Tampere University

KAVEH SAMIEE

Advanced Feature Extraction for Classification of Long-Term Epileptic Electroencephalography Records

Tampere University Dissertations 46

Tampere University Dissertations 46

KAVEH SAMIEE

Advanced Feature Extraction for Classification of Long-Term Epileptic Electroencephalography Records

ACADEMIC DISSERTATION

To be presented, with the permission of

the Faculty Council of the Faculty of Information Technology and Communication Sciences

of Tampere University,

for public discussion in the auditorium TB109

of the Tietotalo, Korkeakoulunkatu 1, Tampere,

on 10 May 2019, at 12 o'clock.

ACADEMIC DISSERTATION
Tampere University,
Faculty of Information Technology and Communication Sciences
Finland

Responsible supervisor and Custos	Professor Moncef Gabbouj Tampere University Finland	
Pre-examiners	Professor Mohamad Sawan Polytechnique Montreal Canada	Professor Mounir Sayadi National Superior School of Engineers of Tunis Tunis
Opponent	D.Sc. (Tech.) Miikka Ermes Tieto Corporation and University of Eastern Finland Finland	

The originality of this thesis has been checked using the Turnitin OriginalityCheck service.

Cover design: Roihu Inc.

ISBN 978-952-03-1052-3 (print)
ISBN 978-952-03-1053-0 (pdf)
ISSN 2489-9860 (print)
ISSN 2490-0028 (pdf)
http://urn.fi/URN:ISBN:978-952-03-1053-0

PunaMusta Oy – Yliopistopaino
Tampere 2019

Abstract

Recent advances in artificial intelligence (AI) offer many opportunities to implement it in a broad range of industries. One of the main ambitious application of AI is in healthcare and patient monitoring. In healthcare industry, unlike the most commercial applications of AI, a missed detection/prediction of a clinical event may result in the death of a patient. Moreover, a high false alarm rate may lead to misdiagnosis causing extra effort and cost for care providers. Thus, in healthcare applications, it is required to strictly minimize the number of false alarms without sacrificing the sensitivity rate of the detection system. This dissertation focuses on epileptic seizure detection and classification in long-term electroencephalogram (EEG) records. More specifically, the two main challenges in supervised EEG seizure detection, the curse of dimensionality and the curse of variability are tackled.

First, a signal decomposition technique, applicable to physiological signals, is devised which can be used as a preliminary step for feature extraction. This is performed by proposing a novel time-frequency transform based on rational functions, namely, rational short time Fourier transform (RSTFT). A sparse decomposition method is then proposed by reconstructing the input signal into several components using non-overlapping sub-sets of the RSTFT coefficients. Sparse representation of signal components is then obtained by inducing L1 regularization penalty on the RSTFT coefficients during the reconstruction phase. The effectiveness of the proposed sparse decomposition method is evaluated in the classification of long-term EEG records for purposes of epileptic seizure detection and sleep stage scoring.

Another part of the thesis investigates the detection of seizures in textural images constructed using a proposed scheme for mapping of EEG signals into gray-scale image domain. The proposed mapping strategy makes it plausible to correlate textural changes of the obtained images with seizure activities in EEG records. The textural analysis is then carried out using the well-known gray-level cooccurrence matrix (GLCM) and Haralick's feature extraction method resulting in a compact representation of EEG epochs. All the methods proposed in this thesis are evaluated using public EEG data-sets freely available on-line. The results obtained by the proposed methods are comprehensively compared with the other conventional dedicated feature extraction techniques using several classifiers.

The main contribution of the thesis is in the adaptation of conventional feature extraction techniques, commonly used in the textural analysis of images, to be applicable in EEG signal analysis. Additionally, the discriminatory power of feature descriptors is improved by representing EEG signals using their sparse rational components. The proposed rational local Gabor binary pattern (LGBP)-width feature outperforms competing methods in both seizure detection and classification problems. Moreover, its perform consistency

in patient/non-patient specific scenarios demonstrates its ability to tackle the curse of variability in seizures.

Preface

This dissertation is a result of research work conducted during the years 2013-2017 in the department of Signal Processing, Tampere University of Technology (TUT), Finland. The research was partly funded by the Finnish Funding Agency for Technology and Innovation, TEKES (currently Business Finland). The financial support is gratefully acknowledged.

First of all, I would like to express my sincerest gratitude to my supervisor Prof. Moncef Gabbouj for providing me the opportunity to work in his research group and for his encouragement and guidance throughout the duration of my thesis. I would also like to thank, Prof. Serkan Kiranyaz for his support during my work at Tampere University of Technology. It has been a pleasure working with the current and former colleagues in the at Multimedia Research Group as well as all the co-authors of the publications. Warm thanks to my office mates, Dr. Jenni Raitoharju, Guanqun Cao, and Honglei Zhang for the graceful working atmosphere and interesting discussions. I would like to specifically thank Dr. Péter Kovács, Dr. Alexandros Iosifidis, Dr. Stefan Uhlmann for their help and support.

Finally, I want to express my deepest gratitude and thanks to my wife Mehrnaz for all her support throughout this entire process. To my parents, my sister, I thank them for their continued encouragement.

Helsinki, April 2018

Kaveh Samiee

Contents

List of Figures

List of Tables

List of acronyms

Acc	accuracy
ADHD	attention deficit hyperactivity disorder
AEEG	ambulatory EEG
AI	artificial intelligence
AppEnt	approximate entropy
AUC	area under receiver operating characteristic curve curve
ANN	artificial neural network
BCI	brain-computer interface
BP	basis pursuit
BPD	basis pursuit denoising
CHB	children's hospital Boston
CNN	convolutional neural network
CP	cluster prominence
CWD	Choi – Williams distribution
CWT	continuous Wavelet transform
ECG	electrocardiogram
EDF	European data format
EEG	electroencephalogram
EMG	electromyogram
EMD	empirical mode decomposition
EOG	electrooculography
EPR	event – related potentials
FDFT	fractional discrete Fourier transform
FISTA	fast iterative shrinkage/thresholding algorithm

FFT fast Fourier transform

FN false negative

FP false positive

FPH false positives per hour

FPR false positive ratio

GLCM gray-level cooccurrence matrix

HG homogeneity

HHT Hilbert – Huang transform

HPSO hyperbolic particle swarm optimization

HR heart – rate

ICA independent component analysis

IIR infinite impulse response

KNN k – nearest neighbors

LBP local binary pattern

LGBP local Gabor binary pattern

Lin-SVM linear – support vector machine

Log-reg logistic – regression

MEG Magnetoencephalography

MT Malmquist – Takenaka

MIT Massachusetts Institute of Technology

MLP multi-layer perceptron

MRF modified rational base function

MSE mean square error

NREM non-rapid eye movement

PPG photo-plethysmogram

PPV positive predictive value

PSD power spectral density

PSG polysomnogram

PSO particle swarm optimization

PV prevalence

REM	rapid eye movement
RF	random forest
RFE	recursive feature elimination
ROB	rational orthogonal basis
ROC	receiver operating characteristic curve
RSTFT	rational short time Fourier transform
SALSA	split augmented Lagrangian shrinkage algorithm
SEM	slow eye movement
Sens	sensitivity
SGD	stochastic gradient descent
Spec	specificity
SNR	signal to noise ratio
STFT	short time Fourier transform
SVM	support vector machine
SWS	slow wave sleep
TDFDF	time-domain frequency-domain features
TP	true positive
TN	true negative
WT	Wavelet transform
WVT	Wigner – Ville transform

List of Publications

[**P1**] Péter Kovács, Kaveh Samiee, Moncef Gabbouj "On application of rational Discrete Short Time Fourier Transform in epileptic seizure classification", *IEEE International Conference on Acoustics, Speech and Signal processing (ICASSP), pp. 5839-5843, May 4-9. 2014.*

[**P2**] Kaveh Samiee, Péter Kovács, Moncef Gabbouj "Epileptic Seizure Classification of EEG Time-Series Using Rational Discrete Short-Time Fourier Transform", *IEEE Transactions on Biomedical Engineering, vol. 62, no. 2, pp. 541-552, Feb. 2015.*

[**P3**] Kaveh Samiee, Serkan Kiranyaz, Moncef Gabbouj, Tapio Saramäki, "Long-term epileptic EEG classification via 2D mapping and textural features", *Expert Systems with Applications, vol. 42, no. 20, pp. 7175-7185, Jun. 2015.*

[**P4**] Kaveh Samiee, Péter Kovács, Serkan Kiranyaz, Moncef Gabbouj, Tapio Saramäki, "Sleep stage classification using sparse rational decomposition of single channel EEG records", *IEEE 23rd European Signal Processing Conference (EUSIPCO), pp. 1905-1909, Aug. 2015.*

[**P5**] Kaveh Samiee, Péter Kovács, Moncef Gabbouj, "Epileptic seizure detection in long-term EEG records using sparse rational decomposition and local Gabor binary patterns feature extraction", *Knowledge-Based Systems, vol. 118, pp. 228-240, Feb . 2017.*

1 Introduction

Patient monitoring systems have been widely deployed to improve the quality of patient care. These systems can provide information needed to make intuitive decisions based on the physiological measurements of patients. In practice, online and real-time processing of psycho-physiological parameters of the human body such as heart rate, respiration rate and blood pressure is crucial in order to make responsive rational clinical decisions. Furthermore, long-term recording of physiological and biomedical signals make it feasible for diagnostic systems to predict or diagnose an anticipated phenomenon with a higher probability using the patient clinical history. Due to the multi-channel and multi-modality nature of physiological signals, the storage and the computational costs needed for processing big biomedical data were the biggest obstacles in utilizing predictive clinical models in the real world. Recently, new advances in distributed and parallel computing and big data analytics have closed this gap. However, in spite of the significant achievements in computer science and artificial intelligence over the past years, it is still challenging to utilize those scientific models in patient monitoring devices. This is due to specific requirements must be fulfilled by each patient monitoring system to be qualified for clinical purposes. One of the most important requirement is the low tolerance for false negative and false positive rates. Missing a potential clinical event may result in the death of the patient while a small increase in the number of false detections/predictions directly affects the quality of care-giving. Another challenge in the utilization of supervised machine learning models in healthcare systems is due to the limited access to annotated data. In fact, visual inspection of physiological signals is a tiresome and time-consuming task and requires the precise attention of clinicians.

Besides, it is essential to develop dedicated biomedical knowledge discovery and data mining models which demand low hardware resources and are able to be integrated into wearable devices. Processing of biomedical signals in embedded and wearable devices is mainly favored in patients with seizures and heart-related diseases where a prompt detection of a critical medical condition is crucial to reduce the fatality risk. On the other hand, computationally extensive models can be executed on a cloud network in order to perform a more detailed analysis. An example of such a system is long-term monitoring of people suffering from seizures. While seizures mostly are not considered hazardous, they cause risky medical problems during and after the occurrence which might cause the death of the patient. Thus, it is crucial in on-line monitoring to raise an alarm at the onset of the seizure. By learning from early seizure events in long-term recordings, it is feasible to predict the occurrence of a seizure prior to its onset.

EEG is the main modality that is commonly used to monitor electrical activity of the brain. Long-term recording of multi-channel EEG signals provides a cost-effective solution for different medical purposes such as diagnosis of epilepsy, retrospective study of neurological disorders like Alzheimer's disease and Parkinson's disease, evaluation of

sleep disorders and more recently for estimation of drivers drowsiness [1]. In non-invasive diagnostic systems, EEG is used in conjugation with other modalities in order to enhance the sensitivity and robustness of the system. Several studies show the effectualness of combining Magnetoencephalography (MEG) and EEG in brain activity monitoring systems and brain-computer interface (BCI) [2, 3, 4]. electrooculography (EOG) can be used for a better removal of eye movement artifacts in EEG channels [5]. Hybrid sleep scoring systems usually combine electromyogram (EMG) and EOG with EEG to improve the robustness of 24 sleep monitoring systems [6, 7]. Video-EEG monitoring is also a common and specialized protocol for pre-surgical evaluation and diagnosis of epilepsy [8].

EEG is still the main component, notwithstanding the number of the other auxiliary modalities utilized in a brain monitoring system. Specifically in the case of epilepsy, a monitoring system solely depended onEEG, is more feasible to implement in contrast to a video-EEG system due to privacy issues. Hence, it is very essential to devise new feature descriptors that can effectively represent signal activities in multi-channel EEG systems.

1.1 Objectives of the thesis

The main objective of this thesis is to address the problem of long-term EEG signals classification by employing RSTFT and other time-frequency representation techniques. For evaluation, two clinical problems of epileptic seizure detection and sleep stage classification are considered. In order to asset the usefulness of the proposed methods in practical real-world clinical scenarios, we only used a minimum amount information, selected from early record history of patient, for purpose of training. Consequently, the objectives of the thesis can be summarized as follows:

- To develop a time-frequency representation of EEG signals using rational functions and to compare the new transform with other conventional time-frequency transforms such as short time Fourier transform (STFT).
- To extend the proposed RSTFT by exploiting basis pursuit (BP) algorithm in order to induce sparsity into the transform coefficients and to come up with a sparse rational decomposition of EEG signals.
- To adapt and modify conventional feature extraction methods often applied in image processing domain and use them in multi-channel EEG signal analysis.
- To demonstrate the superior performance of the devised feature extraction methods in seizure detection and sleep stage classification of EEG signals compared to existing state-of-arts at the time of publication.

A time-frequency transform based on rational functions is proposed for processing EEG and electrocardiogram (ECG) signals. Compared to conventional STFT, the RSTFT is able to adjust itself according to the input, in order to achieve an optimum mean square error (MSE) of the inverse transform. The tuning of the RSTFT is obtained by finding the best combination and place of zeros and poles of the rational function. For this purpose, stochastic gradient descent methods such as particle swarm optimization (PSO) are employed to search for the optimum value of poles and zeros in the unit circle for each epoch of the input signal. Optimization of zeros and poles of the rational function for each epoch makes it feasible for the proposed transform to adapt itself to the morphology

and outlier of input. Experimental results show a lower MSE is achieved using RSTFT in the reconstruction of the input signal compared to conventional STFT. Flexibility and scalability of RSTFT permit exploitation of its combination with other signal processing and feature extraction techniques. A sparse rational decomposition of EEG signals is investigated by grouping the transform coefficients into a number of non-overlapping sets and inducing sparsity into the inverse transform. Furthermore, the application of RSTFT in the analysis of EEG signals is conducted by employing it as the preliminary stage for extraction of conventional morphological, statistical and time/frequency signal features.

The experiments demonstrate the superior discriminatory power of the features extracted from EEG signals represented using RSTFT. Modification and adaptation of existing feature extraction methods such as LGBP in conjugation with RSTFT results in state-of-art performance in epileptic seizure detection and classification of EEG signals. The experimental results show the effectiveness of the proposed feature extraction methods in off-line analysis and exploration of long-term EEG records. The proposed methods can be utilized in sleep profiling of patients or retrospective study of patients with epileptic seizures.

1.2 Thesis Outline

The thesis is organized as follows: the epileptic seizure detection/classification problem is described in Chapter 2. In addition, in this chapter we explain different EEG systems, characteristics of EEG signals and metrics that are used to measure the performance of a seizure detection/classification system. In Chapter 3, the main contribution of the thesis is introduced, namely, RSTFT and sparse RSTFT decomposition methods. Furthermore, the two other contributions of this study are described in Chapter 4. First, the proposed scheme for mapping EEG signals into gray-level image domain is described and next, the rational LGBP-width descriptor is introduced. Chapter 5, summarizes the results obtained using each method and compares the achieved performance with state-of-art methods. Finally, the thesis is concluded in Chapter 6.

1.3 Publications and Author's Contribution

[P1] This publication explores the application of rational functions in time-frequency analysis of epileptic seizure EEG time series. The idea of employing rational functions emerged in a discussion with Péter Kovács when he was a visiting researcher at the Multimedia Research Group in Autumn 2012. The RSTFT was mainly developed and implemented by Péter Kovács while the author helped in adapting it for EEG analysis, performing the experiments and contributing to writing the paper.

[P2] This publication extends the former version of Rational STFT. Here, the impact of the window's length and position of the pole of different rational systems on the performance of STFT were discussed in terms of the reconstruction MSE and classification accuracy. Experimental study was also extended by comparing the performance of the proposed RSTFT against other conventional time-frequency transforms including standard RSTFT and 13 Cohen's class distributions. Moreover, the robustness of the system was evaluated by considering different binary classification problems using time-series of Bonn data-set which is a benchmark for testing and comparing different EEG analysis methods. The candidate as the first

author contributed in designing the work, performing the experiments and writing the manuscript.

[**P3**] A novel method for 2D mapping of multi-channel EEG records into image gray-level domain is suggested in this publication which enables the textural analysis of epileptic seizures. The proposed method utilizes GLCM texture features in order to perform a crisp discrimination between seizure and seizure-free epochs. The results of a comparative study conducted using epileptic records of 23 patients of children's hospital Boston (CHB)-Massachusetts Institute of Technology (MIT) data-set demonstrate the effectiveness of the proposed approach for patient-specific classification of long-term epileptic EEG records. The author was responsible for developing the whole method including implementations, performing all experiments and writing the manuscript.

[**P4**] In this publication, sparse rational decomposition of EEG signals is proposed by inducing sparsity on RSTFT coefficients. The method is then deployed in sleep stage classification of single channel sleep EEG records. Experimental study was conducted using sleep records of 39 subjects of sleep-European data format (EDF) data-set and extracting the energy sum of each EEG sub-band. In total, 670 hours of sleep records were used for testing the performance of the method using random forest (RF) classifier trained in a patient-specific manner. The author is responsible for designing the work, implementing the algorithm and writing the manuscript. Péter Kovács contributed in checking the implementations and also helped in writing the manuscript.

[**P5**] In this publication, a 1D variant of LGBP feature extraction method is introduced to analyze epileptic EEG records. The proposed feature extraction method is then applied on EEG sub-bands obtained using the sparse rational decomposition RSTFT for performing non-patient-specific and patient-specific seizure detection on epileptic records of CHB-MIT data-set. The candidate as the first author is responsible for implementing the method, performing of experiments and writing the manuscript.

2 Background

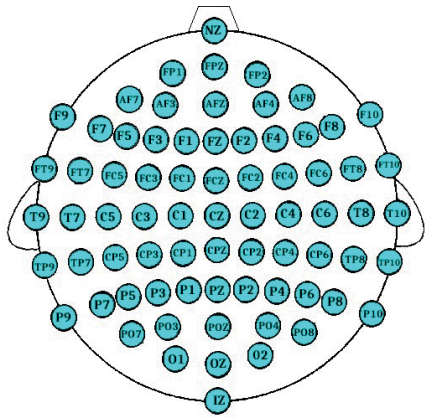
This chapter establishes the background concepts and provides preliminary materials that are used in this thesis and the publications. Since the main scope of this thesis is the off-line detection of seizures in long-term epileptic EEG records, this chapter describes topics related to EEG signals and their properties. Additionally, the problem of seizure detection using analysis of EEG signals is introduced and the metrics used to measure the performance of a seizure detection system are discussed.

2.1 Fundamentals of physiological signals

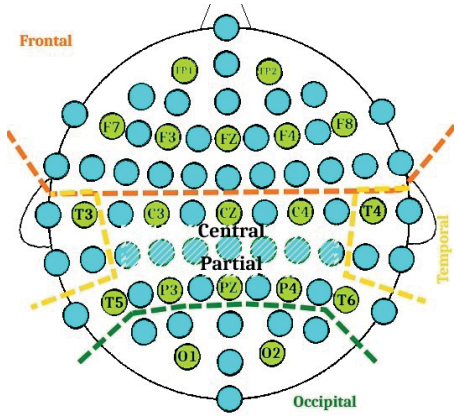
Physiological signals (also known as bio-signals) are the signals acquired from sensors and devices connected to a living body. These signals are used to continuously measure and monitor various physiological mechanisms. The behavior of a physiological system can be expressed using different physiological phenomena and hence a large number of electrical time-varying signals (e.g. EEG, ECG, EMG, EOG, etc) or non-electrical time-varying signals (e.g. temperature, pressure and movement) can be categorized under the term bio-signals. The dynamic of bio-signals varies according to the associated physiological phenomenon. For instance, thermal changes of the body are more steady than changes of heart−rate (HR). Additionally, for most of the bio-signals such as EEG and ECG, the source of excitation is inside the human body, while for some other bio-signals like photo-plethysmogram (PPG) an excitation must be induced from a source outside of the body.

2.2 Electroencephalogram signals

EEG is the prominent technique for monitoring and recording of brain activity. This technique uses electrodes placed on the scalp to measure the electrical activity of the brain. In non-invasive EEG systems, electrodes are placed or attached on the surface of the scalp while in the invasive ones the electrodes must be implanted or penetrated into the scalp via surgery. Electrodes record the voltage changes of the contact surface caused by neural activity over time. The number of electrodes and their montage on the scalp can directly affect the spatial resolution and brain surface coverage of an EEG system. The most common standards for placing electrodes are the international 10-10 and 10-20 systems. The distance and hence the density of the electrodes are determined according to the standard in use. For instance, in 10-20 system, numbers 10 and 20 indicate that the distance between the electrodes must be either 10% or 20% of the total front to back or right to left distance of the skull. Figure 2.1 shows the montage of electrodes in international 10-10 and 10-20 EEG systems. The letters Fp, F, C, O, T, P refer to Frontal partial, Frontal, Central, Occipital, Temporal and Partial representing different locations



(a) 10-10 EEG system



(b) 10-20 EEG system

Figure 2.1: Electrodes montage in different EEG standards

on the skull, respectively. The letter Z is used to indicate an electrode placement along the center. Additionally, even and odd numbers are used to indicate electrodes placement on right and left hemispheres.

2.3 Characteristics of EEG signals

EEG signal possesses specific characteristics which make the interpretation of these signals an intricate problem. It is essential to define these signal properties in order to develop and devise an effective technique for processing of EEG signals. Here, some of the most important EEG properties are briefly described.

2.3.1 Non-stationary and stochastic signal behavior

EEG signals are non-stationary which means their statistical parameters are changing over time. The main reason for the non-stationary behavior of EEG signals is due to brain neural activity which might not be in a coherent structure and thus neural charges/discharges of the same fraction of scalp change with different intensity levels over time, [9]. Nevertheless, different physiological states of the brain such as epilepsy or sleep stages change statistical characteristics of EEG signals in a more interpretive manner.

2.3.2 Low signal to noise ratio

EEG signals usually possess a low signal to noise ratio (SNR) due to the fact that electrodes conductivity on the scalp is affected by body movement, eye blinking, muscles activity or other dynamic changes in the environment.

2.3.3 Non-linearity of EEG signals

Although the human brain is a complex non-linear system, EEG signals generated from brain activity can be represented using as linear model. However, some researchers have shown that EEG signals fit better in non-linear models, [10].

2.4 Brain Waves and EEG signal activity

The oscillatory activity of the brain is a collection of the oscillatory behavior of single neurons in the brain [11]. Every single neuron has an intrinsic tendency to oscillate at multiple frequencies, [12]. The resonating frequency preference of different neurons varies in the range of 0.5-500 Hz, [11]. Brain waves are studied by associating their oscillatory behavior to a frequency sub-band centered at a mean frequency with a limited bandwidth. Each of these frequency sub-bands can represent a mind-state associated with certain brain activities, [13, 14, 15]. The first five frequency sub-bands are as follows:

- delta (0.5-4 Hz): is the lowest active frequency of the brain. Delta waves are mostly associated with deep and dreamless sleep, unconsciousness and deepest meditation. These rhythms are considered to be originated from deep inside the brain, [15].
- theta (4-8 Hz): is considered as the state of dreaming. Brain waves at this frequency range correspond to intuition and light meditation. Study of theta rhythms is useful in the diagnosis and treatment of depression, attention deficit hyperactivity disorder (ADHD), and learning difficulties, [16].
- alpha (8-12 Hz): brain waves in this frequency range are related to learning, concentrating and calmness. These rhythms mostly appear with sinusoidal shape over the occipital region of the brain.
- beta (12-30 Hz): beta frequency sub-band is associated with high level of anxiety and stress. It may be further divided into two sub-bands. Namely, low-beta (12-16 Hz) and hi-beta (15-30 Hz). Beta rhythms mostly appear at frontal and central portions of skull, [17].
- gamma ($\geq$ 30 Hz): is associated with perception and mental tasks related to the cognitive function of the brain. These rhythms, also known as fast-beta rhythms,

mostly appear with a low amplitude in EEG records representing information processing in the brain. Analysis of these rhythms is important in the study of neurological diseases or diagnosis of mentally challenged patients, [18].

2.5 Applications of EEG monitoring systems

EEG monitoring systems are widely used in clinical and commercial applications, and used for pathological purposes such as epilepsy and seizure detection systems as well as physiological state detection and classification such as sleep staging. EEG monitoring systems may also include other modalities such as video, ECG, and EOG. Such systems can be divided into two categories based on the length of recording. Namely, short-term monitoring where brain-activity of subjects is recorded from several minutes to few hours; and long-term monitoring where recording lasts continuously for several days. Short-term video EEG monitoring systems have a proven effectiveness in the diagnosis of pseudo-seizure symptoms, [19]. Non-video short-term EEG monitoring is the cheapest routine for obtaining EEG. However, due to a limited length and number of positive samples, it provides a lower sensitivity. Such system is still useful in the diagnosis of episodic and pseudo-seizure events when the recording trial is repeated multiple times for each subject during a day, [20]. EEG monitoring is tremendously growing in commercial devices and applications such as mind games, yoga, and meditation accessories, etc. In BCI, EEG signals are processed and translated into electrical or computer commands enabling the brain to control devices and interacting with a visual or an audio stimulus, [21]. In most recent years, EEG signals have been employed in music discovery and recommendation systems, [22], as a mean for measuring the arousal and valence levels. Long-term EEG monitoring systems are the main solution for diagnosis of epileptic seizures and other brain symptoms and disorders where a longer recording time is needed to capture positive events such as brain discharges and seizures. Video acquisition can significantly improve the sensitivity of diagnosis and thus long-term EEG-video monitoring is the gold standard for seizure treatment. However, in ambulatory EEG, the video monitoring may not be feasible due to privacy concerns, storage and processing costs, and more importantly, the relevance of captured events. This makes the interpretation of ambulatory EEG signals more challenging and thus it is essential to devise robust techniques for analysis of such signals. Since the main objective of this thesis is seizure detection and sleep stage classification, these scopes are briefly described next.

2.5.1 Seizure detection and classification

A seizure is a transient extravagant electrical discharge of neurons in a brain. This abnormal phenomena may occur partially or evolve generally in the brain. Video monitoring is the most reliable and popular technique for diagnosis of seizure, [23]. However, EEG is widely used for detection, prediction and treatment of seizures. Furthermore, exploring EEG epileptic records enables neurologists to identify seizure's type and its location on the brain. Ictal (seizure) events in EEG appears as low frequency abnormal rhythmic neuronal activities such as spikes which might be predominated with high-frequency oscillations and sudden changes in signal amplitude [24]. Before the onset of abnormal burst of brain electrical activities during a seizure, EEG signal activity is mostly dominated by slow-wave changes and low amplitude irregular spike discharges. Early diagnosis of a seizure, several hours before its onset, is possible by monitoring spike discharge activities and amplitude dynamics of slow-waves during pre-ictal (before seizure) events. However, the visual inspection and annotation of EEG signals are a time-consuming task that must

be conducted by an EEG expert. Hence, it is essential to develop automated seizure detection and prediction methods in order to improve clinical care level and diagnosis.

Epileptic EEG records are classified into three events stating the occurrence of the seizure. Namely, pre-ictal, ictal and post-ictal (after seizure) and inter-ictal (between seizures) events. In the off-line seizure detection applications such as ambulatory EEG (AEEG) monitoring where the aim is to determine seizures' onsets and offsets, EEG epochs can be categorized into two major classes as ictal and seizure-free.

2.5.2 Sleep stage classification

Having an adequate amount of sleep is essential for human health. Lack of sufficient sleep can increase the risk of several health problems such as diabetes and obesity. Additionally, sleep deprivation or poor quality sleep affects conscience and productivity at workplace. The quality of sleep is assessed according to sleep scoring standard of K&R, [25]. At the top level, sleep stage is divided into two major categories. Namely, rapid eye movement (REM) and non-rapid eye movement (NREM) stages. REM constitutes 20 to 25% of total sleep. During REM, the brain is active and most of the dream occur at this stage. Moreover, NREM contains 4 other states indicating the deepness of sleep. First, stage I which is a time of transition between wake-fullness and sleep. It represents 4 to 5% of the whole sleep and slow eye movement (SEM) mostly occurs at this stage. Second, stage II which is considered as the baseline sleep which constitutes 45 to 55% of the sleep. At this stage, awareness from surroundings and conscientiousness are significantly dropped. Third, stage III which is considered as deepest sleep cycle. Almost 20% of the whole sleep of adults consists of stage III where slow waves and delta waves (0.5-4 HZ) are visible in EEG. In this stage, the body is unengaged from outside environment, heart-rate and blood pressure drop and muscle are relaxed. Finally, stage IV, also known as slow wave sleep (SWS), is similar to stage III, but delta waves appear more than 50% of the time period in this stage.

2.6 EEG-driven seizure detection and classification

EEG signals are widely deployed in diagnosis and analysis of seizures. Neurologists visually examine EEG records in order to determine the occurrence of a seizure according to EEG signal shape and patterns changes over time. In order to imitate the visual inspection of neurologists, it is essential to characterize EEG signal activities in each of the inter-ictal, pre-ictal, ictal and post-ictal stages of epileptic EEG signal in the terms of morphology, amplitude and frequency features. Characterization of epileptic EEG signals makes it feasible for signal processing and machine learning techniques to discriminate between seizure and seizure-free epochs. An epileptiform activity in inter-ictal and ictal EEG refers to waves and signal patterns which are distinctive from the background activity. These epileptiform patterns include sharp-wave, spike, spike-slow-wave complex, multiple spike-slow-wave complex, [26]. Sharp-wave and spike patterns indicate a sudden and transient discharge significantly differs from the background. Both of these patterns appear with a sudden attenuation in amplitude of the EEG wave. Sharp waves usually last from 70 to 200 msec while spikes possess a narrower width (<70 msec) and a higher amplitude. Spikes are usually followed by a low voltage slow-wave. Sharp-waves are less localized than spikes and are less commonly followed by slow-waves. The combination of one or more spikes with one or more sharp waves forms spike-slow-wave complexes which are usually followed by a burst in delta frequency rhythms, [27]. The sharpness and

amplitude of these waves and complexes in contrast to their background are the main clues in identification and classification of epileptic records. Localization and type identification of seizures are also performed according to epileptiform patterns and discharges in EEG records. Figure 2.2 shows an epileptic EEG record in inter-ictal and ictal events.

2.7 EEG-driven sleep analysis

Generally, sleep quality assessment is usually performed manually by medical experts using the so-called polysomnogram (PSG) which contains biological signals of non-intrusive sensors such as EEG, EOG and EMG. In practice, each sleep stage can be characterized by a specific EEG activity. For instance, the dominant EEG activities are discharges and spikes in stage I (drowsiness), K-complexes in stage II (light sleep) and delta waves in stages III and IV (deep/very deep sleep).

Several automated sleep classification systems have been proposed over the last decade, imitating the visual inspection of a medical expert. Many of these algorithms utilize time-frequency analysis of the EEG signal to extract descriptive features for sleep scoring. Decomposition of sleep EEG signal into primary frequency sub-bands via 8 levels Wavelet transform (WT) was first introduced in [28]. In this case, 13 features are extracted based on the energy of each sub-band which was used to train a feed-forward artificial neural network (ANN) using the back-propagation algorithm. Another decomposition technique was proposed in [29] based on the Hilbert – Huang transform (HHT) which uses empirical mode decomposition (EMD) to decompose the EEG signal into 7 sub-bands. Similarly, features were extracted based on the energy of the sub-bands. In a more recent study, [30], convolutional neural network (CNN) fed with 2-dimensional vectors constructed for each EEG epoch using WT and fractional discrete Fourier transform (FDFT) of each EEG channel, was used to automatically classify sleep stages.

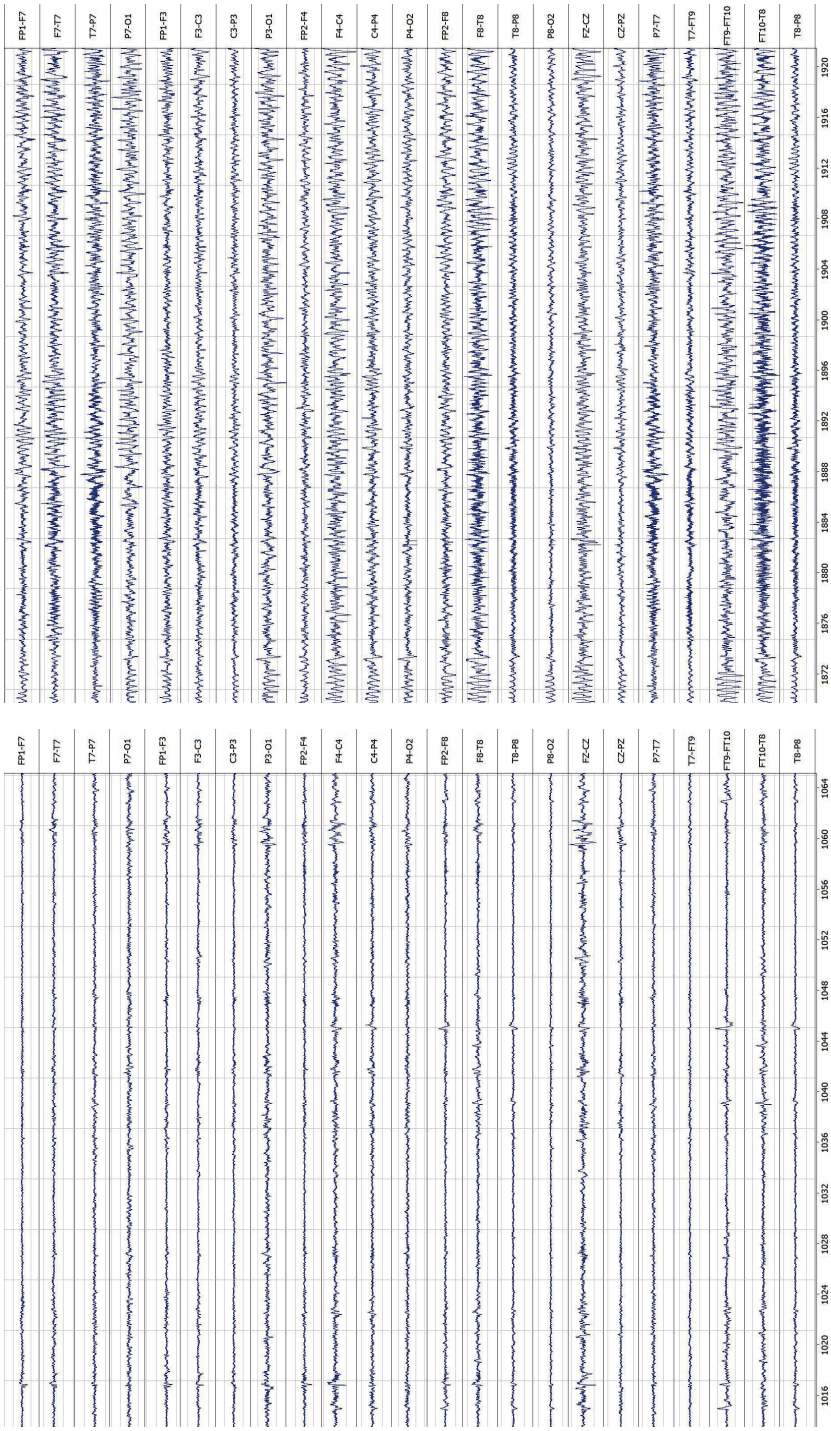
2.8 Qualitative assessment of EEG detection systems

Qualitative evaluation of an EEG system explores all observations in order to define a set of heuristics by finding correlations between features presence or variations during the target event. An example of such an observation is the significant variations in approximate entropy (AppEnt) value between seizure-free and ictal epochs amending detection of seizures by monitoring its trend. Beside the statistical analysis of variables and features derived from EEG channels, qualitative evaluation of a system performance can include preliminary or retrospective questionnaire of patients or caregivers.

2.9 Quantitative assessment of EEG detection systems

Quantitative evaluation of EEG systems is carried out by transforming heuristics from qualitative diagnosis to a classification problem of the target event. In the binary classification of epileptic EEG records, let us denote seizure-free epochs as class **0** and ictal epochs as class **1** indicating negative and positive samples, respectively. Then, for each EEG epoch, the prediction belongs to one of the following categories:

- *truepositive*(TP): a positive sample predicted as positive,
- *truenegative*(TN): a negative sample predicted as negative,



(a) inter-ictal epileptiform patterns

(b) ictal epileptiform patterns

Figure 2.2: epileptiform abnormalities from record 26 of patient 1, CHB-MIT dataset

- $falsepositive$(FP): a negative sample predicted as positive,
- $falsenegative$(FN): a positive sample predicted as negative.

2.9.1 Sensitivity rate

Sens or TP rate represents the accuracy of an EEG monitoring system in detection or classification of positive events. It is defined as:

$$\text{Sens} = \frac{\text{TPs}}{\text{TPs} + \text{FNs}}\,, \tag{2.1}$$

where in the case of seizure event detection, TPs represent seizure events that are correctly detected by the system. Sens is also called as recall.

2.9.2 Specificity rate

Spec or TN rate represents the accuracy of an EEG detection system in differentiating of non-ictal events. It is defined as:

$$\text{Spec} = \frac{\text{TNs}}{\text{TNs} + \text{FPs}}\,. \tag{2.2}$$

2.9.3 False positive rate

False positive rate (FPR) is the probability of missing a seizure event. It is defined as:

$$FPR = \frac{\text{FPs}}{\text{TNs} + \text{FPs}}, \tag{2.3}$$

where FNs represent ictal events that are not detected by the system.

2.9.4 Classification accuracy rate

Classification accuracy (Acc) represents the relationship between Sens and Spec rates. it is defined as:

$$Acc = \text{sensitivity} \times \text{prevalence} + \text{specificity} \times (1 - \text{prevalence})\,, \tag{2.4}$$

where prevalence (PV) is the ratio of positive samples in the test population:

$$PV = \frac{\text{TPs} + \text{TNs}}{\text{TPs} + \text{FNs} + \text{TNs} + \text{FPs}}\,.$$

2.9.5 Positive predictive value

Positive predictive value (PPV) or precision indicate portion of predicted positives (ictal epochs, seizures, etc.) that are correctly positives.

$$PPV = \frac{\text{TPs}}{\text{TPs} + \text{FPs}}\,. \tag{2.5}$$

2.9.6 False positives per hour

In the case of a seizure detection system, (FPH) is defined as the number of false alarms (seizure event detections) per hour. In the field of patient monitoring, FPH is one of the key parameters reflecting the performance of the monitoring device. This parameter must be kept as low as possible since it can directly affect the care level in hospitals.

2.9.7 Area under curve

The area under a receiver operating characteristic curve (ROC) curve is commonly used to measure the discriminatory power of a model. In a classification problem, a ROC curve can be plotted for all Sens and the corresponding FPR values obtained by varying decision threshold. area under curve (AUC) considers similar weights for both Sens and FPR values. AUC of 100% indicates a perfect performance while a AUC of 50% indicates that the performance is not better than a random discrimination.

2.9.8 F1-measure

F1-measure is the harmonic mean between recall and precision and is defined as:

$$F1 = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \,. \tag{2.6}$$

In classification problem such as seizure detection where the classes' distributions are imbalanced, F1-measure can provide a better metric for performance benchmarking of the system in contrast to Acc.

2.10 EEG Datasets

The experiments in the thesis are carried out using several EEG datasets. The details of each dataset and the purpose of its usage is discussed in this section.

2.10.1 Bonn University EEG database

Benchmark EEG database provided by University of Bonn is used for classification of EEG time-series, [31]. The EEG database is freely available online and consists of five sets (A-E). Each set contains 100 segments of single channel EEG with length of 23.6s selected from multi-channel EEG signals. To remove major artifacts caused by eye blinking or muscle activities a visual inspection was performed on the EEG data. Moreover, EEG segments must satisfy a weak stationary criterion. Sets (A) and (B) have been recorded using the standard international 10-20 system for surface EEG recording. Five healthy volunteers were participated in these stets with eyes open and eyes closed in (A) and (B), respectively. For sets (C), (D) and E five epileptic patients were selected for pre-surgical evaluation of epilepsy patients using intracranial electrodes. Depth electrodes were implanted symmetrically to record within the epileptogenic zone (D) and in (C) from hippocampal formation of the opposite hemisphere of the brain. Segments of set E were taken from contacts of all electrodes. In sets (C) and (D), segments contain inter-ictal intervals while seizure activities occur in segments of set (E). All records were acquired using the same the same 128-channel amplifier and each segment was digitalized using a 12-bit analog-to-digital converter at sampling frequency of 173.61 Hz resulting in 4096 samples. Figure 2.3 shows examples of different sets in Bonn dataset.

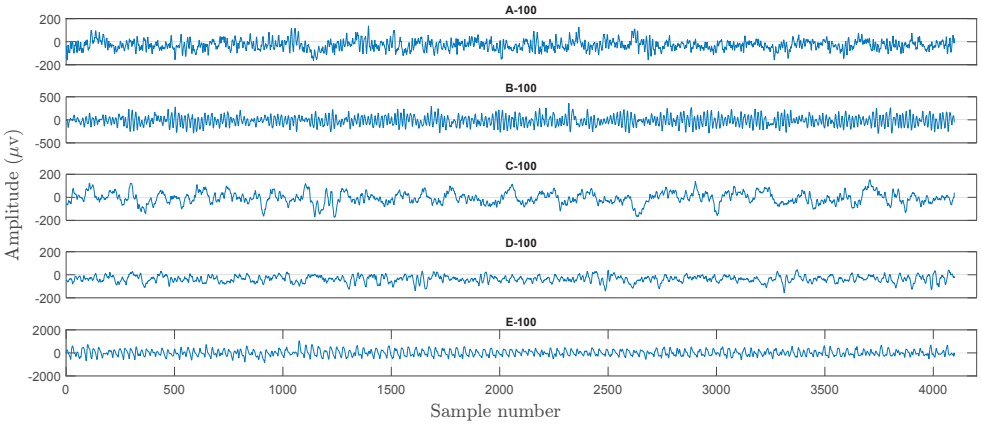


Figure 2.3: record 100 of each EEG set in Bonn dataset

2.10.2 CHB-MIT Scalp EEG Database

For the epileptic seizure detection problem, We use CHB-MIT scalp EEG database which is collected from Children's Hospital in Boston from 23 pediatric subjects with intractable seizures and is publicly available at PhysioNet, [32]. In order to characterize seizures, brain activity of each subject was monitored, after a drug withdrawal, for up to several days in a 10-20 international system. All EEG signals were digitalized with 16-bit resolution with a sampling rate of 256 Hz. In total, the database consists of up to 916 hours of EEG recordings containing 198 seizure events. EEG data for each subject has been segmented into one to four hours long EEG records. In this study, we only collect those segments containing one or more seizure events to reduce the imbalanced ratio of seizure to seizure-free epochs. Furthermore, to imitate the visual inspection of a neurologist, 25% of EEG data of each subject, selected from the beginning of the recording, is used as the training set while the rest is used for testing. The information about each patient and duration of the training and test sets is summarized in Table. 2.1. In total, there are 23 to 26 bipolar EEG channels in each record. We note that patient 12 is discarded owing to the variation of electrodes montage in its records. Additionally, EEG data in experiment 24 was added later to the database with an unknown subject identity and thus we consider it as a separate subject in this study. Hence, the patient cohort contains 5 male, 17 female and 1 unknown subjects with ages ranging from 1.5 to 22 year-old. In all experiments, the power-line frequency of 60 Hz was removed from EEG signals using a notch filter. In addition, a Butterworth bidirectional infinite impulse response (IIR) low-pass filter of order 5 with a cutoff frequency of 42 Hz is applied on each channel in order to discard higher frequency components and muscular artifacts. Since no further prepossessing step on EEG records is performed, EEG signals in the training and the test sets may contain activities such as eye blinking and patient movements in both seizure and seizure-free classes. Figure 2.4 shows an epileptic EEG record from CHB-MIT dataset.

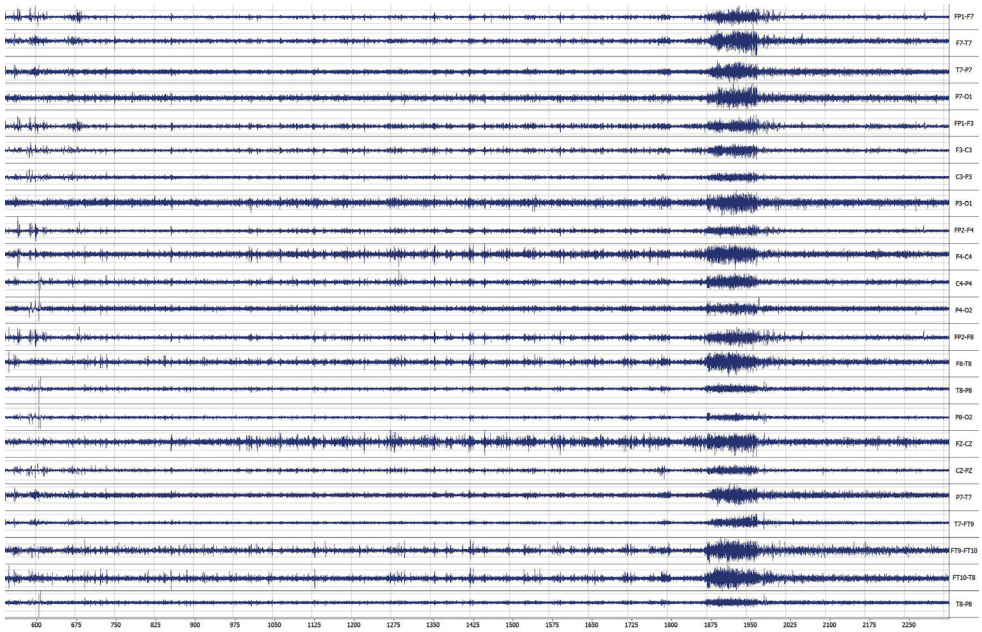


Figure 2.4: EEG record from record 26 of patient 1, CHB-MIT dataset

Table 2.1: Training and test sets used for each patient **[P3]** **[P5]** © 2015 Elsevier

				Training set		Test set	
Patient ID	Sex-Age	No. Seizures	Seizure length (sec)	Seizure (sec)	Seizure-free (sec)	Seizure (sec)	Seizure-free (sec)
1	F-11	7	442	111	5869	331	17612
2	M-11	2	91	23	1117	68	3351
3	F-14	7	402	101	6198	301	18598
4	M-22	4	379	95	9495	284	28485
5	F-7	5	558	140	4360	418	13081
6	F-1.5	7	156	39	23265	117	69795
7	F-14.5	3	325	82	8053	243	24159
8	M-3.5	5	919	230	4271	689	12810
9	F-10	4	277	70	8556	207	25666
10	M-3	6	382	96	10716	286	32148
11	F-12	3	806	202	2314	604	6939
13	F-3	12	539	135	7066	404	21195
14	F-9	8	170	43	6258	127	18772
15	M-16	20	1998	500	12109	1498	36327
16	F-7	10	88	22	5378	66	16134
17	F-12	3	293	74	2633	219	7898
18	F-18	6	317	80	4990	237	14967
19	F-19	3	236	59	2578	177	7732
20	F-6	8	296	74	4935	222	14805
21	F-13	4	199	50	3398	149	10193
22	F-9	3	204	51	2649	153	7947
23	F-6	7	428	107	7957	321	23869
24	-	16	519	130	10671	389	32010

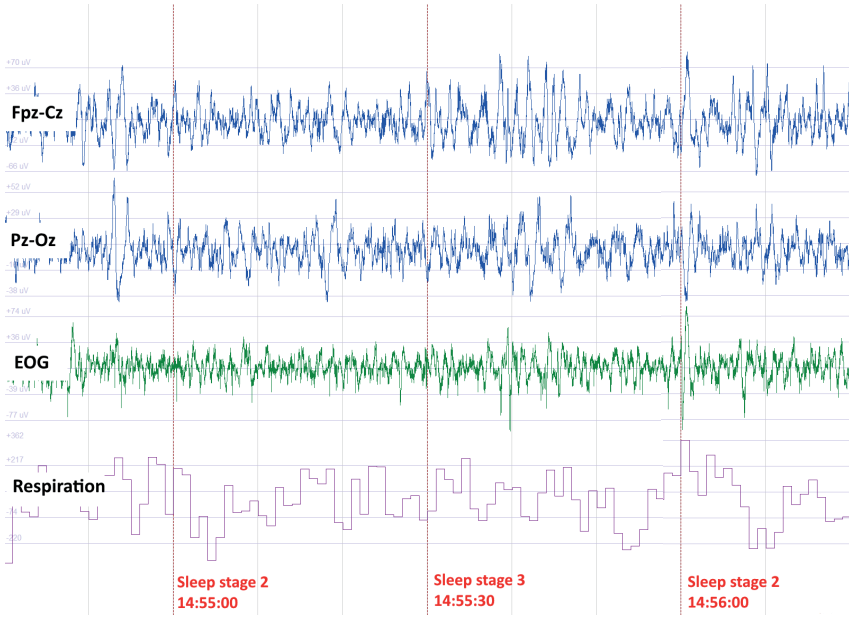


Figure 2.5: A snapshot of the sleep record of subject SC4112E0 from Sleep-EDF dataset

2.10.3 Sleep-EDF Database

The public sleep-EDF database is used as the experimental data for sleep stage detection task. The database is a part of the PhysioNet data bank, [32]. Sleep records were obtained from two different groups of subjects. First group contains 79 healthy Caucasians subjects aged 25-101, without any sleep-related medication. The second group of subjects had sleep difficulty and were under influence of temazepam medication during the recording. In this study we only consider the first group of subjects. Each sleep record was obtained using different modalities including: one horizontal EOG, two EEG channels (Fpz-Cz and Pz-Oz), sub-mental EMG envelope, oronasal airflow, and rectal body temperature. EEG and EOG signals were sampled at 100 Hz. Additionally, all records were annotated by trained experts and start and stop time of each different sleep stages were marked in a separate annotation file for each record using eight letters. These scoring letters are namely, **1**, **2**, **3**, **R** (rapid eye movements), **W** (wakefulness), **M** (movement time) and **?** (undefined). The frontal channel (Fpz-Cz) of the EEG record in this study to detect sleep stages. A snapshot of a sleep record from the dataset presenting 2 EEG channels, one frontal EOG channel, respiration rate and annotations regarding onset of sleep stages can be seen in Figure 2.5.

3 Time-Frequency representation of biomedical signals

In this section, the theory behind STFT and rational functions is briefly described. Furthermore, the generalized RSTFT is proposed as a variant of STFT leveraging rational functions in order to obtain an adjustable time-frequency representation of biomedical signals. In addition, a signal processing decomposition model is introduced based on the generalized RSTFT which is then deployed in Chapter 4 as the preliminary step for EEG signal modeling and feature extraction.

3.1 Short time Fourier transform

Fourier transform is a well-known tool for analyzing the frequency distribution of a signal. Unfortunately, the time information has been lost during this transformation. As a consequence, sudden changes of the signal cannot be localized in time. However, there is a wide range of applications where both information is necessary. On the other hand, several techniques like STFT, Wigner – Ville transform (WVT) and Choi – Williams distribution (CWD), are attempted to solve this problem. Most of these algorithms operate on shorter segments of the signal which makes it possible to localize the frequencies in time. On the other hand, there is a huge tradeoff between the resolutions of the domains. It means that any attempt to increase the time resolution causes a decrease in the frequency resolution. The well-known continuous Wavelet transform (CWT) gives a good solution for this issue. In the case of CWT the frequency domain is logarithmically scaled by dynamically changing the window size. So, the lower frequencies have fine frequency distribution while the higher ones are well localized in time. Nevertheless, the wavelet basis is not an adaptive system. Hence, the required number of coefficients can be high for achieving an acceptable level of accuracy.

Let $\mathbb{C}$ stand for the set of complex numbers and $\mathbb{D} := \{ z \in \mathbb{C} : |z| < 1 \}$ for the open unit disc. Having a signal f defined in the space of square integrable functions on $\mathbb{D}$ ($f \in L^2(\mathbb{D})$), STFT of signal f is defined as

$$\mathcal{F}_g f(t, \omega) = \int_{\mathbb{R}} f(\tau - t)\overline{g}(\tau) e^{-2\pi i \tau \omega} d\tau \quad (t, \omega \in \mathbb{R}), \tag{3.1}$$

where $\overline{g}$ is the complex conjugate of $g \in L^2(\mathbb{D})$. In practical applications g is compactly supported when it is called as window function. In this case the integral above can be evaluated over the support of g. Furthermore, this algorithm can be interpreted as a successive evaluation of Fourier transforms over short segments of the whole signal. Additionally, the frequencies can be visually represented by displaying the squared magnitude

of the Fourier coefficients at each section. This diagram is called the spectrogram of the signal f.

Let us denote the uniformly sampled $f(t)$ and $g(t)$ functions by $f[n]$ and $g[n]$. Then the discrete (D) STFT over a compactly supported g window function can be written as

$$\mathcal{F}_g^{\epsilon} f[n,k] = \sum_{m=0}^{M-1} f[m-n]\overline{g}[m]\overline{\epsilon}_k[m] \quad (n \in \mathbb{N}), \tag{3.2}$$

$$f[m-n] \approx \frac{1}{M\overline{g}[m]} \sum_{k=0}^{M-1} \mathcal{F}_g^{\epsilon} f[n,k]\epsilon_k[m] \quad (n \in \mathbb{N}), \tag{3.3}$$

where $\epsilon_k[m] = e^{-2\pi m \frac{k}{N}}$, M is the window length of g and N is the number of samples in signal f.

3.2 Rational functions

Let $\mathbb{N} := \{1, 2, 3, \dots\}$ and $\mathbb{T} := \{ z \in \mathbb{C} : |z| = 1 \}$ stand for the set of natural numbers and the torus, respectively. The basic rational functions are defined as follows

$$r_{a,k}(z) = \frac{1}{(1-\overline{a}z)^k}, \qquad (a \in \mathbb{D}, k \in \mathbb{N}). \tag{3.4}$$

The parameter a is referred to as *inverse pole* (because $1/\overline{a}$ is a pole in the standard sense), k is said to be the *order* of the basic function. Using a terminology similar to the trigonometric case, the value $k = 1$ corresponds to the *fundamental tone* and $k > 1$ the *overtones*. Furthermore, it can be easily shown that any function f, that is analytic on the closed unit circle, can be represented by an infinite linear combination of basic rational functions. Namely,

$$f = \sum_{k=0}^{\infty} c_k r_{a,k}, \qquad (a \in \mathbb{D}), \tag{3.5}$$

for an appropriate set of complex coefficients $c_k \in \mathbb{C}$. In the case of real signals we will use the truncated sum of Eq. (3.5) to approximate them as a real valued function. Additionally, we will restrict the $r_{a,k}$ elementary waves onto the unit circle such that

$$[-\pi, \pi) \ni t \to e^{it} \to \mathrm{Re}\big(r_{a,k}(e^{it})\big). \tag{3.6}$$

However, the basic rational functions $\{ r_{a,k} : k \in \mathbb{N} \}$ are linearly independent, but do not form an orthogonal set, so it is difficult to compute the c_k coefficients in Eq. (3.5). On the other hand, we can easily solve this problem by applying Gram–Schmidt orthogonalization to the basic rational functions. The corresponding rational function system is the so-called Malmquist–Takenaka (MT) system. A handy property of the MT system is that the elements can be expressed as Blaschke products. Namely, taking the basic functions for a given $n \in \mathbb{N}$ and the sequence of inverse poles $a_0, \dots, a_n \in \mathbb{D}$ the orthogonalized MT system can be written as:

$$\Phi_k(z) = \frac{\sqrt{1-|a_k|^2}}{1-\overline{a}_k z} \prod_{j=0}^{k-1} B_{a_j}(z), \tag{3.7}$$

with $0 \le k \le n$, where $B_a(z)$ is the so-called Blaschke function defined by

$$B_a(z) := \frac{z-a}{1-\overline{a}z} \qquad (z \in \mathbb{C} \setminus \{\, 1/\overline{a} \,\}) \,. \tag{3.8}$$

Although, now we have an orthonormal set of functions, the time localization property of the basic rational form has been lost. Fortunately, the biorthogonal rational functions are attempted to treat this issue by keeping the orthogonality but avoiding the bad properties of the MT system. This kind of biorthogonal system can be defined by taking $n+1$ different inverse poles $a_0, \ldots, a_n$ with multiplicities $m_0, \ldots, m_n$ and the corresponding modified rational base function (MRF)s

$$\varphi_{k,i}(z) = \frac{z^{i-1}}{(1-\overline{a}_k z)^i} \qquad (k = 0, \ldots, n,\ i = 1, \ldots, m_k) \,. \tag{3.9}$$

The system of $r_{a,k}$ and $\varphi_{a,k}$ span the same subspaces of the functions analytic on the closed unit circle which is denoted by $\mathfrak{R}$. For the definition of the biorthogonal system we will need the following functions

$$\Omega_{\ell n}(z) = \frac{1}{(1-\overline{a}_\ell z)^{m_\ell}} \prod_{i=0, i\neq\ell}^{n} \left(\frac{z-a_i}{1-\overline{a}_i z} \right)^{m_i} , \tag{3.10}$$

$$\omega_{\ell n}(z) = \frac{\Omega_{\ell n}(a_\ell)}{\Omega_{\ell n}(z)} \tag{3.11}$$

where $(0 \le \ell \le n)$.

By Theorem 1 in [33] the functions

$$\Psi_{\ell,j}(z) = \frac{\Omega_{\ell n}(z)(z-a_\ell)^{j-1}}{\Omega_{\ell n}(a_\ell)} \sum_{s=0}^{m_\ell - j} \frac{\omega_{\ell n}^{(s)}(a_\ell)}{s!} (z-a_\ell)^s \tag{3.12}$$

$(0 \le \ell \le n,\ 1 \le j \le m_\ell)$ are biorthogonal to $\varphi_{k,i}$ with respect to the scalar product defined as follows

$$\langle F, G \rangle = \frac{1}{2\pi} \int_{-\pi}^{\pi} F(e^{it}) \overline{G}(e^{it}) \, \mathrm{d}t \quad (F, G \in H^2(\mathbb{D})) \,. \tag{3.13}$$

where $H^2(\mathbb{D})$ denotes the Hardy space of square integrable functions that are analytic on the open unit circle $\mathbb{D}$. More precisely,

$$\langle \Psi_{\ell r}, \varphi_{ks} \rangle = \delta_{k\ell}\delta_{rs} \,, \tag{3.14}$$

where $(1 \le r \le m_\ell,\ 1 \le s \le m_k,\ 0 \le k, \ell \le n)$ and δ_{ij} is the well-known Kronecker delta symbol.

As a result of the previous statements the MT and the biorthogonal expansions of an $f \in \mathfrak{R}$ function can be easily calculated as follows

$$\begin{aligned} P_\Psi^N f &= \sum_{k=0}^{n} \sum_{i=1}^{m_k} \langle f, \Psi_{ki} \rangle \, \varphi_{ki} \,, \\ P_\Phi^n f &= \sum_{k=0}^{n} \langle f, \Phi_k \rangle \, \Phi_k \,, \end{aligned} \tag{3.15}$$

where $N = m_0 + m_1 + \cdots + m_n$. This can be interpreted as a projection onto the N or n dimensional subspace of $\mathfrak{R}$.

The advantages of using rational function systems over other transformation methods are as follows

- flexibility in the sense that not only the coefficients but also the system itself can be varied which means the system can be adapted to the EEG signal;
- the coefficients give a compressed representation of the signal, so in further processing steps they can be used as a feature;
- elementary waves are localized in time and basic functions can carry time-frequency information;
- this is a simple analytic representation of the original signal, which means the whole army of analytic tools can be exercised on the representation;
- only a couple of arithmetic operations are required to recover the signal.

We note that the MT and the biorthogonal systems Φ and Ψ with the basic rational functions φ are referred as the rational orthogonal basis (ROB) in the literature. The construction of these generalized orthogonal basis was introduced by [34].

3.3 Rational Discrete Fourier Transform

Using the same terminology as in Eq. (3.2) we can define a similar representation of the signal by replacing the trigonometric bases ϵ_k with the elements of the ROB. More precisely, let us consider a single inverse pole a_0 with multiplicity $m_0 = M$, and an $f \in H^2(\mathbb{D}))$ uniformly sampled function. Then the generalized RSTFT can be written as

$$\begin{aligned}
\mathcal{R}_{\psi}\mathcal{F}_{g}f[n,k] &= \sum_{m=0}^{M-1} f[n-m]\overline{g}[m]\psi_k[m]\,, \\
\mathcal{R}_{\varphi}\mathcal{F}_{g}f[n,k] &= \sum_{m=0}^{M-1} f[n-m]\overline{g}[m]\varphi_k[m]\,, \\
\mathcal{R}_{\phi}\mathcal{F}_{g}f[n,k] &= \sum_{m=0}^{M-1} f[n-m]\overline{g}[m]\phi_k[m]\,,
\end{aligned} \tag{3.16}$$

where $\psi_k[m] = \Psi_{0,k+1}(e^{-2\pi\frac{m}{M}}), \varphi_k[m] = \varphi_{0,k+1}(e^{-2\pi\frac{m}{M}})$ and $\phi_k[m] = \Phi_k(e^{-2\pi\frac{m}{M}})$.

3.4 Sparse rational decomposition

In case of RSTFT, we use the first N coefficients in each window to represent the original EEG signal f. Note that the mth sample of the nth segment $f[m-n]$ can be approximated by using Eq. (3.3). In order to minimize the MSE of this rational approximation, we applied the hyperbolic variant of hyperbolic particle swarm optimization (HPSO), [35]. Hence, we get an optimal inverse pole and coefficient vector $\mathbf{a_o^n} := (a_0^n, \ldots, a_{N-1}^n) \in \mathbb{D}^N$ and $\mathbf{c_o^n} \in \mathbb{C}^N$ where n denotes the index of the corresponding segment and $c_k^n := \mathcal{F}_g^{\phi} f[n,k]$ for $0 \leq k < N$. Each coefficient of the RSTFT spectra points to a specific frequency

range in the t-f domain. More precisely, coefficients with larger magnitudes indicate the dominant signal activity in a specific frequency range. In order to decompose the signal into these sub-bands, we first sort the coefficients in a descending order of magnitude. Then, we use the set of rearranged rational components $S^n := \{c_k^n \Phi_k^n : 0 \leq k < N\}$, which is related to the nth segment. This is followed by partitioning S^n into L distinct subsets, which contain the components of the corresponding sub-bands. Thus, the ith sub-band of the nth segment for $0 \leq i < L$ can be defined as,

$$f_i[m-n] = \frac{1}{M\overline{g}[m]} \sum_{k=i\ell}^{(i+1)\ell-1} c_{\sigma(k)}^n \Phi_{\sigma(k)}^n \,, \tag{3.17}$$

where $\ell = N/L$ and σ denotes the permutation of the indices corresponding to the rearrangement of the coefficients. Samples of the reconstructed signal can be obtained by the sum of these sub-bands.

In order to induce the sparsity constraint to the RSTFT decomposition in Eq. (3.17), we employ the well-known BP algorithm, [36]. The BP convex optimization problem is as follows:

$$\min_x \|x\|_1 \quad \text{subject to} \quad Ax = b \,, \tag{3.18}$$

where b is an univariate signal, A is an over-complete dictionary and x is the coefficient vector of the transform. Additionally, the basis pursuit denoising (BPD) as a variant of the original BP problem can be obtained as,

$$\min_x \frac{1}{2} \|b - Ax\|_2^2 + \lambda \cdot \|x\|_1 \,, \tag{3.19}$$

where $\lambda > 0$ is the so-called regularization parameter.

In case of RSTFT, A is the matrix whose columns are the synthesis functions Φ_k of the transform, x contains the coefficients c_k^n and b is the nth segment of the EEG signal. The optimization problem in Eq. (3.19) can be solved by using the fast iterative shrinkage/thresholding algorithm (FISTA), [37]. Here, we applied the split augmented Lagrangian shrinkage algorithm (SALSA), which was proven to converge faster than FISTA and other alternative algorithms, [38]. In what follows, we will use a single inverse pole $a_0 \in \mathbb{D}$, which is repeated N times and optimized for each epoch. Then, the corresponding coefficients of each segment will be partitioned into L distinct subsets. Based on the experiments, we found that setting $N = 64$ coefficients and $L = 8$ sub-bands is a good trade-off between performance and computational complexity. We will also utilize the results of our previous work, **[P2]**, where we showed that the optimal window size is $M = 256$.

Although EEGs are real valued signals, the inverse RSTFT, i.e., the right hand sides of Eqs. (3.3)-(3.17) are complex functions. In order to resolve this problem, we will use the analytic signal $F := f + i\mathcal{H}f$, where $\mathcal{H}$ denotes the well-known Hilbert transformation. In this case, the real parts of the right hand sides in Eqs. (3.3)-(3.17) approximate the real part of F, which is equal to the original signal f.

4 Signal analysis of long-term EEG records

In this chapter, first, the general categories of EEG feature extraction methods are described. Then, in Section 4.2.1, the proposed rational LGBP-width feature extraction technique is explained. Finally, in Section 4.3, the details of the proposed scheme for mapping EEG signals into 2D gray-level image domain are described.

4.1 Characterization of multi-channel EEG signals

A large set of feature extraction techniques have been proposed in the literature for characterization EEG signals in multi-channel monitoring/recording systems. In general, three feature types can be extracted from EEG data:

1. uni-variate features that are computed from each monopolar/bipolar channel,
2. bi-variate features that are extracted from each channel pair
3. multi-variate features that are associated with two or more channels from different brain regions.

The uni-variate class includes most of the conventional feature extraction techniques which are used to characterize epileptiform waves. This class can be further divided into three subcategories based on the characteristic which is identified by the feature. These three categories are:

1. Chaosity or randomness measures,
2. Morphological features,
3. Frequency-domain features.

4.1.1 Chaosity analysis

Chaosity analysis of EEG signals can help identify different epileptiform patterns since an EEG is considered a non-stationary signal. Entropy as a measure of randomness is the most conventional statistical measure to analyze chaosity. Entropy of a finite signal $X = [x(0), x(1), ..., x(N-1)]$ is defined as:

$$Entropy = -\sum_{i=0}^{N-1} P(x_i) \log_2 P(x_i)\,,$$

where $P(x_i)$ is the probability density function of X.

AppEnt, [39], was first introduced for regularity measurement of heart-rate trends and then found a wide application in chaosity analysis of EEG signals. Let $\mathbb{U}$ denote a set of real numbers, $\mathbb{U} = [u(1), u(2), \ldots, u(N)]$ which are equally spaced in time data points of vector X and $\mathbf{x}(i) \in \mathbb{R}^m := \{\, u(i), u(i+1), \ldots, u(i-m+1) \,\}$ denote a sequence of vectors where m is a positive integer. Let us define a parameter $C_i^m(r)$ such that:

$$C_i^m(r) = (number \quad of \quad j \quad such \quad that \quad d\,|x(i), x(j)| \leq r)/(N-m+1)\,, \tag{4.1}$$

where r is a positive real number and d denotes the maximum difference between corresponding points in $x(i)$ and $x(j)$ and is defined as follows:

$$d\,|x(i), x(j)| = \max_{k=1,2,\ldots,m} (|u(i+k-1) - u(j+k-1)|)\,. \tag{4.2}$$

Using (4.1) and defining a statistical parameter $\Phi^m(r)$ such that:

$$\Phi^m(r) = (N-m+1)^{-}1 \sum_{i=1}^{N-m+1} \log C_i^m(r)\,, \tag{4.3}$$

AppEnt is defined as follows:

$$AppEnt(m, r, N) = \Phi^m(r) - \Phi^{m+1}(r)\,. \tag{4.4}$$

It was shown that the AppEnt value of EEG signals dramatically decreases during epileptic seizure epochs and thus it can be used to discriminate between ictal and inter-ictal patterns in an automatic seizure detection system. In addition, it is demonstrated that entropy features possess lower values during pre-ictal and ictal epochs due to the periodicity of patterns in the discharge of neurons. Moreover, it was observed that entropy values slightly increase during an ictal event. Due to this fact, entropies as statistical measures can also help in differentiating between pre-ictal and ictal epochs, [40].

4.1.2 Morphological analysis

Morphological analysis of EEG aims at the characterization of EEG signal waves and spike discharges using standard first-order statistics of waveforms and spikes such as amplitude, shape, width, and duration. EEG signal changes are then detected by splitting them into several parts containing background and current signal states. It was shown in a recent study, [41], that false positive rate is reduced by considering morphological features extracted from background activities. Morphological features are mostly extracted in the time domain. However, their application can be extended by introducing frequency sub-bands or other signal decomposition techniques such as half-wave decomposition, [42]. In addition, morphological features can be extracted for second-order statistics by representing EEG signals in different structures such as co-occurrence matrices. Co-occurrence matrices are mainly used in the textural analysis of monochrome images and are known as GLCM, [43]. We describe GLCM in more details further in Section 4.3.2.1.

4.1.3 Spectral analysis

Spectral analysis techniques aim at characterizing spectral contents of signals using descriptors such as power spectral density (PSD) which represents the power distribution

of a signal over frequencies, [44]. PSD can be defined in different ways. Let us denote a zero mean discrete-time signal by $x(t)$ where $t = \{ 0, \pm 1, \pm 2, \ldots, \infty \}$, the covariance function of $x(t)$ is defined as:

$$r(k) = \mathbb{E} \{ x(t) \overline{x}(t-k) \} \, , \tag{4.5}$$

where $\mathbb{E}$ is the expectation operator and $\overline{x}$ is the complex-conjugate of x. PSD of x is defined as:

$$\phi(\omega) = \sum_{k=-\infty}^{\infty} r(k) e^{-i\omega k} \, , \tag{4.6}$$

where ω is the frequency in terms of radians per sampling interval.

Assuming $r(k)$ is a rapidly declining signal in such a way that

$$\lim_{N \to \infty} \frac{1}{N} \sum_{k=-N}^{N} |k| \, |r(k)| = 0 \, , \tag{4.7}$$

The second definition of PSD is defined as:

$$\phi(\omega) = \lim_{N \to \infty} \mathbb{E} \left\{ \frac{1}{N} \left| \sum_{t=1}^{N} r(k) x(t) e^{-i\omega t} \right|^2 \right\} \, . \tag{4.8}$$

PSD estimators are mostly based on periodogram and correlogram methods derived from the definitions of PSD in (4.6) and (4.8). A periodogram spectral estimator is limited to the samples obtained for $t = \{ 1, 2, \ldots, N \}$ and relies on PSD definition in (4.8), so that

$$\hat{\phi}_p(\omega) = \frac{1}{N} \left| \sum_{t=1}^{N} r(k) x(t) e^{-i\omega t} \right|^2 \, . \tag{4.9}$$

Correlogram spectral estimators are derived according to the correlation-based definition of PSD in (4.6) so that

$$\hat{\phi}_c(\omega) = \sum_{k=-(N-1)}^{(N-1)} \hat{r}(k) e^{-i\omega k} \, . \tag{4.10}$$

where $\hat{r}(k)$ denotes the covariance lag $r(k)$ that is estimated for available samples in $t = \{ 1, 2, \ldots, N \}$.

Both of the correlogram and periodogram-based spectral estimators provide a poor estimate of PSD. This is due to large statistical variability in the estimated spectrum. Several modified variants of correlogram and periodogram-based spectral estimators were introduced in order to address the problem of high variance by employing different smoothing and averaging approaches, i.e. [45], [46] and [47] methods. Welch method calculates periodogram over overlapped segments of data. Furthermore, these segments are computed prior to periodogram calculation. Let us denote $x_j(t)$ the $j_t h$ windowed segment of $x(t)$ as follows:

$$x_j(t) = x((j-1)K + t) \, ,$$

where $t = 1, ..., M$, $j = 1, ..., S$, S is number of segments and K number of overlapping samples. The windows periodogram corresponding to $x_j(t)$ is then computed as:

$$\hat{\phi}_j(\omega) = \frac{1}{MP} \left| \sum_{t=1}^{M} v(t) y_j(t) e^{-i\omega t} \right|^2 , \tag{4.11}$$

where P denotes power of temporal window $v(t)$:

$$P = \frac{1}{P} \sum_{t=1}^{M} |v(t)|^2 . \tag{4.12}$$

Welch estimation of PSD is obtained by averaging over all windows periodogram estimated in (4.11):

$$\hat{\phi}_W(\omega) = \frac{1}{S} \sum_{t=1}^{S} \hat{\phi}_j(\omega) . \tag{4.13}$$

Welch PSD estimation can be calculated using fast Fourier transform (FFT) and is easy to implement. It has been widely used in spectral analysis of EEG signals and feature extraction for different EEG classification problems, [48, 49, 50].

4.1.4 Frequency sub-band activities

Time-frequency representation of EEG signals is mostly used for analyzing EEG signal activities in different time-frequency scales. The most common technique is to decompose EEG signals into different frequency sub-bands. Signal decomposition can be performed using WT or band-pass filtering, [51, 52]. The main advantage of this decomposition technique is that once the sub-band decomposition is performed, other common features can then be extracted from each frequency sub-band, [53].

4.2 Uni-variate feature extraction using rational short time Fourier transform

4.2.1 1D local Gabor binary patterns operator

The local binary pattern (LBP) operator was originally proposed for texture analysis, [54], which was combined with Gabor filtering in [55]. There are various methods to compute LGBP features of a signal. Generally, a set of multi-scale and multi-orientation Gabor filters is first applied to obtain a multi-resolution representation in different domains (spatial, temporal, frequency, etc.). Then, the real, imaginary, magnitude or phase part of the Gabor filtered complex valued signal can be used as the input of the LBP operator. The resulting feature is the so-called LGBP, which is widely used in face recognition and head pose estimation, [55, 56, 57, 58, 59, 60]. It is worth mentioning that LGBP features are usually adopted to 2D and 3D problems, which utilize 2D and 3D Gabor filters, [61, 62].

4.2.1.1 Gabor Filter

Gabor transform, [63], is a special type of STFT, which provides an adjustable band-pass filter that can be used to determine the response of a localized signal to certain frequencies.

It uses a Gaussian envelope multiplied by a complex sinusoidal carrier to generate a joint *t-f* representation of the signal centered at a specific frequency and orientation. The 1D Gabor filter can be defined as follows:

$$g(x, \sigma, x_0, \omega_0) = \frac{1}{2\pi\sigma} e^{\frac{-(x-x_0)^2}{2\sigma^2}} e^{2\pi i \omega_0 x}\,, \tag{4.14}$$

where σ and x_0 denote the scale and the location of the peak of the Gaussian envelope, and ω_0 is the center frequency of the sinusoidal carrier. Gabor filters can be applied on EEG frequency sub-bands. For the sake of simplicity, we denote these filters by $g_j(x) := g(x, \sigma_j, x_j, \omega_j)$, where $\sigma_j = x_j = 2^{4-j}$ and $\omega_j = j$ $(1 \leq j < 5)$. Thus, the Gabor filtered sub-bands are obtained via the discrete convolution operator:

$$G_{ij}[m] = (f_i * g_j)[m] \qquad (0 \leq i < L, 1 \leq j < 5)\,. \tag{4.15}$$

Then, we take average of the magnitude responses of the complex valued filtered signals, which serve as the input to the 1D LBP operator:

$$AVG_i[m] = \frac{1}{4} \sum_{j=1}^{4} |G_{ij}[m]| \qquad (0 \leq i < L)\,. \tag{4.16}$$

4.2.1.2 1D local binary patterns

LBP transform provides a robust feature extractor for texture analysis. The original LBP transform is a 2D operator, which thresholds the pixels using the center value of the neighborhood mask. In order to make it applicable to time-series, we define a simple 1D extension of the LBP operator. Let f stand for the discrete time series that contains the absolute values of the corresponding real signal, and μ the unit step function:

$$\mu[m] = \begin{cases} 1, & \text{if } m \geq 0 \\ 0, & \text{else} \end{cases}\,. \tag{4.17}$$

Then, we set the threshold at the mth point to:

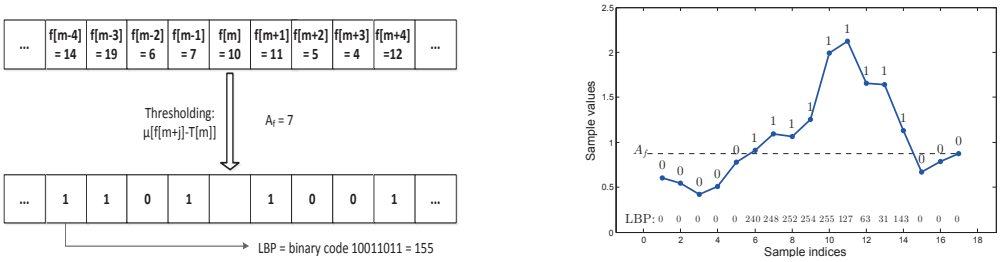
$$T[m] = \begin{cases} \infty, & \text{if } A_f > f[m] \\ A_f, & \text{else} \end{cases}\,, \tag{4.18}$$

where A_f is the average amplitude of the whole EEG signal f. The LBP pattern at the mth sample can then be obtained by assigning the factor 2^j to each neighborhood as follows:

$$\mathrm{LBP}[m] = \sum_{\substack{j=-4 \\ j \neq 0}}^{4} \mu[f[m+j] - T[m]] \cdot 2^{4+j-\mu[j]}\,. \tag{4.19}$$

Hence, similar to 2D LBP operator, 1D LBP operator also assigns a value between 0 and 255 to each signal value. Note that in our interpretation, due to the thresholding operator T, low signal samples have zero LBP values (i.e., only higher values or relevant activities are considered during the feature extraction). This is illustrated in Figure 4.1, where we demonstrate the mechanism of the 1D LBP operator.

The LBP operator is applied on each of the average Gabor filtered sub-band. In order to reduce the computational complexity, we will down-sample the LGBP vector. Namely, we take the mode value of the LGBPs in each one-second-long EEG epoch. We call the resulting features as LGBP-mode, which are the input of the LGBP-width operator.



(a) The modified 1D LBP operator. **(b)** LBP values of an EEG epoch.

Figure 4.1: Examples for the 1D LBP operator **[P5]** © 2017 Elsevier

4.2.1.3 local Gabor binary pattern-width

The LGBP is supposed to detect the lower frequency rhythmic discharges and signal oscillations in the rational components. Hence, we expect the largest LGBP value, 255, to occur during seizure events. However, the largest LGBP value may also occur during inter-ictal sections. To discriminate between these two LGBP groups, we consider a new coding called LGBP-width. More precisely, we search for all the LGBPs, which are equal to 255. Then, we count the number of instances that the same value constantly occurs in the left and the right neighborhoods. In each direction, the counting stops when it reaches a value lower than 255. Thus, the resulting numbers record the widths of 255 constant sequences. Finally, the resulting widths are assigned to the corresponding samples. In fact, the rational LGBP-width provides a sparse version of LGBP by suppressing the isolated largest LGBPs carried out by assigning zero value to them. As it can be seen in Figure 4.2, the proposed LGBP-width can efficiently reduce the similarity between ictal and inter-ictal events in the LGBP vector, and so it explicitly increases the discrimination power of the system. For instance, Figure 4.3, shows the results of the proposed feature extraction method on the 2nd rational component of a real EEG signal. The signal belongs to channel T8-P8 of record 26 of patient 1 in the CHB-MIT Scalp EEG database. Note that the largest LGBP-width values point to 90% of the seizure's length, which is observed between $1862 - 1963$ seconds (dashed lines).

4.3 Multi-variate feature extraction of 2D mapped EEG records

In this section, a 2D mapping scheme of EEG signals is described. The method is used to map 1D EEG signals into gray-level image domain enabling the use of textural feature extraction methods from image processing domain. The extracted features from the constructed image can then describe the spatio-temporal characteristics of EEG channels and their frequency sub-bands.

4.3.1 The curse of dimensionality

In machine learning, the curse of dimensionality referes to the phenomena when the training set consists of a limited number of samples with a high dimensional feature space. In fact, each data sample is represented with an optimal number of extracted features and increasing the dimensionality of the feature space reduces the generalization power of the classifier via over-fitting. In the classification of multi-channel EEG records, the curse of dimensionality is one of the main challenges since, in most of the applications, the

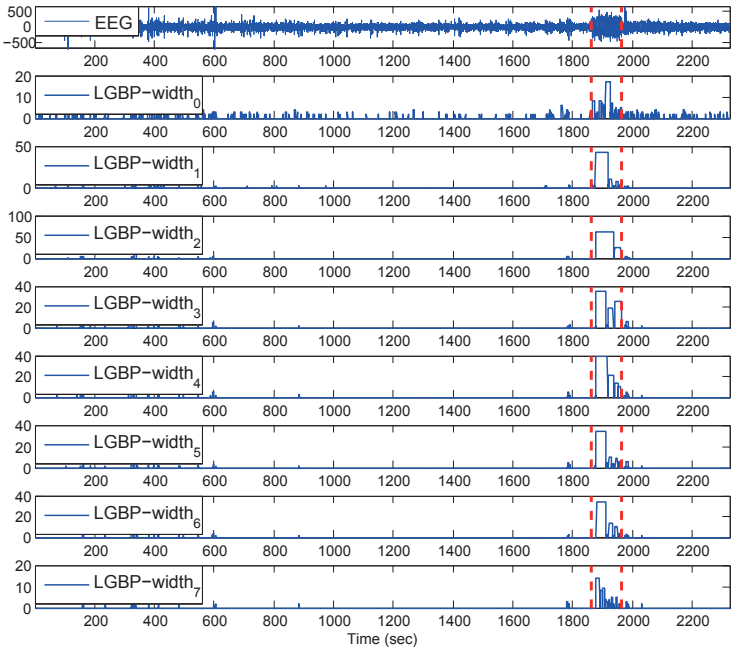


Figure 4.2: LGBP-width features for the $8th$ sparse rational components of EEG record *chb01_26*, channel *T8-P8* **[P5]** © 2017 Elsevier

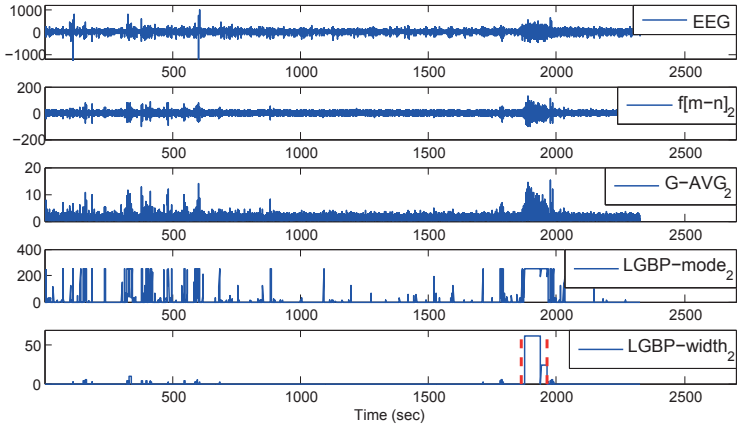


Figure 4.3: The results of the rational LGBP-width feature extraction technique for EEG record *chb01_26*, channel *T8-P8* **[P5]** © 2017 Elsevier

size of the training set is small and conventional feature extraction methods result in a large dimensional feature space depending on the number of EEG channels. Multi-variate features aid to overcome this phenomena by construction of a more compact feature space.

4.3.2 2D mapping of EEG signals

In order to map EEG signals into a grey-level image, a 2D matrix is constructed using all EEG signals and their associated frequency sub-bands. Each row of the matrix indicates one EEG signal while the columns represent the amplitudes of the EEG signals as a function of time. More explicitly, element *(i,j)* in the 2D matrix indicates the EEG value of *i_th* EEG signal at time *j*. The final gray-level image, I, is obtained by scaling the absolute values of the 2D matrix between 0 and 255. In CHB-MIT dataset, there are 23 bipolar channels and EEG records have been divided into 1 to 4 hours segments. Therefore, 5 sub-bands and the original channel will make 6 distinct EEG signals per channel, and the overall, image I is composed of $6 \times 23 = 138$ rows and more than $921K$ columns depending on the total duration of the signal. To analyze the EEG activity in time, we consider EEG epochs with length of 1.25 seconds and 25% overlap. Similarly, image I can be divided into a sequence of frames with size of 138×320 pixels, representing a texture pattern associated with an EEG activity as a function of time. During seizures, EEG activities are monitored by a set of characteristics such as variant frequency and amplitude changes, rhythmic discharges and non-stationary spiking which result in a different texture pattern in image I compared to the normal epochs.

4.3.2.1 Gray-level Co-occurrence Matrices

In order to discriminate between these texture patterns occurring during seizures from other frames in image I, we use GLCM introduced in [64] and [43]. In image I with p pixels, GLCM is constructed for each pixel and represents the variations of gray values between the pixel of interest and its neighborhoods. Neighboring pixels are selected using a rectangular window w with a width of l pixels. In a fixed sliding window w in image I, GLCMs count the number of times that different combinations of gray values occurring between the pixel of interest at the center of w and other pixels in w. Computation of GLCMs is followed by setting a number of parameters including: orientation $(0°, 45°, 90°, ...)$, which defines in which direction the combination of gray values must happen; and offset $(d = 1, 2, ..., (l-1)/2)$, which defines the pixel gap between occurrences in each direction. For an image with k gray levels, each set of parameters results in a $k \times k$ GLCM matrix. Figure 4.4 illustrates possible combinations of pixel distances and orientations in computation of GLCMs GLCMs in a 5×5 window. An example of a GLCM calculated for $d = 1$ and orientation of $-180°$ is shown in Figure 4.5.

However, for gray-levels ranging from 0 to 255, there might be a large number of possible combinations that can increase the computational cost of GLCMs. Hence, a quantization step is performed in order to reduce the number of possible combinations of pixel values.

Several GLCMs can be extracted for different sets of parameters. Statistical information from the obtained GLCMs can be used as features describing texture patterns. Haralick, introduced 14 second-order statistical features that can be extracted from each GLCM. This features set is widely used in texture analysis and image processing. More features were proposed for GLCMs by other researchers afterwards, [65]. Here, we extract 16 of these statistical features that have been empirically found to be more discriminative in our work.

4.3.2.2 Sorting EEG signals

As it was discussed in Section 4.3.2.1, texture patterns in the resulting image I appear with more intensity variations during a seizure event. This aids to obtain additional

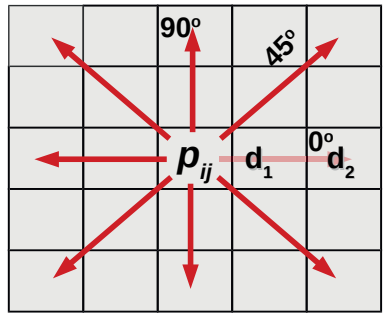


Figure 4.4: Illustration of possible orientations and pixel distances for computation of GLCMs in a 5×5 window

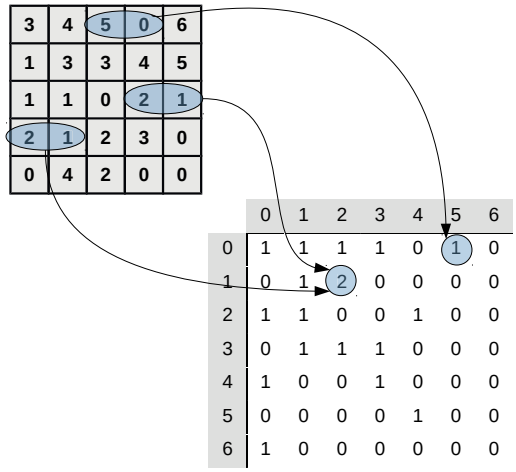


Figure 4.5: An example of a GLCM computed for $d = 1$ and orientation of $-180°$

possible combinations of gray values in the computation of GLCMs and thus achieving discriminatory Haralick's features. It is important to note that these texture patterns vary depending on how EEG signals have been sorted when forming the rows of the image I. Hence, the order of EEG signals in the texture Image I must be fixed for both training and test sets to preserve the coherency between all mapped EEG records of each patient. For each training set, the first row of image I is fixed with channel (FP1-F7) and the other EEG signals are sorted according to their similarity with respect to the first row. Consequently, for each patient, all upcoming records are sorted according to the order determined using the training set. The most common method for measuring the linear dependency and similarity between signal pairs is based on linear cross-correlation. The cross-correlation between a signal pair is computed as:

$$C_{i,j}(l) = \frac{1}{N} \sum_{n=0}^{N-1} x_i(n+l) x_j^*(n), \quad |l| \leq (N-1), \tag{4.20}$$

where N is the maximum length of signals x_i and x_j and l is the delay between two signals. Based on (4.20), for a signal pair (x_i, x_j), the maximum value of normalized

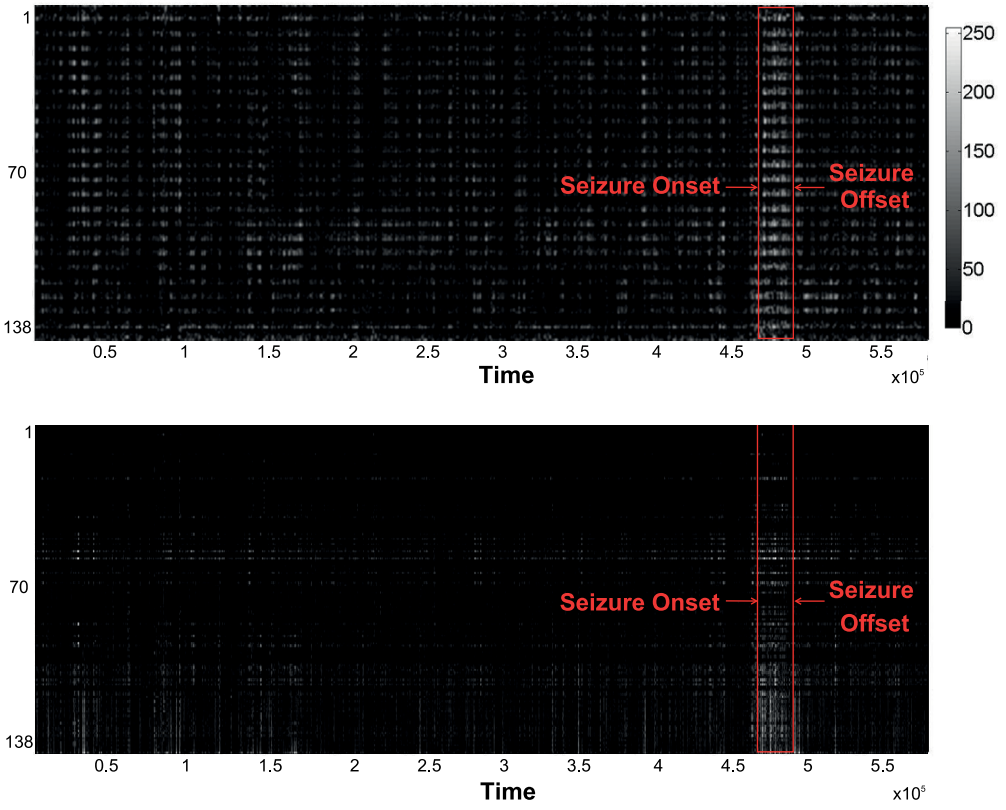


Figure 4.6: Results of mapping different EEG records into gray-level image domain using the proposed method. EEG record 26 from patient 1 in CHB-MIT dataset **[P3]** © 2017 Elsevier

cross-correlations for delays ranging between $-(N-1)$ and $(N-1)$ can be defined as, [66]:

$$C_{max} = \max_{l} \left\{ \left| \frac{C_{i,j}(l)}{\sqrt{C_{i,i}(0) \cdot C_{j,j}(0)}} \right| \right\}. \tag{4.21}$$

We use the maximum cross correlation values between the channel (FP1-F7) and all other EEG signals of the channels and their associated frequency sub-bands to sort them as the rows of image I, in a descending order from top to bottom. Since the first row is always fixed with (FP1-F7), the signals (rows) at the bottom are less similar to the channel (FP1-F7). As it can be seen in Figure 4.6, regardless whether the channel (FP1-F7) includes a seizure or not, during a seizure event texture patterns exhibits brighter pixels with significant gray-level variations ascending from top to bottom of the image I . This can be explained by the likelihood that EEG signals, which obtained a lower cross-correlation value with the channel (FP1-F7) for delays ranging between $-(N-1)$ and $(N-1)$, are less similar to each other. Hence, Hence, more chaotic texture patterns are expected at the bottom of image I.

5 Summary of the experimental results

This Chapter summarizes the results obtained by each method in the classification of EEG records and discusses the effectiveness of the rational feature extraction methods. The EEG datasets used in experiments of the thesis were briefly described in Section 2.10.

5.1 Features preprocessing

For classification tasks in all experiments, training and test sets are normalized using a transform function calculated according to sample values in the training set. In cases where features have different units, Z-score normalization is mainly used. Z-score normalization centralizes each feature in such a way that the new distribution has a mean of 0 and a standard deviation of 1 ($\mu = 0, \sigma = 1$) so that:

$$z_i = \frac{x_i - \mu_i}{\sigma_i}, \tag{5.1}$$

where x_i is the i_{th} feature element in the feature vector with mean and standard deviation of μ_i and σ_i, respectively.

Alternatively, when feature elements in the feature vector are all with the same unit, features are normalized using the minimum and maximum of features:

$$x_{norm} = \frac{x - x_{min}}{x_{max} - x_{min}}, \tag{5.2}$$

where x_{min} and x_{max} are minimum and maximum values in the training set, respectively.

5.2 Classification performance of single channel EEG time series

To assess the effectiveness of different RSTFT systems in the classification of epileptic EEG records, several binary classification problems are defined according to EEG time series sets in EEG-Bonn dataset. The dataset is widely used in the literature and can be considered as a benchmark for performance evaluation of different epilepsy detection methods. As described in Section 2.10.1, among the five different EEG sets existing in the dataset, only set (E) contains EEG signal patterns recorded during seizure. Hence, we define the following epilepsy classification tasks:

- (E - A): classification of set E in the presence of set A;
- (E - B): classification of set E in the presence of set B;
- (E - C): classification of set E in the presence of set C;
- (E - D): classification of set E in the presence of set D;
- (E - A, C): classification of set E in the presence of sets A and C;
- (E - A, B, C, D): classification of set E in the presence of sets A, B, C and D.

In all classification tasks, each EEG epoch is represented using a feature vector consisting of the absolute values of RSTFT coefficients and five statistical measures:

- absolute mean value;
- absolute median value;
- absolute standard deviation;
- absolute maximum value;
- absolute minimum value of the coefficients.

As the first experiment, the impact of the inverse pole position in the unit circle is examined for each of the rational systems in the classification task of (E - A) sets. For this purpose, the average classification accuracy of each rational system was calculated for a set of inverse poles located in a 50×50 grid of the unit circle. The results of the experiment for window size of $M = 176$ samples, $N = 16$ rational coefficients and a training/test splitting percent of $50 : 50$ are reported in Table. 5.1. MLP classifier with $N + 5$ neurons in the input layer, $(N + 6)/2$ neurons in the hidden layer, and two neurons in the output layer representing seizure and seizure-free classes was used in all classification tasks. As it can be seen, MT rational system outperforms the other variants. It can also outperform STFT when the inverse pole is located at $a_0 = -0.1 + 0.1i$. According the expansion of rational terms in 3.16, when the location of the inverse pole tends closer to the torus $\mathbb{T}$, the coefficients represent higher frequencies in RSTFT. In general, inverse poles that are closer to the center of the unit circle, result in a similar performance to STFT. The result of this experiment demonstrates the importance of choosing the optimum value for the pole in order to minimize the transform error.

Similar to STFT, there is a trade-off between time and frequency resolutions in RSTFT. Thus, it is expected to achieve a better resolution in frequency domain by increasing the length of windows resulting in a lower time resolution [67]. The impact of time-frequency resolution in the classification performance is examined in the second experiment. For this purpose, the six different classification tasks are performed for all possible combinations of time-frequency regularization parameters. Namely, four different time resolutions corresponding to epoch sizes $M = 173, 256, 512, 1024$ and three different frequency resolutions obtained for various number of MT coefficients $N = 8, 16, 32$. The classification results are summarized in Table. 5.2. The highest average Sens, Spec and accuracy are achieved for $N = 32$ MT RSTFT coefficients and an epoch length of $M = 256$ sample which is roughly equal to 1.5 s. As expected, having a better frequency resolution in a fixed epoch length results in a better classification performance for each of the tasks

Table 5.1: Comparison of average and best classification accuracies obtained by different rational coefficients in the 50×50 grid **[P2]** © 2015 IEEE

Rational System	Avg Acc (%)	Best Acc (%) (non-zero poles)
MT	**97.9**	**99.4**
RF	94.6	97.3
MRF	96.8	99.2
BRF	94.4	99.3
rational STFT	-	99.0

since, with a larger number of coefficients, RSTFT can approximate the input signal more precisely. On the other hand, an increase in the size of epoch declines the classification performance. This is due to the fact that RSTFT uses a limited number of base functions for the approximation of the input signal. Hence, RSTFT is not able to fit the input signal in the presence of high-frequency components in a long epoch.

5.2.1 Comparation with other conventional feature descriptors

The classification performance of MT RSTFT is compared with conventional uni-variate feature extraction methods using the same MLP classifier and the training/test ratio of 50%. These methods are namely:

1. Wavelet analysis, using Daubechies (db2) wavelet transform to decompose each EEG epoch into five frequency subbands. The feature space is constructed by maximum, minimum, average, and standard deviation of the wavelet coefficients of each subband, [68, 69].

2. Entropy-based feature extraction using Shannon entropy, log-energy entropy, and sample entropy, [70].

3. Spectral analysis, Welch PSD, [49].

4. Approximate entropy-based feature extraction, [71].

The results of the comparative study are reported in Table. 5.4. For each feature extraction method, classification tasks are repeated 10 times in order to address the bias-variance trade-off. As it can be seen, MT RSTFT outperforms both the entropy and the approximate entropy-based feature extraction methods in all classification tasks. Overall, MT RSTFT together with the wavelet-based method are one of the best performing feature extraction methods in terms of classification accuracy. For each EEG epoch, MT RSTFT extracts 37 features resulting in a more compact features space in contrast to the high dimensions of wavelet-based and Welch PSD feature spaces. This helps to avoid the curse of dimensionality when the size of the training set is limited. In the classification task 6, the ratio of the samples between minority and majority classes is 25%. High performance obtained by MT RSTFT in this task demonstrates the discrimination power of the features extracted with RSTFT.

Table 5.2: Classification results of MT RSTFT for different epoch sizes and number of coefficients [**P2**] © 2015 IEEE

Classification task		Window size + Number of coefficients											
		173			**256**			**512**			**1024**		
		8	**16**	**32**	**8**	**16**	**32**	**8**	**16**	**32**	**8**	**16**	**32**
E-A	**Sens%**	99.5	99.0	99.0	99.9	100.0	99.9	99.5	100.0	100.0	97.9	96.9	100.0
	Spec%	99.1	99.9	99.9	98.1	99.4	99.6	98.4	98.4	99.5	96.6	99.0	97.6
	Acc%	99.3	99.5	99.5	99.0	99.7	99.8	99.0	99.3	99.8	97.3	98.0	98.8
E-B	**Sens%**	92.5	98.9	98.9	99.1	94.9	99.6	98.5	99.8	99.0	99.5	88.1	99.0
	Spec%	98.3	98.4	98.4	94.3	98.9	99.0	95.6	92.8	98.2	91.3	98.6	95.7
	Acc%	95.4	98.7	98.7	96.8	96.9	99.3	97.1	96.4	98.6	95.3	93.5	97.3
E-C	**Sens%**	96.9	96.8	96.8	95.5	97.3	99.3	87.9	99.5	98.3	96.4	94.8	99.5
	Spec%	96.7	97.3	97.3	97.1	99.0	97.7	95.3	93.5	98.4	78.7	91.3	90.8
	Acc%	96.8	97.0	97.0	96.3	98.1	98.5	91.5	96.6	98.4	87.3	93.0	95.0
E-D	**Sens%**	94.5	94.1	94.1	91.3	93.7	95.6	86.0	95.9	96.9	83.4	82.9	90.7
	Spec%	94.3	95.9	95.9	94.5	94.7	94.1	88.6	89.4	92.5	86.0	93.7	92.8
	Acc%	94.4	95.0	95.0	92.9	94.2	94.9	87.3	92.8	94.8	84.8	88.5	91.8
E-A, C	**Sens%**	98.8	98.7	98.7	98.2	99.1	99.2	94.6	99.3	97.8	94.0	97.7	100.0
	Spec%	95.0	96.6	96.6	96.7	97.6	98.9	94.7	97.7	97.5	91.6	92.6	94.1
	Acc%	97.5	98.0	98.0	97.7	98.6	99.1	94.7	98.8	97.7	93.2	96.0	98.0
E-A, B, C, D	**Sens%**	98.3	99.6	99.6	97.4	97.1	99.2	98.3	97.8	98.7	97.3	98.5	98.2
	Spec%	91.6	87.6	87.6	93.7	94.3	93.8	78.7	91.1	91.6	66.8	81.3	88.3
	Acc%	96.9	97.2	97.2	97.4	96.5	98.1	94.4	96.5	97.3	90.8	94.8	96.1
Average	**Sens%**	96.8	97.9	97.9	96.9	97.0	**98.8**	94.1	98.7	98.5	94.8	93.2	97.9
	Spec%	95.8	96.0	96.0	95.7	**97.3**	**97.2**	91.9	93.8	96.3	85.2	92.8	93.2
	Acc%	96.7	97.6	97.6	96.7	97.3	**98.3**	94.0	96.7	97.8	91.5	94.0	96.2

Table 5.3: Top five most significant feature elements for each classification task **[P2]** © 2015 IEEE

	Classification task					
	E-A	E-B	E-C	E-D	E-A,C	E-A,B,C,D
# top 5	36	36	36	20	36	36
	12	5	18	19	10	10
	10	8	11	21	37	11
	11	7	16	18	12	32
	5	11	19	31	16	9

5.3 Performance on sleep stage classification

As discussed in Section 3.4, the sparse rational decomposition technique provides a baseline enhancing the discriminatory power of common feature extraction methods. To assess the effectiveness of the decomposition technique, it is applied on single-channel EEG records of the sleep-EDF dataset in order to perform classification of sleep stages. The total energy of each sparse rational component is extracted as single features to distinguish between different sleep stages. Three classifiers are used in order to conduct the experiments, namely, 1) a single hidden layer MLP classifier with 8 input neurons and 25 neurons in the hidden layer, 2) a RF classifier with a maximum number of 8 trees, 3) a linear– (Lin-SVM) classifier with L1 penalty, a regularization parameter $C = 1.0$ and a tolerance of 10^{-3}. The 5-fold cross validation technique over each training set is used to tune hyper parameters of the classifiers. Table. 5.5, shows the overall classification results obtained using different classifiers. As it can be seen, RF classifier outperforms Lin-SVM and MLP classifiers. However, the performance gap among the classifiers is not significant. The detailed classification results for each subject in the dataset are listed in Table. 5.6.

5.3.1 Comparation with state-of-art results

The comparison of the proposed method with the results of other state-of-art methods reported in the literature for the same dataset and the same classification problem can be seen in Table.5.7. The competing methods were evaluated using smaller subsets of the sleep-EDF dataset containing only a limited number of subjects and no information regarding the subjects IDs were reported by authors. As subjects IDs have been reported in [72], the performance of the sparse rational components using the same subset is also reported in Table.5.7. The superior performance of sparse rational components over the whole dataset using only one feature descriptor and a limited training/test rate demonstrates its ability in improving the discriminatory power of conventional feature extraction methods.

5.4 Performance in classification of multi-channel epileptic EEG records

Compared to single-channel EEG time series of Bonn dataset, long-term multi-channel EEG records of CHB-MIT dataset contain more intra/inter class variations. For each patient, the seizure type and its location vary over time. The variability of dataset increases

Table 5.4: Classification results obtained by different feature extraction methods using MLP classifier average over 10 repetitions **[P2]** © 2015 IEEE

Classification task		Feature extraction				
		Wavelet Analysis	Entropy Based	Welch PSD	Approximate Entropy	MT rational DSTFT
E-A	Sens%	100.0	92.1	98.7	87.3	99.9
	Spec%	100.0	99.9	100.0	89.6	99.6
	Acc%	**100.0**	96.0	99.4	88.4	99.8
E-B	Sens%	99.2	90.5	98.2	91.1	99.6
	Spec%	99.1	99.4	100.0	94.0	99.0
	Acc%	99.2	94.9	99.1	92.6	**99.3**
E-C	Sens%	98.7	89.6	92.3	0.0	99.3
	Spec%	97.9	97.6	99.9	100.0	97.7
	Acc%	98.3	93.6	96.1	50.3	**98.5**
E-D	Sens%	95.6	85.2	86.7	2.8	95.6
	Spec%	97.4	94.4	99.6	99.4	94.1
	Acc%	**96.5**	89.8	93.2	51.3	94.9
E-A, C	Sens%	97.5	93.6	98.3	0.0	99.2
	Spec%	99.7	98.0	99.7	100.0	98.9
	Acc%	99.0	96.5	**99.3**	66.3	99.1
E-A, B, C, D	Sens%	96.0	90.3	95.0	0.0	99.2
	Spec%	98.7	96.7	98.7	100.0	93.8
	Acc%	**98.1**	95.5	97.9	79.8	**98.1**

Classifier	Sleep Stage				Overall	
	Wake	Stage 1 + REM	Stage 2	Stage 3&4		
	recall %	recall %	recall %	recall %	Acc%	F1%
Lin-SVM	97.71	50.33	79.73	39.30	88.26	86.43
MLP	97.53	69.75	82.55	75.89	91.63	91.46
RF	**98.26**	**73.51**	**82.69**	**77.17**	**92.50**	**92.39**

Table 5.5: Overall classification results obtained for all 39 patients using different classifiers [**P4**] © 2015 IEEE

Table 5.6: Sleep stage classification results of subject in Sleep-EDF dataset

record ID	Wake		Stage I + REM		Stage II		Stage III & IV		Overall	
	recall%	F1%	recall%	F1%	recall%	F1%	recall%	F1%	recall%	F1%
SC4001E0	99.0	98.2	71.1	75.4	77.5	80.6	91.4	9.0	94.4	94.2
SC4002E0	98.9	98.5	76.4	77.7	81.0	79.9	86.5	8.9	93.1	93.1
SC4011E0	98.5	97.3	82.7	78.5	85.1	89.6	72.8	7.9	93.4	93.3
SC4012E0	97.6	97.1	79.2	74.3	86.7	89.9	72.2	7.4	92.4	92.5
SC4021E0	98.0	97.2	61.1	67.2	88.1	85.7	70.1	7.6	91.9	91.6
SC4022E0	98.2	97.2	75.2	74.5	76.7	79.9	77.0	8.0	91.0	90.9
SC4031E0	98.3	98.5	88.1	87.7	93.4	93.2	83.8	8.2	96.3	96.3
SC4032E0	99.4	98.9	81.3	83.4	85.5	85.9	78.4	7.9	94.6	94.5
SC4041E0	96.6	96.9	83.5	77.1	84.0	87.5	69.2	6.8	91.2	91.4
SC4042E0	97.3	97.5	85.7	80.1	82.9	86.0	76.1	7.9	92.2	92.3
SC4051E0	98.5	97.4	68.4	72.7	75.0	78.9	69.0	7.4	94.1	93.9
SC4052E0	97.0	96.1	72.3	75.4	86.5	87.7	88.1	8.5	91.6	91.5
SC4061E0	99.2	98.9	68.9	72.2	85.7	85.9	89.4	8.8	95.0	94.9
SC4062E0	97.5	97.8	75.0	71.8	85.5	86.2	88.8	9.2	93.2	93.3
SC4071E0	99.4	98.2	77.1	82.0	87.9	86.7	78.8	8.5	94.1	94.0
SC4072E0	99.6	98.6	85.0	85.0	84.9	88.1	89.8	9.2	95.4	95.3
SC4081E0	98.4	97.5	65.5	72.1	75.0	75.4	88.6	8.8	92.7	92.5
SC4082E0	97.7	96.5	70.9	72.5	70.0	69.7	80.5	8.5	89.4	89.3
SC4091E0	99.6	98.4	66.7	68.8	81.4	81.5	76.8	8.3	91.3	91.1
SC4092E0	96.5	95.0	74.7	75.7	88.2	88.2	70.0	7.9	89.4	89.2
SC4101E0	97.5	95.9	71.8	72.4	83.0	86.2	0.0	0.0	91.2	91.0
SC4102E0	96.9	96.3	79.3	77.8	87.5	89.8	55.6	6.3	92.6	92.6
SC4111E0	96.9	97.4	40.6	48.9	83.8	76.7	71.0	7.5	89.2	89.0
SC4112E0	99.9	99.1	77.1	77.1	84.6	87.0	73.3	7.8	95.3	95.2
SC4121E0	98.3	98.0	70.4	71.3	83.5	83.0	84.0	8.8	92.1	92.0
SC4122E0	98.3	97.8	75.0	71.9	64.7	70.6	89.7	9.0	91.6	91.5
SC4131E0	99.7	99.1	86.4	81.5	82.4	87.3	92.0	9.0	95.0	94.9
SC4141E0	98.2	98.1	89.4	85.5	88.1	89.6	84.9	8.9	95.2	95.2
SC4142E0	99.5	98.2	81.8	84.3	88.6	90.2	83.2	9.0	95.5	95.4
SC4151E0	99.3	98.7	77.2	77.7	79.1	83.1	94.4	9.1	94.1	94.1
SC4152E0	99.5	99.1	76.2	80.1	82.0	83.2	92.0	8.6	93.6	93.5
SC4161E0	96.7	96.4	80.4	77.9	84.3	84.8	73.4	8.0	91.1	91.1
SC4162E0	97.7	96.0	66.9	71.8	78.9	81.0	72.2	7.6	90.6	90.4
SC4171E0	96.4	96.5	75.0	74.5	81.2	75.4	65.2	7.3	90.0	90.0
SC4172E0	98.8	96.7	47.6	54.9	81.6	79.4	58.0	6.7	87.5	86.9
SC4181E0	99.3	98.2	40.7	51.1	82.8	77.0	79.9	8.6	91.8	91.4
SC4182E0	98.6	98.0	76.0	79.7	80.8	78.5	87.1	9.1	93.9	93.8
SC4191E0	98.0	95.2	68.5	69.5	82.9	86.2	83.9	8.7	88.7	88.5
SC4192E0	97.6	95.4	77.7	81.4	84.2	88.8	72.5	6.9	91.7	91.5
Average	98.3	97.5	73.5	74.7	8.3	83.7	77.2	8.0	92.5	92.4

Author(s)	# of subjects	Acc (%)
Zhu et al. (2012) [72]*	4	83.19
Sparse rational energy*	4	**93.35**
Phan et al. (2013) [73]	4	86.30
Liu et al. (2010) [29]	7	89.3
Ebrahimi et al. (2008) [74]	7	**93.00**
Li et al. (2009) [75]	8	81.73
Sparse rational energy	39	**92.50**

Table 5.7: The comparison of the sparse rational decomposition technique with other state-of-the-art methods for the 4-class sleep scoring performed on the Sleep-EDF dataset **[P4]**

* results reported for the subjects: SC4002E0, SC4012E0, SC4102E0 and SC4112E0.

considering the diversity in patients and existence of multiple channels. Hence, CHB-MIT dataset provides a benchmark for a more realistic epilepsy detection/classification problem. As it was discussed in Section 2.10.2, for each patient, the training set is picked from early records with a training/test splitting percent of 25 : 75. The constraint in the selection of the training set of each patient aids to mimic two main challenging phenomena which appear in epilepsy seizure monitoring in the real world, namely, the curse of dimensionality which is more probable to happen with a limited number of samples in the training set, and the curse of variability that occurs in EEG records of patients suffering from several types of seizure.

To assess the robustness of the rational LGBP-width and Haralick-textural feature extraction methods, several different classifiers are employed in the experiments, and their classification performances are compared against other dedicated feature extraction methods proposed in the literature. The feature extraction methods in the comparative study include: (1) feature extraction based on RSTFT coefficients, **[P2]**. For each EEG epoch, 64 RSTFT coefficients are concatenated with five statistical measures containing the maximum, minimum, mean, median and standard deviation of the absolute values of the coefficients; (2) spectral analysis of each EEG epoch using Welch FFT power spectral density (PSD) [49]; (3) Wavelet based feature extraction method; (4) Entropy-based feature extraction; (5) Approximate entropy-based feature extraction; (6) a method based on several time-domain frequency-domain features (TDFDF), [76], containing three frequency domain features formerly proposed by [77] and two time domain features, namely spike rhythmicity and relative spike amplitude, [78]. In addition, similar to the feature extraction method in Section 5.3 and **[P4]**, the total energy of each sparse rational component of EEG channels is extracted in order to examine the effectiveness of LGBP features.

All algorithms in this comparison are applied on all bipolar channels in CHB-MIT dataset. Finally, the features, computed individually for each channel, are concatenated into one vector to construct the final feature vector. Eventually, each classifier is fed by a $P \times Q$ data matrix where P is the size of the feature vector and Q is the total number of EEG segments in the training set. The dimension of the feature vector of each method obtained for each EEG segment is listed in Table.5.8. According to Tables.2.1 and 5.8, for patients 1 with 23 bipolar channels, rational LGBP and Haralick-textural feature extraction methods yield training sets with the size of 184×5980 and 102×5980, respectively.

Tables. 5.9 and 5.10, summarize classification results obtained using different classifiers

Table 5.8: Feature vector dimension of each feature extraction method [**P3**] © 2015 Elsevier

FEX method	FV dimension
Rational LGBP-width	No. of EEG channels×8
Sparse rational energy	No. of EEG channels×8
Rational coefficients	No. of EEG channels×69
Welch PSD	No. of EEG channels×50
Wavelet coefficients	No. of EEG channels×20
App Ent	No. of EEG channels×1
Entropy based	No. of EEG channels×3
TDFDT	No. of EEG channels×5
Haralick-textural	7×16

Table 5.9: The average Sens and Spec rates obtained for Haralick-textural features and the other features extraction methods using different classifiers [**P5**] © 2015 Elsevier

FEX method	Classifier									
	SGD-SVM		Log-Reg		KNN		Naive Bayes		RF	
	Sens%	Spec%	Sens%	Spec%	Sens%	Spec%	Sens%	Spec%	Sens%	Spec%
Haralick-textural	**70.19**	97.74	**58.74**	97.94	44.36	98.96	**78.50**	72.27	53.55	96.38
Wavelet coefficients	62.48	98.22	40.65	97.66	**52.46**	98.59	72.70	71.22	**57.81**	**98.65**
Entropy based	54.17	96.87	26.57	96.35	37.73	98.93	76.10	**74.83**	46.18	98.10
TDFDF	36.36	**98.91**	34.68	**98.62**	35.32	**99.26**	74.08	70.82	50.62	96.95
AppEnt	32.31	95.14	25.93	97.67	36.14	94.34	61.61	73.21	26.70	94.85

for Halarick and rational LGBP features, respectively. We note that the same training and test sets, described in Table. 2.1, are used in all the experiments corresponding to CHB-MIT dataset. Furthermore, the k-fold cross validation technique over each training set is utilized to estimate optimal hyper parameters of each classifier. As it can be seen, Haralick's features outperform other competing feature extraction methods in terms of Sens rate for 3 out of 5 classifiers, while it is ranked as second or third best performing method in terms of Spec rate. Rational LGBP feature extraction method shows even more robustness to the classifier type as it obtains the highest AUC and F1 rates using all classifiers, while it ranked as the best performing method in terms of Sens and Spec rates for 2 out of 3 classifiers. The consistency of classification performances achieved by the proposed methods demonstrates their discriminatory power in differentiating between ictal and non-ictal EEG signal patterns. Haralick's features set and rational LGBP obtain their best performance using stochastic gradient descent (SGD)-support vector machine (SVM) and logistic−regression (Log-reg) classifiers, respectively. We note that the slight differences in performance of feature extraction methods using RF and Log-reg classifiers in Tables. 5.9 and 5.10 are due to different implementations and parameter settings in the classifiers. In all the classification experiments on CHB-MIT dataset, L1 and L2 regularization penalties were used in Log-reg and SVMs classifiers, respectively. Furthermore, the maximum number of trees in RF classifier was set to the number of feature elements. Besides, the k−nearest neighbors (KNN) classifier was tuned using the Euclidean distance and the number of neighbors of $k = 3$.

As it can be seen in Table. 5.10, the rational LGBP-width and the sparse rational energy feature descriptors outperform rational coefficients. This shows that EEG signals decomposed into sparse rational components are more amenable for extraction of useful information using common feature descriptors. Moreover, the performance improvement obtained by the deployment of the LGBP-width descriptor, compared to the sparse rational energy, demonstrates the effectiveness of the feature descriptor.

Table 5.10: overall classification results of the rational LGBP-width and the other different feature extraction methods using three classifiers, top two best performance of each classifier are highlighted in bold **[P5]** © 2015 Elsevier

FEX method	Classifier											
	Log-reg				RF				Lin-SVM			
	Sens%	Spec%	AUC%	F1%	Sens%	Spec%	AUC%	F1%	Sens%	Spec%	AUC%	F1%
Rational LGBP-width	**70.39**	**99.09**	**85.41**	**98.85**	**66.35**	99.29	**82.79**	**98.62**	60.42	**99.49**	**83.00**	**98.89**
Sparse rational energy	42.83	84.69	64.02	86.23	10.07	**99.65**	54.47	97.08	37.26	79.88	58.28	79.68
Rational coefficients	56.12	91.81	73.68	93.06	48.69	96.47	70.31	96.14	47.94	84.88	69.44	**88.08**
Welch PSD	**71.74**	89.95	**81.23**	91.98	**60.17**	96.70	**75.55**	96.40	**61.69**	81.56	73.96	83.97
Wavelet coefficients	69.32	90.28	80.89	92.14	56.30	**98.69**	74.98	**98.10**	**60.75**	83.29	**75.08**	85.51
App Ent	47.40	90.51	67.38	88.75	26.54	95.33	59.62	95.13	40.00	80.69	61.19	80.80
Entropy based	59.20	**92.08**	75.57	**93.93**	44.74	97.92	70.20	97.45	47.60	**85.36**	67.58	85.45
TDFDT	45.85	89.26	69.43	91.04	49.40	97.02	70.64	96.62	40.69	83.99	63.49	84.36

Table 5.11: Top five feature elements in the extracted Haralick's feature set **[P3]** © 2015 Elsevier

# rank	GLCM	feature type
1	offset: 5, horizontal	HG
2	offset: 5, horizontal	DEN
3	offset: 5, horizontal	CP
4	offset: 1, horizontal	DVA
5	offset: 5, horizontal	DS

5.4.1 Feature analysis

In order to find the most discriminative feature element in each of the rational LGBP-width and the Haralick's feature sets, a supervised feature ranking algorithm, [79], is employed on the training sets used previously for patient-specific seizure classification. The recursive feature elimination (RFE) technique uses SVM weights assigned to each feature element in order to rank the most informative ones. The results of feature ranking on Haralick's feature set is summarized in Table. 5.11. As it can be seen, the most contributing feature in all classification tasks is the homogeneity (HG) feature calculated for a GLCM with a distance offset of 5 pixels and an orientation of 0°. Moreover, the other high ranked feature elements also belong to the horizontal orientation of 0° with various distance offsets. This implies the importance of spatial exploration of EEG signal in different frequency sub-bands in order to differentiate between ictal and non-ictal patterns. On the other hand, due to the sorting of EEG sub-bands in the construction of the mapped image, the obtained textural pattern possesses smoother changes in the vertical direction. Hence, with a small distance offset, it is more likely to achieve less variability in GLCMs calculated for vertical orientations.

In order to evaluate the average contribution of each Haralick's feature set using all GLCMs obtained in different orientations and pixel distances, SGD-SVM is trained using the training set with a feature vector limited to elements of only one Halarlick's feature set in a patient-specific manner. The weighted average classification results for each Halarlick's feature set is listed in Table. 5.12. HG feature set obtained the best classification performance together with the cluster prominence (CP). It can be seen in comparison to the overall classification performance using all Haralick's features in Table. 5.9 that the overall Sens is significantly improved while the overall Spec dropped by roughly 2% compared to Spec rates obtained for each individual feature set.

RFE feature ranking method is also applied to rational LGBP-width features in order to

Table 5.12: Average performance of each Haralick's feature set on the test set **[P3]** © 2017 Elsevier

Haralick's feautre set	Sens %	Spec %	AUC %
AC	30.17	99.75	62.14
CO	42.37	99.39	71.72
CS	26.94	99.83	60.38
DS	46.14	99.39	73.71
UF	42.38	99.50	69.40
EN	41.10	99.47	68.64
HG	47.56	99.52	**74.20**
MP	43.18	99.51	70.15
VA	32.39	99.39	63.32
SAV	33.81	99.39	64.03
SVA	33.19	99.70	64.72
SEN	40.48	99.48	68.17
DVA	42.38	99.39	71.73
DEN	46.32	99.50	73.49
CP	47.05	99.40	**74.03**
CR	43.87	99.38	72.50

rank the importance of each feature. In addition, by selecting the top five feature elements of each patient and finding the corresponding EEG channel and frequency sub-band, it is possible to estimate the average contribution of each EEG channel in seizure classification of all patients.

Figure 5.1 shows the probability of each EEG channel being one of the most informative channels over the entire test set. The numbers above each bar show the three most significant frequency sub-bands of the corresponding channel. As it can be seen, channels $FP1 - F7$, $T7 - P7$, $F7 - T7$ and $P4 - O2$ obtain the highest contribution in the classification of seizure segments. These channels cover frontal, parietal, temporal and occipital lobes of the brain, respectively. We note that seizure activities and their location in the brain vary from patient to patient with a non-uniform distribution in the dataset. As a result, it is not easy to correlate the relevance of each frequency sub-band to a seizure or seizure-free event.

5.4.2 Comparison with state-of-art results

To assess the robustness of rational LGBP-width and Haralick's textural feature extraction methods, their performance in patient-specific seizure detection/classification problem is compared with the results of other competing methods reported in the literature for CHB-MIT dataset. Table. 5.13 summarizes results of the comparison. Among the methods listed in Table. 5.13, results of the rational LGBP-width and the Haralick's textural methods are achieved for identical training and test sets (reported in Table.2.1) using Log-reg and SGD-SVM classifiers, respectively. The method in [80] also uses the same training/test ratio while the other methods benefit from larger training ratios. The lower training/test ratio causes the sub-optimality of the rational LGBP-width compared to the other competitors. Besides, the constraint of picking the training samples only from early records of each patient makes it more difficult for the classifier to generalize the model in the test set. In fact, with such a constraint, there is more intra variability between seizure events in training and test sets of each patient. This phenomenon can be seen in patients 13–16 and 24 for whom the seizures type and locality on the brain change over time. Nevertheless, due to the high specificity rate, the proposed method can

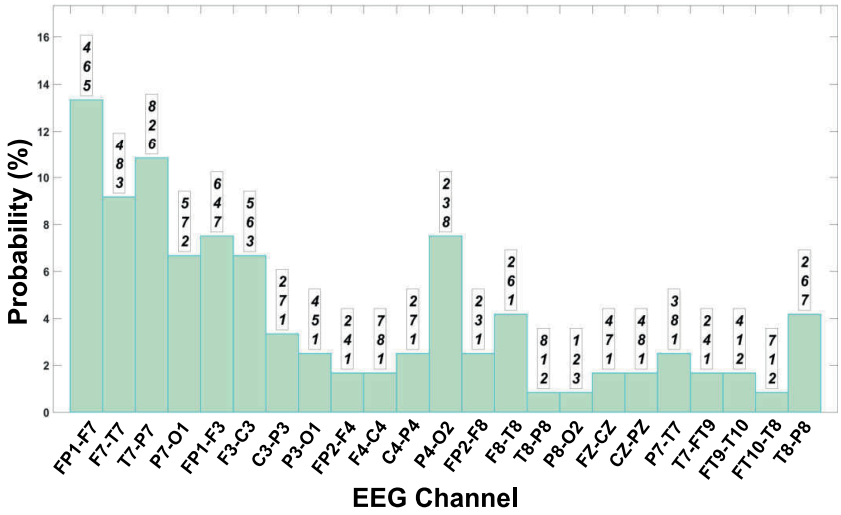


Figure 5.1: Probability distribution of each channel being one of the top 5 relevant channels over the entire test set. Numbers in boxes indicate top 3 most significant frequency sub-bands of each channel **[P5]** © 2017 Elsevier

still avoid producing high false alarm rates. In the seizure segment classification task, in contrast to methods used the same training rate, the rational LGBP-width achieves a better performance in terms of specificity rate. In addition, it outperforms the Haralick's textural features.

Table 5.13: Performance comparison of different seizure detection methods reported for CHB-MIT dataset **[P5]** © 2015 Elsevier

Seizure segment detection

Author(s)	No. Subjects	training/test	Sens%	Spec%	Acc%	AUC%
[81]	24	80-20%	-	-	80.16	-
[80]	21	25-75%	89.00	93.00	-	-
[40]	18	10-fold CV	98.27	98.36	98.31	-
[82]	24	80%-Holdout	93.00	94.00	-	98.00
[83]	24	70-20%	-	-	90.00	-
Haralick's textural, **[P3]**	23	25-75%	70.19	97.74	-	-
Rational LGBP-width, **[P5]**	23	25-75%	70.40	99.10	-	85.40

Seizure event detection

Author(s)	No. Subjects	training/test	Sens%	AUC%	Lat (sec)	Fph
[84]	24	variable	96.00	-	4.60	0.13
[85]	10	80-20%	100.00	-	0.7	0.55
[86]	24	60-40%	98.50		1.76	-
[87]	6	Leave-one-out CV	100.00	-	3.36	21.29-183.40
[88]	22	50-50%	97.10		2.80	-
[89]	23	Leave-one-out CV	-	97.70	9.33	1.52
Rational LGBP-width, **[P5]**	23	25-75%	91.10	99.10	5.56	0.35

6 Conclusion

Analysis and classification of long-term epileptic EEG records is a challenging problem as it must fulfill clinical requirements and standards. First, each EEG detection system must yield a high true positive rate and a low false alarm rate. Additionally, there is usually a limited amount of annotated data available for each patient which is insufficient for deployment of advanced supervised machine learning models such as recurrent neural networks. Moreover, due to the non-stationary behavior of EEG signals and the high variability of EEG signal patterns, generic and non-patient specific models may result in a deficient performance. In machine learning, the curse of dimensionality phenomenon can occur when the dimension of extracted features is significantly high in contrast to the number of samples available in the training set. In the binary classification of multi-channel EEG records with a limited number of annotated data, this phenomenon is mostly the main reason for over-fitting of the model. These problems are investigated and have been addressed by proposing a novel sparse decomposition method of EEG signals which aids to enhance the discriminatory power of conventional feature extraction techniques.

The suggested method in [**P3**] addressed the curse of dimensionality problem in classification of multi-channel EEG records by reducing the dimension of the feature space extracted from a mapped 2D input EEG space. The input space is constructed by mapping EEG channels and their frequency sub-bands into a 2D "image" gray-scale domain in such a way that EEG signals originating from different channels were sorted based on their correlation with a fixed frontal channel. Mapping each EEG epoch into a gray-scale image makes it possible to adapt and extract features commonly used in image analysis, resulting in a more compact feature space. Besides, by employing textural feature extractors such as GLCMs, it is possible to obtain features capable of describing the semantics of signals of all channels of each epoch simultaneously. The effectiveness of the proposed feature extraction method was examined in a patient-specific classification task of epileptic long-term EEG records of CHB-MIT data-set by considering a small training rate of 25% for each patient in the data-set. High specificity and sensitivity rates obtained by the proposed method demonstrate the possibility of synthesis imaging of multi-channel epileptic EEG records in order to semantically analyze the chaosity and irregularity of ictal and non-icatl EEG signal patterns. The proposed 2D representation provides a spatio-temporal representation which can be directly fed into a CNN and other deep learning techniques for further analysis. However, the proposed feature extraction method was not successful to outperform state-of-art results in non-epileptic EEG classification problems where similar texture appears in the 2D input space due to quantization and scaling error induced during the mapping of the signal amplitudes to gray-level values.

In [**P2**] a novel time-frequency transform was introduced using different rational systems. Similar to STFT, it was shown that each EEG signal epoch in this transform can be

encoded into a compact combination of poles and zeros of the rational system. However, unlike discrete STFT, the optimum place of poles and zeros in the unit circle can be searched for each epoch in order to achieve a lower reconstruction error. The application of such a system was investigated in the classification of single-channel epileptic EEG time-series. Besides, the relation between the EEG epoch size and the number of coefficients of RSTFT with the performance of seizure detection system were examined in terms of classification accuracy. However, despite the lower reconstruction MSE and superior classification performance of the proposed method compared to standard STFT, further research must be conducted to evaluate the effectiveness of RSTFT in a real-world application with long-term multi-channel EEG recordings and a lower training rate.

This technique enables to characterize each EEG epoch by signal activities in each sub-band. Inspired by this, A signal decomposition method was introduced in [**P4**] using RSTFT and BP algorithm. The method was employed to decompose each EEG signal into several sparse signal components by imposing L1 minimization constraint on coefficients of RSTFT. This helps to characterize each channel of a EEG epoch. However, similar to independent component analysis (ICA) and EMD methods, the frequency bands of signal components achieved using this approach may not be limited to standard frequency sub-bands, namely delta, theta, alpha, beta, and gamma. The robustness of the proposed method was evaluated on the single-channel classification of over 670 hours EEG sleep records of 39 subjects of Sleep-EDF data-set. By extracting the total energy of each sub-band as the input to a RF classifier trained on 25% of sleep record of each subject, the overall accuracy of 92.50% on test sets was obtained outperforming the state-of-art results.

The robust results of sleep stage classification of EEG records in [**P4**] propelled to evaluate the sparse decomposition method in a more complex EEG classification problem. Thus, a new feature extraction method was devised in [**P5**] which adopts the idea of combining Gabor filters bank and LBP operator that are widely used in image processing face recognition research domains. The proposed feature extraction method, LGBP, yields a single feature for each signal sub-band of each EEG channel with a value ranging between 0 and 255, similar to LBP operator. Moreover, since the EEG sub-band obtained from the sparse decomposition are sparse, the feature vector for each EEG epoch has a sparse distribution which also results in a faster training phase of a classifier when a gradient descent method is used for optimization. The proposed method outperformed the state-of-art results in a patient-specific classification manner. Furthermore, The robustness of the proposed feature extraction method was also demonstrated in a non-patient-specific classification manner where training sets of 11 randomly selected patients were used for training the model. The proposed method in [**P5**] suppressed the results of the other proposed methods in [**P2**], [**P3**] and [**P4**].

The main drawbacks of such a system are due to its extensive computational cost and the delay in detection of the seizures' onset. Hence, parallel or cloud implementation of the methods can be taken into account. Additionally, deep learning techniques can be deployed to overcome the limitations of GLCM and other classical textural analysis methods in order to achieve a better discrimination from the 2D mapped EEG signals. Notwithstanding the high computational costs, the proposed methods are a suitable candidate for retrospective analysis of epileptic EEG records. Contrary to the 2D representation in [**P3**] where the information regarding the channels placement on the scalp is lost, it is possible to correlate seizure signal activities, detected by the rational LGBP-width method, to the corresponding locations on the skull.

In conclusion, the feature extraction methods proposed in this thesis provide a compact and compressed representation of multi-channel EEG records suitable for discrimination and classification techniques. Moreover, the limited number of extracted features helps to overcome the curse of dimensionality phenomena and makes it easier for a classifier to find each class's distribution. Even though the performance of the proposed methods were demonstrated in epilepsy detection and sleep stage scoring problems, the application scope is not limited to those in this thesis and can be extended to the other brain activity analysis such as rehabilitation and restoration, prevention and prediction, on-board diagnostics and wearable devices, event – related potentials (EPR), BCI, social interaction and media recommendation systems. In the future, it is aimed to explore the application of RSTFT in other clinical problems. More specifically, the sparse decomposition method using RSTFT offer a generic and multi-purpose time-frequency scheme applicable to wide range of biomedical and physiological signals like ECG and impedance respiration.

Bibliography

[1] D. Wu, V. J. Lawhern, S. Gordon, B. J. Lance, and C.-T. Lin, "Driver drowsiness estimation from eeg signals using online weighted adaptation regularization for regression (owarr)," *IEEE Transactions on Fuzzy Systems*, vol. 25, no. 6, pp. 1522–1535, 2017.

[2] Ü. Aydin, J. Vorwerk, P. Küpper, M. Heers, H. Kugel, A. Galka, L. Hamid, J. Wellmer, C. Kellinghaus, S. Rampp *et al.*, "Combining eeg and meg for the reconstruction of epileptic activity using a calibrated realistic volume conductor model," *PLoS One*, vol. 9, no. 3, p. e93154, 2014.

[3] G. L. Barkley and C. Baumgartner, "Meg and eeg in epilepsy," *Journal of clinical neurophysiology*, vol. 20, no. 3, pp. 163–178, 2003.

[4] S. Knake, E. Halgren, H. Shiraishi, K. Hara, H. Hamer, P. Grant, V. Carr, D. Foxe, S. Camposano, E. Busa *et al.*, "The value of multichannel meg and eeg in the presurgical evaluation of 70 epilepsy patients," *Epilepsy research*, vol. 69, no. 1, pp. 80–86, 2006.

[5] D. V. Moretti, F. Babiloni, F. Carducci, F. Cincotti, E. Remondini, P. Rossini, S. Salinari, and C. Babiloni, "Computerized processing of eeg–eog–emg artifacts for multi-centric studies in eeg oscillations and event-related potentials," *International Journal of Psychophysiology*, vol. 47, no. 3, pp. 199–216, 2003.

[6] M. E. Tagluk, N. Sezgin, and M. Akin, "Estimation of sleep stages by an artificial neural network employing eeg, emg and eog," *Journal of medical systems*, vol. 34, no. 4, pp. 717–725, 2010.

[7] S. Nonoue, M. Mashita, S. Haraki, A. Mikami, H. Adachi, H. Yatani, A. Yoshida, M. Taniike, and T. Kato, "Inter-scorer reliability of sleep assessment using eeg and eog recording system in comparison to polysomnography," *Sleep and Biological Rhythms*, vol. 15, no. 1, pp. 39–48, 2017.

[8] R. M. Pressler, S. Seri, N. Kane, T. Martland, S. Goyal, A. Iyer, E. Warren, L. Notghi, P. Bill, R. Thornton *et al.*, "Consensus-based guidelines for video eeg monitoring in the pre-surgical evaluation of children with epilepsy in the uk," *Seizure*, vol. 50, pp. 6–11, 2017.

[9] W. Klonowski, "Everything you wanted to ask about eeg but were afraid to get the right answer," *Nonlinear Biomedical Physics*, vol. 3, no. 1, p. 2, 2009.

[10] W. Klonowski, "Application of new non-linear dynamics methods in biosignal analysis," in *World Medical Conference*, vol. 5975, 2006, pp. 335–344.

[11] E. Başar, C. Başar-Eroglu, S. Karakaş, and M. Schürmann, "Gamma, alpha, delta, and theta oscillations govern cognitive processes," *International journal of psychophysiology*, vol. 39, no. 2, pp. 241–248, 2001.

[12] B. Hutcheon and Y. Yarom, "Resonance, oscillation and the intrinsic frequency preferences of neurons," *Trends in neurosciences*, vol. 23, no. 5, pp. 216–222, 2000.

[13] G. Pfurtscheller and F. L. Da Silva, "Event-related eeg/meg synchronization and desynchronization: basic principles," *Clinical neurophysiology*, vol. 110, no. 11, pp. 1842–1857, 1999.

[14] H. Adeli, S. Ghosh-Dastidar, and N. Dadmehr, "A wavelet-chaos methodology for analysis of EEGs and EEG subbands to detect seizure and epilepsy." *IEEE Trans. Biomed. Eng.*, vol. 54, no. 2, pp. 205–11, mar 2007.

[15] A. Von Stein and J. Sarnthein, "Different frequencies for different scales of cortical integration: from local gamma to long range alpha/theta synchronization," *International journal of psychophysiology*, vol. 38, no. 3, pp. 301–313, 2000.

[16] W. Klimesch, "Eeg alpha and theta oscillations reflect cognitive and memory performance: a review and analysis," *Brain research reviews*, vol. 29, no. 2-3, pp. 169–195, 1999.

[17] P. K. Sangra, "Classification of electroencephalography signals using mixture of features," Ph.D. dissertation, 2011. [Online]. Available: http://ethesis.nitrkl.ac.in/2591/http://ethesis.nitrkl.ac.in/2591/1/107EC023.pdf

[18] A. K. Engel and W. Singer, "Temporal binding and the neural correlates of sensory awareness," *Trends in cognitive sciences*, vol. 5, no. 1, pp. 16–25, 2001.

[19] M. Bhatia, P. Sinha, S. Jain, M. Padma, and M. Maheshwari, "Usefulness of short-term video eeg recording with saline induction in pseudoseizures," *Acta Neurologica Scandinavica*, vol. 95, no. 6, pp. 363–366, 1997.

[20] M. Salinsky, R. Kanter, and R. M. Dasheiff, "Effectiveness of multiple eegs in supporting the diagnosis of epilepsy: an operational curve," *Epilepsia*, vol. 28, no. 4, pp. 331–334, 1987.

[21] F. Lotte, M. Congedo, A. Lécuyer, F. Lamarche, and B. Arnaldi, "A review of classification algorithms for eeg-based brain–computer interfaces," *Journal of neural engineering*, vol. 4, no. 2, p. R1, 2007.

[22] S. Koelstra, C. Muhl, M. Soleymani, J.-S. Lee, A. Yazdani, T. Ebrahimi, T. Pun, A. Nijholt, and I. Patras, "Deap: A database for emotion analysis; using physiological signals," *IEEE Transactions on Affective Computing*, vol. 3, no. 1, pp. 18–31, 2012.

[23] A. J. Rowan and E. Tolunsky, *A primer of EEG: with a mini-atlas*. Butterworth-Heinemann Medical, 2003.

[24] C. S. Herrmann, D. Strüber, R. F. Helfrich, and A. K. Engel, "Eeg oscillations: from correlation to causality," *International Journal of Psychophysiology*, vol. 103, pp. 12–21, 2016.

[25] A. Rechtschaffen, "A manual of standardized terminology, techniques and scoring system for sleep stages of human subjects," *Public health service*, 1968.

[26] S. Noachtar, C. Binnie, J. Ebersole, F. Mauguiere, A. Sakamoto, and B. Westmoreland, "A glossary of terms most commonly used by clinical electroencephalographers and proposal for the report form for the eeg findings. the international federation of clinical neurophysiology." *Electroencephalography and clinical neurophysiology. Supplement*, vol. 52, p. 21, 1999.

[27] L. J. Greenfield, J. D. Geyer, and P. R. Carney, *Reading EEGs: A practical approach*. Lippincott Williams & Wilkins, 2012.

[28] E. Oropesa, H. L. Cycon, and M. Jobert, "Sleep stage classification using wavelet transform and neural network," *International computer science institute*, 1999.

[29] Y. Liu, L. Yan, B. Zeng, and W. Wang, "Automatic sleep stage scoring using Hilbert-Huang transform with BP neural network," in *Proceedings of the 4th International Conference on Bioinformatics and Biomedical Engineering (iCBBE)*, Jun. 2010, pp. 1–4.

[30] N. Liu, Z. Lu, B. Xu, and Q. Liao, "Learning a convolutional neural network for sleep stage classification," in *Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI), 2017 10th International Congress on*. IEEE, 2017, pp. 1–6.

[31] R. G. Andrzejak, K. Lehnertz, F. Mormann, C. Rieke, P. David, and C. E. Elger, "Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state," *Physical Review E*, vol. 64, no. 6, 1, 2001.

[32] A. L. Goldberger, L. A. N. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C. K. Peng, and H. E. Stanley, "PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, pp. 215–220, 2000.

[33] S. Fridli and F. Schipp, "Biorthogonal systems to rational functions," *Annales Univ. Sci. Budapest., Sect. Comp*, vol. 35, pp. 95–105, 2011.

[34] P. S. C. Heuberger, P. M. J. Van den Hof, and B. Wahlberg, *Modelling and Identification with Rational Orthogonal Basis Functions*. London, UK: Springer-Verlag, 2005.

[35] P. Kovács, S. Kiranyaz, and M. Gabbouj, "Hyperbolic particle swarm optimization with application in rational identification," in *Signal Processing Conference (EUSIPCO), 2013 Proceedings of the 21st European*, Sep. 2013, pp. 1–5.

[36] S. S. Chen, D. L. Donoho, and M. S. Saunders, "Atomic decomposition by basis pursuit," *SIAM Journal on Scientific Computing*, vol. 20, no. 1, pp. 33–61, 1998.

[37] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM Journal on Imaging Sciences*, vol. 2, no. 1, pp. 183–202, 2009.

[38] M. V. Afonso, J. M. Bioucas-Dias, and M. A. T. Figueiredo, "Fast image recovery using variable splitting and constrained optimization," *IEEE Transactions on Image Processing*, vol. 19, no. 9, pp. 2345–2356, 2010.

[39] S. M. Pincus, "Approximate entropy as a measure of system complexity." *Proceedings of the National Academy of Sciences*, vol. 88, no. 6, pp. 2297–2301, 1991.

[40] J. Xiang, C. Li, H. Li, R. Cao, B. Wang, X. Han, and J. Chen, "The detection of epileptic seizure signals based on fuzzy entropy," *Journal of Neuroscience Methods*, vol. 243, pp. 18–25, Mar. 2015.

[41] R. Yadav, A. Shah, J. Loeb, M. Swamy, and R. Agarwal, "Morphology-based automatic seizure detector for intracerebral eeg recordings," *IEEE Transactions on Biomedical Engineering*, vol. 59, no. 7, pp. 1871–1881, 2012.

[42] J. Gotman and P. Gloor, "Automatic recognition and quantification of interictal epileptic activity in the human scalp eeg," *Electroencephalography and clinical neurophysiology*, vol. 41, no. 5, pp. 513–529, 1976.

[43] R. Haralick, K. Shanmugam, and I. Dinstein, "Textural features for image classification," *IEEE Transactions on Systems, Man and Cybernetics*, vol. SMC-3, no. 6, pp. 610–621, 1973.

[44] P. Stoica, R. L. Moses *et al.*, *Spectral analysis of signals*. Pearson Prentice Hall Upper Saddle River, NJ, 2005, vol. 1.

[45] M. S. Bartlett, "Periodogram analysis and continuous spectra," *Biometrika*, vol. 37, no. 1/2, pp. 1–16, 1950.

[46] R. B. Blackman and J. W. Tukey, "The measurement of power spectra from the point of view of communications engineering—part i," *Bell Labs Technical Journal*, vol. 37, no. 1, pp. 185–282, 1958.

[47] P. Welch, "The use of fast fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms," *IEEE Transactions on audio and electroacoustics*, vol. 15, no. 2, pp. 70–73, 1967.

[48] J. del R Millan, J. Mouriño, M. Franzé, F. Cincotti, M. Varsta, J. Heikkonen, and F. Babiloni, "A local neural classifier for the recognition of eeg patterns associated to mental tasks," *IEEE transactions on neural networks*, vol. 13, no. 3, pp. 678–686, 2002.

[49] K. Polat and S. Güneş, "Classification of epileptiform EEG using a hybrid system based on decision tree classifier and fast Fourier transform," *Applied Mathematics and Computation*, vol. 187, no. 2, pp. 1017–1026, Apr. 2007, 00144.

[50] C.-J. Lin and M.-H. Hsieh, "Classification of mental task from eeg data using neural networks based on particle swarm optimization," *Neurocomputing*, vol. 72, no. 4-6, pp. 1121–1130, 2009.

[51] A. Zandi, M. Javidan, G. Dumont, and R. Tafreshi, "Automated real-time epileptic seizure detection in scalp EEG recordings using an algorithm based on wavelet packet transform," *IEEE Transactions on Biomedical Engineering*, vol. 57, no. 7, pp. 1639–1651, Jul. 2010.

[52] H. Adeli, Z. Zhou, and N. Dadmehr, "Analysis of EEG records in an epileptic patient using wavelet transform," *Journal of Neuroscience Methods*, vol. 123, no. 1, pp. 69–87, Feb. 2003.

[53] L. Guo, D. Rivero, and A. Pazos, "Epileptic seizure detection using multiwavelet transform based approximate entropy and artificial neural networks," *Journal of Neuroscience Methods*, vol. 193, no. 1, pp. 156–163, Oct. 2010.

[54] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 7, pp. 971–987, 2002.

[55] T. Senechal, V. Rapp, H. Salam, R. Seguier, K. Bailly, and L. Prevost, "Combining AAM coefficients with LGBP histograms in the multi-kernel SVM framework to detect facial action units," in *Proceedings of the IEEE International Conference on Automatic Face Gesture Recognition and Workshops (FG)*, 2011, pp. 860–865.

[56] W. Zhang, S. Shan, W. Gao, X. Chen, and H. Zhang, "Local Gabor binary pattern histogram sequence (LGBPHS): a novel non-statistical model for face representation and recognition," in *Proceedings of the 10th IEEE International Conference on Computer Vision (ICCV)*, vol. 1, 2005, pp. 786–791.

[57] A. Higashi, T. Yasui, Y. Fukumizu, and H. Yamauchi, "Local Gabor directional pattern histogram sequence (LGDPHS) for age and gender classification," in *Proceedings of the IEEE Statistical Signal Processing Workshop (SSP)*, 2011, pp. 505–508.

[58] W. Hu, B. Ma, and X. Chai, "Head pose estimation using simple local Gabor binary pattern," in *Proceedings of the 6th Chinese Conference on Biometric Recognition*, 2011, pp. 74–81.

[59] S. Xie, S. Shan, X. Chen, and W. Gao, "V-LGBP: Volume based local Gabor binary patterns for face representation and recognition," in *Proceedings of the 19th International Conference on Pattern Recognition (ICPR)*, 2008, pp. 1–4.

[60] H. V. Nguyen, L. Bai, and L. Shen, "Local gabor binary pattern whitened PCA: A novel approach for face recognition from single image per person," in *Advances in Biometrics*, ser. LNCS, M. Tistarelli and M. S. Nixon, Eds. Germany: Springer-Verlag Berlin Heidelberg, 2009, vol. 5558, pp. 269–278.

[61] T. R. Almaev and M. F. Valstar, "Local Gabor binary patterns from three orthogonal planes for automatic facial expression recognition," in *Proceedings of the 5th Humaine Association Conference on Affective Computing and Intelligent Interaction (ACII)*, 2013, pp. 356–361.

[62] D. J. Heeger, "Model for the extraction of image flow," *Journal of the Optical Society of America-A*, vol. 4, no. 8, pp. 1455–1471, 1987.

[63] D. Gabor, "Theory of communication. Part I: The analysis of information," *Journal of the Institution of Electrical Engineers – Part III: Radio and Communication Engineering*, vol. 93, no. 26, pp. 429–441, 1946.

[64] R. M. Haralick and K. S. Shanmugam, "Combined spectral and spatial processing of ERTS imagery data," *Remote Sensing of Environment*, vol. 3, no. 1, pp. 3–13, 1974.

[65] L. Soh and C. Tsatsoulis, "Texture analysis of SAR sea ice imagery using gray level co-occurrence matrices," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 37, no. 2, pp. 780–795, Mar. 1999.

[66] F. Mormann, T. Kreuz, C. Rieke, R. G. A., A. Kraskov, P. David, C. E. Elger, and K. Lehnertz, "On the predictability of epileptic seizures," *Clinical Neurophysiology*, vol. 116, no. 3, pp. 569 – 587, Mar. 2005.

[67] N. Kehtarnavaz, *Digital Signal Processing System Design (Second Edition)*. Academic Press, 2008.

[68] I. Güler and E. D. Übeyli, "Adaptive neuro-fuzzy inference system for classification of EEG signals using wavelet coefficients," *Journal of Neuroscience Methods*, vol. 148, no. 2, pp. 113 – 121, Oct. 2005.

[69] K. A. H. Kulasuriya and M. U. S. Perera, "Forecasting epileptic seizures using EEG signals, wavelet transform and artificial neural networks," in *Proceedings of the International Symposium on IT in Medicine and Education (ITME)*, vol. 1, 2011, pp. 557 – 562.

[70] S. Aydin, H. M. Saraoğlu, and S. Kara, "Log energy entropy-based EEG classification with multilayer neural networks in seizure," *Annals of biomedical engineering*, vol. 37, no. 12, pp. 2626 – 2630, 2009.

[71] V. Srinivasan, C. Eswaran, and N. Sriraam, "Approximate entropy-based epileptic EEG detection using artificial neural networks," *IEEE Transactions on Information Technology in Biomedicine*, vol. 11, no. 3, pp. 288 – 295, 2007.

[72] G. Zhu, Y. Li, and P. Wen, "An efficient visibility graph similarity algorithm and its application on sleep stages classification," in *Proceedings of the International Conference on Brain Informatics*, ser. LNCS, Z. F. Massimo, et al., Ed. Heidelberg, Germany: Springer-Verlag Berlin, 2012, vol. 7670, pp. 185 – 195.

[73] H. Phan, Q. Do, T. L. Do, and D. L. Vu, "Metric learning for automatic sleep stage classification," in *Proceedings of the 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2013, pp. 5025 – 5028.

[74] F. Ebrahimi, M. Mikaeili, E. Estrada, and H. Nazeran, "Automatic sleep stage classification based on EEG signals by using neural networks and wavelet packet coefficients," in *30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2008. EMBS 2008*, Aug. 2008, pp. 1151 – 1154.

[75] Y. Li, F. Yingle, L. Gu, and T. Qinye, "Sleep stage classification based on EEG Hilbert-Huang transform," in *Proceedings of the 4th IEEE Conference on Industrial Electronics and Applications (ICIEA)*, 2009, pp. 3676 – 3681.

[76] V. Srinivasan, C. Eswaran, and N. Sriraam, "Artificial neural network based epileptic detection using time-domain and frequency-domain features," *Journal of Medical Systems*, vol. 29, no. 6, pp. 647 – 660, Dec. 2005.

[77] J. Gotman and L. Y. Wang, "State-dependent spike detection: concepts and preliminary results," *Electroencephalography and Clinical Neurophysiology*, vol. 79, no. 1, pp. 11 – 19, Jul. 1991.

[78] V. P. Nigam and D. Graupe, "A neural-network-based detection of epilepsy," *Neurological Research*, vol. 26, no. 1, pp. 55 – 60, Jan. 2004.

[79] I. Guyon, J. Weston, S. Barnhill, and V. Vapnik, "Gene selection for cancer classification using support vector machines," *Machine Learning*, vol. 46, no. 1-3, pp. 389–422, Jan. 2002.

[80] S. Kiranyaz, T. Ince, M. Zabihi, and D. Ince, "Automated patient-specific classification of long-term Electroencephalography," *Journal of Biomedical Informatics*, vol. 49, pp. 16–31, Jun. 2014.

[81] N. Rafiuddin, Y. Khan, and O. Farooq, "Feature extraction and classification of EEG for automatic seizure detection," in *2011 International Conference on Multimedia, Signal Processing and Communication Technologies (IMPACT)*, Dec. 2011, pp. 184–187.

[82] P. Fergus, D. Hignett, A. Hussain, D. Al-Jumeily, and K. Abdel-Aziz, "Automatic Epileptic Seizure Detection Using Scalp EEG and Advanced Artificial Intelligence Techniques," *BioMed Research International*, vol. 2015, p. e986736, Jan. 2015, 00000.

[83] M. Behnam and H. Pourghassem, "Singular Lorenz Measures Method for seizure detection using KNN-Scatter Search optimization algorithm," in *2015 Signal Processing and Intelligent Systems Conference (SPIS)*, Dec. 2015, pp. 67–72.

[84] A. H. Shoeb, "Application of machine learning to epileptic seizure onset detection and treatment," Thesis, Massachusetts Institute of Technology, 2009, 00093.

[85] Y. U. Khan, O. Farooq, P. Sharma, and N. Rafiuddin, "Latency Study of Seizure Detection," in *Advances in Computer Science, Engineering & Applications*, ser. Advances in Intelligent and Soft Computing, D. C. Wyld, J. Zizka, and D. Nagamalai, Eds. Springer Berlin Heidelberg, Jan. 2012, no. 166, pp. 129–136.

[86] N. Ahammad, T. Fathima, and P. Joseph, "Detection of Epileptic Seizure Event and Onset Using EEG," *BioMed Research International*, vol. 2014, Jan. 2014.

[87] A. Supratak, L. Li, and Y. Guo, "Feature extraction with stacked autoencoders for epileptic seizure detection," in *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Aug. 2014, pp. 4184–4187, 00000.

[88] X. Chen, Z. Syed, and A. Hero, "EEG Spatial Decoding and Classification with Logit Shrinkage Regularized Directed Information Assessment (L-SODA)," *arXiv:1404.0404 [cs, math]*, Apr. 2014, 00000 arXiv: 1404.0404.

[89] A. Van Esbroeck, L. Smith, Z. Syed, S. Singh, and Z. Karam, "Multi-task seizure detection: addressing intra-patient variation in seizure morphologies," *Mach. Learn.*, vol. 102, no. 3, pp. 309–321, mar 2016.

Publications

PUBLICATION
I

On application of rational Discrete Short Time Fourier Transform in epileptic seizure classification

Péter Kovács, Kaveh Samiee, Moncef Gabbouj

IEEE International Conference on Acoustics, Speech and Signal processing (ICASSP), pp. 5839-5843, May 4-9. 2014
https://doi.org/10.1109/ICASSP.2014.6854723

PUBLICATION
II

Epileptic Seizure Classification of EEG
Time-Series Using Rational Discrete Short-Time Fourier Transform

Kaveh Samiee, Péter Kovács, Moncef Gabbouj

IEEE Transactions on Biomedical Engineering, vol. 62, no. 2, pp. 541-552, Feb.2015
https://doi.org/10.1109/TBME.2014.2360101

PUBLICATION III

Long-term epileptic EEG classification via 2D mapping and textural features

Kaveh Samiee, Serkan Kiranyaz, Moncef Gabbouj, Tapio Saramäki

Expert Systems with Applications,
vol. 42, no. 20, pp. 7175-7185, Jun. 2015
https://doi.org/10.1016/j.eswa.2015.05.002

PUBLICATION IV

Sleep stage classification using sparse rational decomposition of single channel EEG records

Kaveh Samiee, Péter Kovács, Serkan Kiranyaz, Moncef Gabbouj, Tapio Saramäki

IEEE 23rd European Signal Processing Conference (EUSIPCO),
pp. 1905-1909, Aug. 2015
https://doi.org/10.1109/EUSIPCO.2015.7362706

PUBLICATION V

Epileptic seizure detection in long-term EEG records using sparse rational decomposition and local Gabor binary patterns feature extraction

Kaveh Samiee, Péter Kovács, Moncef Gabbouj

Knowledge-Based Systems, vol. 118, pp. 228-240, Feb . 2017
https://doi.org/10.1016/j.knosys.2016.11.023