall and ROC curves,” *Proceedings of the 23rd International Conference on Machine Learning - ICML '06*, pp. 233–240, 2006.
- [44] C. Manning and H. Schütze, *Foundations of Statistical Natural Language Processing*. MIT Press, 1999.
- [45] S. Theodoridis and K. Koutroumbas, *Pattern Recognition*. Academic Press, 1999.