Rochester Institute of Technology

RIT Scholar Works

Theses

9-2015

Automatic Localization of Epileptic Spikes in EEGs of Children with Infantile Spasms

Supachan Traitruengsakul
sxt2021@rit.edu

Follow this and additional works at: <https://scholarworks.rit.edu/theses>

Recommended Citation

Traitruengsakul, Supachan, "Automatic Localization of Epileptic Spikes in EEGs of Children with Infantile Spasms" (2015). Thesis. Rochester Institute of Technology. Accessed from

This Thesis is brought to you for free and open access by RIT Scholar Works. It has been accepted for inclusion in Theses by an authorized administrator of RIT Scholar Works. For more information, please contact ritscholarworks@rit.edu.

R·I·T

**Automatic Localization of Epileptic Spikes in
EEGs of Children with Infantile Spasms**

by

Supachan Traitruengsakul

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of
Master of Science in Electrical Engineering

Department of Electrical and Microelectronic Engineering
Kate Gleason College of Engineering

Rochester Institute of Technology
Rochester, New York
September 2015

Automatic Localization of Epileptic Spikes in EEGs of Children with Infantile Spasms

by

Supachan Traitruengsakul

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of Master of
Science
in Electrical Engineering

Supervised by

Dr. Behnaz Ghoraani
Department of Biomedical Engineering
Kate Gleason College of Engineering
Rochester Institute of Technology
Rochester, New York
September 2015

Co-supervisor: Dr. Laurie E. Seltzer, D.O.
University of Rochester Medical Center, Rochester, NY

Approved by:

Dr. Behnaz Ghoraani, Assistant Professor
Thesis Advisor, Department of Biomedical Engineering

Dr. Daniel Phillips, Department Head and Associate Professor
Committee Member, Department of Biomedical Engineering

Dr. Sohail Dianat, Department Head and Professor
Committee Member, Department of Electrical Engineering

Thesis Release Permission Form

Rochester Institute of Technology
Kate Gleason College of Engineering

Title:

Automatic Localization of Epileptic Spikes in EEGs of Children with Infantile Spasms

I, Supachan Traitruengsakul, hereby grant permission to the Wallace Memorial Library to reproduce my thesis in whole or part.

Supachan Traitruengsakul

Date

© Copyright 2015 by Supachan Traitruengsakul
All Rights Reserved

Dedication

To my parents, Wichai and Sopa Traitruengsakul, who said...
 “Where there’s a will, there’s a way”
 and
 “Every obstacle is surmountable.”

Acknowledgments

The research for this thesis was supported by my advisor, Dr. Behnaz Ghorrani, who drives me to achieve my goals and has been an extreme inspiration to me in various ways. I would like to acknowledge her and the committees, Drs. Dan Philips, and Sohail Dianat, for agreeing on my thesis. In addition, I also would like to acknowledge Dr. Laurie E. Seltzer who corroborates with me. Finally, many thanks to my parents, families, and staffs who work at Office of Educational Affairs, Royal Thai Embassy, Washington DC.

Abstract

Automatic Localization of Epileptic Spikes in EEGs of Children with Infantile Spasms

Supachan Traitruengsakul

Supervising Professor: Dr. Behnaz Ghoraani

Infantile Spasms (ISS) characterized by electroencephalogram (EEG) recordings exhibiting hypsarrhythmia (HYPS) are a severe form of epilepsy. Many clinicians have been trying to improve ISS outcomes; however, quantification of discharges from hypsarrhythmic EEG readings remains challenging.

This thesis describes the development of a novel method that assists clinicians to successfully localize the epileptic discharges associated with ISS in HYPS. The approach includes: construct the time-frequency domain (TFD) of the EEG recording using matching pursuit TFD (MP-TFD), decompose the TFD matrix into two submatrices using non-negative matrix factorizations (NMF), and employ the decomposed vectors to locate the spikes.

The proposed method was employed to an EEG dataset of five ISS individuals, and identification of spikes was compared with those which were identified by the epileptologists and those obtained using clinical software (Persyst). Performance evaluations showed results based on classification techniques: thresholdings, and support vector machine (SVM). Using the thresholdings, average true positive (TP) and false negative (FN) percentages of 86% and 14% were achieved, which represented a significant improvement over the use of Persyst, which only achieved average TP and FN percentages of 4% and 96%, respectively. Using SVM, the percentage of area under curve (AUC) of receiver operating characteristic (ROC) was significantly improved up to 98.56%.

In summary, the proposed novel algorithm based on MP-TFD and NMF was able to successfully detect the epileptic discharges from the dataset. The development of the proposed automated method can potentially assist clinicians to successfully localize the epileptic discharges associated with ISS in HYPS. The quantitative assessment of spike detection, as well as other features of HYPS, is expected to allow a more accurate assessment of the relevance of EEG to clinical outcomes, which is significant in therapy management of ISS.

Contents

Dedication	iv
Acknowledgments	v
Abstract	vi
1 Introduction	1
1.1 Motivation	1
1.2 Objectives	3
1.3 Contribution	3
1.4 Thesis Organization	4
2 Background and Fundamentals	6
2.1 Part A: Signal Analysis	6
2.1.1 Electrode Placements of Electroencephalogram	6
2.1.2 Characteristics of HYPS	6
2.1.3 A Signal in Time and Frequency Representations	9
2.1.4 Introduction to Joint Time-Frequency Representations	9
2.2 Part B: Matrix Decomposition	13
2.3 Part C: Feature Extraction	15
2.4 Part D: Classification and Performance	16
2.4.1 Classifiers	16
2.4.2 Assessment Methods	16
2.4.3 Performance Measures	17
3 Methods	19
3.1 Part A: Data Acquisition	20
3.2 Part B: Proposed Methods	21
3.2.1 Method I	22
3.2.2 Method II	24
3.2.3 Method III	29

4	Results and Discussion	33
4.1	Method I	33
4.2	Method II	37
4.3	Method III	44
4.4	Results Summary	44
5	Conclusions	47
	Bibliography	49

List of Tables

2.1	A confusion matrix applied for spike detection algorithms	18
3.1	Proposed methods for this thesis.	21
3.2	Two probabilistic models based on automatic determination relevance for β -NMF [59]. μ_V denotes mean of matrix V	25
4.1	Results of detection of spike locations are calculated in Eqs.(3.12)-(3.14) in Chapter 3. True Positive (TP). False Negative (FN). False Positive (FP). The number in each parenthesis indicates the number of spikes. . . .	36
4.2	Comparison results of performance between Methods I and II . Method II uses parameters $\beta = 1$, $a = 0.5$ and thresholds $Th_w = 8-19$ Hz, and $Th_{Aw} = 12$. Parentheses indicate number of spikes. Shadow areas on the table represent significant performance.	39
4.3	Results of performance based on K-Fold with different σ and comparison between F1 , F2 . Acc, Prec, Sens, Spec, AUC are denoted accuracy, precision, sensitivity, specificity and the area under the curve of ROC, respectively.	45

List of Figures

1.1	A representative bipolar montage EEG recording exhibiting HYPS. The EEG appears highly disorganized with high amplitude waveforms, and contains multifocal discharges with variable morphologies.	3
1.2	Flow chart of background and fundamentals in this thesis. More details are described in Chapter 2.	5
2.1	The general montages according to international 10-20 system are employed in clinical practice [48]. (a) Electrode placement. (b) Longitudinal bipolar montages are employed to this thesis for further analysis. (c) Transverse bipolar montages. (d) Referential montages.	7
2.2	Comparison between normal and abnormal (HYPS) EEGs in bipolar montages. (a) Normal EEG characterizes low voltage, and well-organized (uniform), while (b) HYPS performs high voltages or discharges, and disorganized (non-uniform and chaotic).	8
2.3	Illustration of two chirps, $x_1(t)$ and $x_2(t)$. Over time, (a) normalized frequency of the chirp, $x_1(t)$, increases from 0.05 to 0.30 while (b) normalized frequency of the chirp, $x_2(t)$, decreases from 0.30 to 0.05. The results of power spectrum, $X_1(f)$ and $X_2(f)$, as depicted in (c) and (d). The results of spectrograms, $STFT_{x_1}(t, f)$ and $STFT_{x_2}(t, f)$, using width $\frac{N}{4}$ of Hamming window as shown in (e) and (f).	10
2.4	The results of using different widths of a window and each sub-rectangle is called a logon. (a) Using short window provides better time resolution; however, frequency resolution is not well localized and (b) vice versa. (c) Example of a logon which is related to Heisenberg's uncertainty principle depicted from (b).	11
2.5	Comparison the resolutions between STFT and WVD of two chirps. (a) Resolution of STFT using width, $\frac{N}{4}$, of Hamming window provides the auto-terms as blurred image, but it does not appear cross-term. (b) WVD's resolution displays auto-terms as ridge image, but having extra cross-terms.	13

2.6	Comparison TF resolutions of the EEG using several techniques of TFDs. The signal is plotted on the top of each subfigure. (a) STFT with long Hamming window length $\frac{N}{4}$. (b) STFT with short Hamming window length $\frac{N}{10}$. (c) WVD provides multiple cross-terms and perplex interpretation. (d) MP exhibits no cross-terms, and information about transients of the signal, such as spikes.	14
2.7	General theme of NMF shows decomposition of MP into spectral patterns (basis) and temporal activations (coefficient), W and H , respectively.	15
2.8	Assessment methods including (a) K-Fold and (b) leave-one-out cross validations. Exp_k , where $k = 1, \dots, K$, denotes k^{th} experiment.	17
2.9	Example of two classes ($\circ$ and $\bullet$). (a) Feature vectors of the classes are not well separated in low dimensional space. (b) Using transformation ϕ maps the feature vectors into higher dimensional space in order to be well-separated with linear hyperplane [67].	18
3.1	General scheme of the proposed methods.	19
3.2	Block diagram of Method I	23
3.3	The illustration of feature extraction for Methods II and III shows how to calculate and tabulate the features. (a) Features of spectral vector (w_k) contains area (A_w) and mean frequency (MF_w) values. To calculate area of w_k , the vector has to be normalized in order to keep the feature because of magnitude-invariance among the entire data. (b) The vector with two distributions has multiple features denoted as A_d , B_d , C_d , where d is a number of distribution (in this example, $d = 2$). The thick-dashed vertical line relied on the first distribution and associated with spike marked by the epileptologist is labeled as class “1”. Otherwise, class “0” indicates non-spike. (c) Table contains features of spectral and temporal vectors, including class labels.	28
3.4	Illustration of classification models as a preprocessing step of classifier. Each of spikes and non-spikes having multiple features are split into two subsets for training (80%) and testing (20%) using K-Fold model.	32
3.5	ROC curve represents different levels of performance indicated by direction of the arrow (from the worst to better performance).	32

- 4.1 The step by step illustration of the first proposed algorithm. (a) 10-second EEG recording. (b) The MP-TFD of the EEG recording. (c) and (e) The decomposed spectral and temporal matrices, respectively. The table shows the MF value of each spectral vector and w_3^* which has the highest MF value. (d) The temporal vector corresponding to h_3^* represented by the dashed lines indicate the spike locations. 34
- 4.2 The top displays EEG signal ($x(t)$), the bottom represents h_k^* . Spike locations detected by the h_k^* peaks that are greater than threshold Th_h . The dashed lines indicate the proposed locations of the spikes and arrows on top of the $x(t)$ indicate the spikes, which were marked by the epileptologist. 35
- 4.3 Differences between original MP and approximation of decomposed factors. (a) Normalized MP is decomposed and each element values are very small as shown in color bar. (b) The results of using β -NMF based on l_1 -ARD provides ill-reconstruction error of $\mathbf{W}$, $\mathbf{H}$ when initial $K = 30$ and the resulting $K_{\text{eff}} = 1$. For further analysis of **Method II**, all MPs are multiplied by 1000, and (c) the results show better reconstruction error when using initial $K = 30$ and resulting $K_{\text{eff}} = 10$. The parameters were set $\{a, tol, I\} = \{0.5, 10^{-6}, 3500\}$. (d) using NMF based on projected gradient (PG), decomposition of normalized MP shows improved reconstruction error for **Method III**. The parameters were set $\{K_{\text{eff}}, tol, I\} = \{10, 10^{-15}, 3500\}$ 38
- 4.4 Results of number of K_{eff} versus hyper parameter, $a = \{0.5, 100, 500, 1000\}$. Vertical and horizontal axes of each subplot represent K_{eff} and a , respectively. Each row indicates each patient, and columns from left to right indicate number of spikes $\{0, 1, 2, \dots, 6\}$. The title of each sub-figure shows the number of samples per ten seconds. 40

- 4.5 Result of using NMF based on l_1 -ARD with parameters $\beta = 1$, $a = 0.5$. (a) Original MP with spike locations indicated as dashed lines marked by the epileptologist. MP is decomposed to $\mathbf{W}$, $\mathbf{H}$ as shown (b), (c), respectively. (b) Spectral vectors sorted decrement values of MF, and the ranges of threshold (Th_w) are indicated by two horizontal dashed lines. Horizontal and vertical axes represent index k^{th} of w , and frequency [Hz]. (c) Temporal vectors are reordered corresponding to the index number of w_k and show spike locations as dashed lines. Horizontal and vertical axes represent time [sec], and index k^{th} of h . (d) Table contains index number, $MF(w_k)$, and area, $A(w_k)$. Notice that, after initial K was 30, K_{eff} is 10 and the remains are removed due to irrelevance. 41
- 4.6 Result of another example based on l_1 -ARD with parameters $\beta = 1$, $a = 0.5$. Following descriptions are similar to Figure 4.5. (a) Original MP. (b) Spectral vectors. (c) Temporal vectors. (d) Table contains index number, $MF(w_k)$, and area, $A(w_k)$. As seen in the table, w_4 and w_7 are in the conditions of the thresholds, and h_4 and h_7 are possibly indicated as spikes. . . . 42
- 4.7 The result of common temporal vectors between h_4 and h_9 , depicted from Figure 4.5(c), as shown in (a), (b). (c) The result of the common h of the two vectors using Eqs.(3.20)-(3.22) in algorithm of feature extraction, and shows six correct spikes, three non spikes. 43
- 4.8 The result of another example with common temporal vectors between h_4 and h_{17} , depicted from Figure 4.6(c), as shown in (a), (b). (c) The result of common h displays three correct spikes, three non spikes and one missed spike. 43
- 4.9 Results of ROC curves using K-Fold cross validation with parameter $\sigma = 10$ per feature set, **F1** and **F2**. From the worst to the best performance, AUCs of **F1** (a) and **F2** (b) are (90.38%-98.51%), and (93.14%-98.56%), respectively. 46

Chapter 1

Introduction

1.1 Motivation

The incidence of infantile spasms (ISS), which is a severe epilepsy, occurs in range between 2 to 3.5 per 10,000 live births [29], [56]. In 1841, ISS or West syndrome was originally described by Dr William James West who diagnosed his son as a patient with this disorder [66]. West observed that, “(ISS) causes a complete heaving of the head forward towards his knees, and then immediately relaxing into the upright position ... these bowings and relaxings would be repeated alternately at intervals of a few seconds, and repeated from 10 to 20 or more times at each attack, with each attack not continuing more than 2 or 3 minutes; he sometimes has 2, 3 or more attacks in the day [66].”

In other words, ISS relates to contractions of muscle of neck, trunk, and extremities and often happens in durations of five second clusters. West reported that ISS, if left untreated, can lead to developmental regression and mental retardation [66]. Even though ISS was described 174 years ago, assessment, diagnosis, and management for the disorder are very enigmatic, and physicians have not reached consensus on objective protocols for diagnosing ISS and require the need for consensus [53]. To improve the outcomes, the Infantile Spasms Workshop Group (ISWG) has been established in the USA, and its objectives are to review the consistency of our understanding of ISS, assess the scientific evidence, and reach a consensus on protocols for diagnostic workup and management of ISS [53]. However, ISWG still requires improvement of ISS outcomes, such as, electroencephalogram (EEG) reading. Most clinicians read EEG signals to assess whether an individual is diagnosed as a normal or abnormal case at the first line of diagnosis for ISS.

ISS refers to a catastrophic form of epilepsy and is diagnosed based on the searching of hypsarrhythmia (HYPS) which is characterized as chaotic and high voltage background with multifocal, discharges described by Gibbs *et al.* [20] in EEG recording combined with epileptic spasms. Hrachovy *et al.* [25] studied diversities of the original pattern or modified HYPS from prolonged 24-hour EEG recordings of 67 patients with ISS, and they observed that variations of HYPS challenge clinicians or neurologists to be enigmatic for diagnosis of ISS. In particular, an expert electroencephalographer interprets an EEG by inspecting and approximating the characteristics of HYPS subjectively rather than through objective quantification. Due to the complex nature of these signals, even experienced EEG readers tend to interpret HYPS differently, which can have serious implications in the treatment

of the infant [2], [27]. The aspects of the EEG signals that are amenable to quantification include frequency, amplitude, symmetry, and presence of stereotypical discharge patterns.

This thesis describes the development of a novel method that may assist clinicians to successfully localize the epileptic discharges associated with ISS in HYPS. Such a method is deemed to be necessary for reliable evaluation of features associated with HYPS, which could potentially improve the assessment of infantile spasms, which is of significant importance in the therapy, management, and ultimately the success of the prescribed treatment.

Several algorithms have been developed to detect epileptic discharges during epilepsy. Some algorithms including temporal analysis based on template matching, and mimetic analysis methods or time-frequency (TF) method based on wavelet analysis are summarized:

- **Methods based on template matching techniques:** This technique requires an EEG reader to pre-specify spike characteristics such as amplitude and duration of the discharges. Once those values are defined, the waveform is detected and recognized by the algorithm. In addition, when obtaining new data, the algorithm can adapt with the spike variations [28], [38], [43].
- **Methods based on mimetic techniques:** Mimetic methods describe the characteristic of waveforms in EEG recordings. For example, a spike contains multiple features such as amplitude, slope, duration, *etc.* These specifications are defined by an expert EEG reader [31], [42], [68].
- **Methods based on time-frequency techniques:** This method represents transformation of a signal in the time domain to the joint time-frequency domain as an image. Using a wavelet transform technique provides a scalogram showing coefficient values of a signal onto scale and translation. For example, while the higher scale shows low resolution in the time domain, the lower scale represents a high temporal resolution and is suitable to spike detection [49].

However, those methods have been developed for epileptic discharge detection in EEG signals associated with other types of epilepsy, but not in the presence of HYPS. The existing methods are either based on supervised classifiers in which the presence of true spikes has to be readily available and identifiable to train the algorithm during a learning phase, or they are template based. Given the chaotic appearance of EEG during HYPS (see Figure 1.1), the manual localization of true spikes is not always possible. Characteristics of the spikes are non-uniform, which presents a challenge for temporal-based methods. Therefore, there is a need to develop an unsupervised feature extraction and classification method to assist in localization of spikes with multiple foci and varying morphologies during HYPS.

A novel algorithm, being explained in Chapters 2 and 3, applies time-frequency domain (TFD) analysis and non-negative matrix factorization (NMF) to extract TF features that can locate epileptiform discharges associated with ISS in the presence of HYPS.

1.2 Objectives

The objectives of this thesis are: (i) to identify and quantify the variations of spikes of HYPs in time-frequency representation of EEG signals, (ii) to extract features associated with spike from decomposed factors using NMF, and (iii) to classify classes between spikes and non-spikes in order to evaluate performance of the proposed methods.

1.3 Contribution

The main contribution of this thesis is to detect spikes in HYPS EEG and extract features from TF domain using NMF. The following contributions are described:

1. Enhanced TF resolution of hypsarrhythmic EEG pattern using Matching Pursuit (MP).
2. Extracted HYPS spikes related TF features using NMF.
3. Determined the NMF model parameter in order to get more efficient decomposition.
4. Classified spike and non-spikes using two methods, *i.e.*, thresholdings and support vector machine (SVM).
5. This work was presented at the 37th annual international conference of the IEEE Engineering in Medicine and Biology Society MiCo - Milano Conference Center - Milan, Italy, August 25-29, 2015.

Publication of this work: S. Traitruengsakul, L. E. Seltzer, A. R. Paciorkowski, and B. Ghoraani, Automatic Localization of Epileptic Spikes in EEGs of Children with Infantile Spasms, Accepted at the 37th Annual International IEEE EMBS Conference, 2015.

1.4 Thesis Organization

This thesis consists of five chapters:

- **Chapter 1** (Introduction) provides introduction of HYPS, motivation, general ideas, objectives, and methods of this thesis.
- **Chapter 2** (Background) describes general knowledge about EEG, TF representations of natural signals, and characteristics of discharges with epileptic form and hypsarrhythmic pattern.
- **Chapter 3** (Methods) explains concepts of the proposed methods including with TF domain, feature extractions, and performance evaluation and how to employ them according to the objectives of this thesis.
- **Chapter 4** (Results and discussions) illustrates and evaluates the results obtained from the studies of the objectives.
- **Chapter 5** (Conclusions) summarizes the thesis and explains how the proposed methods accomplish the goals.

As depicted in Figure 1.2, general frameworks of this thesis are described:

- **Part A: Signal Analysis** introduces EEG and characteristic of discharges in HYPS and explains representations of signals analysis in time, frequency, and joint TF domains. In addition, it also accounts for a reason why signal analysis in joint TF domains provides meaningful information rather than the other domains exclusively, and how to select the best resolution of a TF representation being suitable for non-stationary signals, such as, EEG.
- **Part B: Matrix Decomposition** provides general information about data factorization, acquaints the reader with NMF, and explains what representative factors such as coefficient and basis matrices decomposed by NMF are associated to spike in HYPS.
- **Part C: Feature Extraction** informs the reader what features relate to the decomposed matrices, how to calculate mean frequency, geometries of each spectral vector, and how to find spike locations of temporal vector.
- **Part D: Classification and Performance** finally evaluates our performance by calculating true positive rate (% correct spikes), false positive rate (% non-spikes) and false negative rate (% missed spikes), and plotting receiver operating characteristic (ROC-curve) based on cross validations.

Figure 1.2: Flow chart of background and fundamentals in this thesis. More details are described in Chapter 2.

Chapter 2

Background and Fundamentals

This chapter presents the reader general knowledge of the concepts employed within this thesis. Fundamentals and background depicted in Figure 1.2 contain four parts as follows:

2.1 Part A: Signal Analysis

2.1.1 Electrode Placements of Electroencephalogram

The EEG is a tool to measure electrical activity of the brain that records voltages from multiple electrodes located on the scalp depicted in Figure 2.1(a), and also assists neurologists to assess individuals who have malfunctions in their brains [48], [50]. In clinical practice, electrode connections commonly consist of various types of montages: longitudinal bipolar, transverse bipolar, and referential montages as shown in Figures 2.1(b)-(d), respectively.

As seen in Figures 2.1(b), (c), bipolar montages measure different voltages of two electrodes as inputs in referential montages. For example, $P_4 - O_2$ in longitudinal bipolar montages indicates differences between P_4 and O_2 in referential montages (see the arrow lines indicated connections between two electrodes in Figures 2.1(b), (c)). Montage representations allow the neurologists to evaluate their patients by viewing the EEG signals in spatial distribution across the cortex.

2.1.2 Characteristics of HYPs

Figures 2.2 (a), (b) show comparison between normal and abnormal (HYPs) EEGs. Observe that the former characterizes low voltage and well-organized (or uniform) while the latter performs high voltages ($> 150 \mu V$, $< -150 \mu V$), discharges and disorganized (or non-uniform). Reading HYPs provides epileptologists challenges to quantify and identify the discharges because it is time-consuming. Therefore, there is a need to develop an unsupervised feature extraction and classification method to assist in localization of spikes with multiple foci and varying morphologies during HYPs.

Figure 2.2: Comparison between normal and abnormal (HYPS) EEGs in bipolar montages. (a) Normal EEG characterizes low voltage, and well-organized (uniform), while (b) HYPS performs high voltages or discharges, and disorganized (non-uniform and chaotic).

2.1.3 A Signal in Time and Frequency Representations

Signal analysis traditionally requires two classical representations such as time and frequency domains. The former represents how a signal varies with time. For example, Figures 2.3 (a), (b) show two chirps (linear frequency modulation: LFM), $x_1(t)$ and $x_2(t)$, which have frequencies increasing and decreasing over time, but these signals in time domain do not indicate frequency components. The latter (or spectrum) represents the frequency content of a signal located in the range of frequencies and is obtained by employing a Fourier transform (FT) to the signal as following equation:

$$X(f) = \int x(t)e^{-j2\pi ft} dt. \quad (2.1)$$

The above equation integrates the input signal, $x(t)$, over time duration and can be considered if the signal is stationary, which its frequency components are constant over time. In the real world, in contrast, natural signals are non-stationary such as the chirps. Figures 2.3 (c), (d), for instance, display the indifferent results of the spectrum of the chirps, $X_1(f)$ and $X_2(f)$. In fact, it can lead to potential misinterpretations. In other words, the spectrum of the signals provides no indication on how the frequency content of the signals changes with time. Therefore, signal analysis cannot be analyzed individually in time or frequency domains [7], [14].

To overcome this deficiency, joint time-frequency representations (TFRs) such as Short-Time Fourier Transform (STFT), Wigner-Ville Distribution (WVD), and Matching Pursuit (MP) provide a two-dimensional energy representation of the signal in terms of their temporal and spectral content [7], [14]. As seen in Figures 2.3 (e), (f) of the chirps, the results of STFT allow viewers to visualize how frequencies of the signals vary with time.

2.1.4 Introduction to Joint Time-Frequency Representations

A TF representation provides significant information of a signal in both time and frequency domains and has been widely pertained to non-stationary signals in many applications such as speech [19], [54], electrocardiogram (ECG) [57], [1], electroencephalogram (EEG) [8], [24], *etc.*

Short-Time Fourier Transform

Short-Time Fourier Transform (STFT) performs an attempt to apply the Fourier transform (FT) for a sliding short-time analysis window and its equation can be expressed:

$$STFT_x(t, f) = \int x(\tau)h^*(\tau - t)e^{-j2\pi f\tau} d\tau, \quad (2.2)$$

where $x(t)$ is an input signal, $h(t)$ is a window function, and the superscript (*) is complex conjugate. A window function extracts a finite-length portion of the sequence $x(t)$ and

Figure 2.3: Illustration of two chirps, $x_1(t)$ and $x_2(t)$. Over time, (a) normalized frequency of the chirp, $x_1(t)$, increases from 0.05 to 0.30 while (b) normalized frequency of the chirp, $x_2(t)$, decreases from 0.30 to 0.05. The results of power spectrum, $X_1(f)$ and $X_2(f)$, as depicted in (c) and (d). The results of spectrograms, $STFT_{x_1}(t, f)$ and $STFT_{x_2}(t, f)$, using width $\frac{N}{4}$ of Hamming window as shown in (e) and (f).

Figure 2.4: The results of using different widths of a window and each sub-rectangle is called a logon. (a) Using short window provides better time resolution; however, frequency resolution is not well localized and (b) vice versa. (c) Example of a logon which is related to Heisenberg's uncertainty principle depicted from (b).

can be chosen by users such as Hamming, Hanning, Bartlett, *etc.*, [51]. The spectrogram defines as the squared magnitude of the STFT:

$$SPEC_x(t, f) = |STFT_x(t, f)|^2. \quad (2.3)$$

The STFT has drawback of resolution in time and frequency domains according to the window length. In the example of Figures 2.4(a), (b), if the window length is narrow, it produces high resolution in the time domain, but blurred resolution in the frequency domain and vice versa, respectively, [7], [14]. In Gabor's paper [18], resulting different resolutions derive from the time-frequency plane which is divided into an array of rectangles using a filter bank. A logon depicted in Figure 2.4(c) represents each sub-rectangle and has dimensions of decay time and tuning width according to Eq.(2.4) of Heisenberg's uncertainty principle:

$$\Delta t \Delta f \geq \frac{1}{4\pi}, \quad (2.4)$$

where Δt and Δf represent the effective duration and bandwidth of the logon, respectively. In the transform domain, in general, Δt indicates how well the time duration can be isolated from each other while Δf indicates how well the spectral contents can be isolated from each other. In fact, these parameters of the Heisenberg's box are fixed so that they cannot be made smaller in the same time. Even though STFT can be applied to a stationary signal, it cannot be suitable for non-stationary signals due to the trade-off between time and frequency resolutions.

Wigner-Ville Distribution

Wigner-Ville Distribution (WVD) originated by Ville in 1948 [64] provides high energy concentration and resolution in the TF domain for a mono-component signal having one

frequency component. Its equation is defined as:

$$W_z(t, f) = \int z(t + \frac{\tau}{2}) z^*(t - \frac{\tau}{2}) e^{-j2\pi f\tau} d\tau, \quad (2.5)$$

where $z(t)$ is analytic associated with the signal, $x(t)$, obtained by Hilbert transform, $H[s(t)]$, given below [6], [7], [10]:

$$z(t) = x(t) + jH[x(t)]. \quad (2.6)$$

Unlike the resolution of STFT, a limitation of the WVD resolution is cross-terms of a multicomponent signal that may lead to potential misinterpretation. For instance, given two chirps, displayed in Figure 2.5, increasing normalized frequencies of $s_1(t)$, and $s_2(t)$ from 0.05 to 0.25 and from 0.15 to 0.45, respectively, and analytic associated with the signals $z(t) = z_1(t) + z_2(t)$, then

$$W_z(t, f) = W_{z_1}(t, f) + W_{z_2}(t, f) + 2Re[W_{z_1, z_2}(t, f)], \quad (2.7)$$

where $W_{z_1}(t, f)$, $W_{z_2}(t, f)$ are auto-terms, and $W_{z_1, z_2}(t, f)$ is cross-terms having oscillation and located between the auto-terms in TF domain [7], [9].

Many authors have employed several methods, such as smoothed Pseudo Wigner-Ville Distribution (SPWVD) [23], Choi-William Distribution (CWD) [12], modified-Beta Distribution (MBD) [46], and *etc.*, to suppress the cross-terms while retaining the auto-terms of multicomponent signals. They apply various kernel functions with double convolution of time and frequency domain into the Eq.(2.5) as follows:

$$\rho(t, f) = K(t, f) ** W_z(t, f), \quad (2.8)$$

where $K(t, f)$ is a kernel function, $\rho(t, f)$ is a quadratic time-frequency distribution (QTFD) and $(**)$ is double convolution in time and frequency domains [7], [9]. Notice that WVD applies $K(t, f)$ equal to 1 as same as Eq.(2.5). In fact, some non-stationary signals, EEG, cannot be applied to some QTFDs, WVD, SPWVD and CWD, because of the appearance of the cross-terms and low resolutions.

Matching Pursuit

In order to suppress the cross-terms, retain the auto-terms, and improve TF resolutions, Matching Pursuit-TFD (MP-TFD) developed by Mallat and Zhang [47] is suitable for non-stationary signals. MP algorithm conveys an optimal atom generated by Gabor function from a redundant dictionary to match the subsignal. Then, the atom in the time domain is transformed into TF domain by Eq.(2.5). Once all atoms match to the whole signal, each WVD of each atom is combined together as a MP-TFD. MP implementation is comprehensively explained later in Chapter 3.

In this thesis, HYPSE EEG is applied to different TFDs as shown in Figure 2.6 in order

Figure 2.5: Comparison the resolutions between STFT and WVD of two chirps. (a) Resolution of STFT using width, $\frac{N}{4}$, of Hamming window provides the auto-terms as blurred image, but it does not appear cross-term. (b) WVD's resolution displays auto-terms as ridge image, but having extra cross-terms.

to compare TF resolutions. As seen in Figure 2.6, MP is suitable for EEG signal analysis while the other TFDs provide the lack of information about transient of the signal. In this case, it leads to difficult interpretations.

2.2 Part B: Matrix Decomposition

There are various techniques for dimensionality reduction including Singular Value Decomposition (SVD), Principle Component Analysis (PCA), and Non-Negative Matrix Factorization (NMF). Their common objective aims to determine a low-rank factorization of a large data set into low dimensional factors; however, their results of factorizations are different based on the employed constraints. For instance, using SVD or PCA with no non-negative constraints to decompose an image provides their factors having both arbitrary positive and negative components. In this case, it leads to difficult interpretation when those methods are entailed by summing up some basis factors and subtracting others, called holistic representation. On the other hand, NMF with non-negative constraints provides decomposed matrices having non-negative elements. The results of the factors lead a natural interpretation because each factor is analyzed by adding them together as a linear combination. NMF has been successfully employed in time-frequency feature extraction applications because features of the decomposed matrices of NMF represent a TF matrix with better temporal and spectral localization [19], [32], [60], [4], [3].

In general, NMF originated by Paatero and Tapper [52] decomposes non-negative data into two non-negative factors such as coefficient matrix ($\mathbf{H}$) and basis matrix ($\mathbf{W}$). As shown in Eq.(2.9), after NMF is applied to the resulting time-frequency matrix, $\mathbf{MP}$ with non-negative elements, it is possible to approximate it in terms of spectral and temporal matrices, $\mathbf{W}$ and $\mathbf{H}$, respectively. The following equation of NMF can be written as a linear

Figure 2.6: Comparison TF resolutions of the EEG using several techniques of TFDs. The signal is plotted on the top of each subfigure. (a) STFT with long Hamming window length $\frac{N}{4}$. (b) STFT with short Hamming window length $\frac{N}{10}$. (c) WVD provides multiple cross-terms and perplex interpretation. (d) MP exhibits no cross-terms, and information about transients of the signal, such as spikes.

combination of each latent (k) of the decomposed matrices:

$$\mathbf{MP}_{F \times N} \approx \mathbf{W}_{F \times K} \cdot \mathbf{H}_{K \times N} \approx \sum_{k=1}^K w_k h_k^T, \quad (2.9)$$

where non-negativity of $\mathbf{W}$ and $\mathbf{H}$ is the decomposition constraint, F denotes the number of elements in the frequency domain, N denotes the length of signal in time domain, and $K \ll \min(F, N)$ is the NMF decomposition factor, and $[\cdot]^T$ denotes transposition of the vector h_k . Figure 2.7 illustrates NMF in Eq.(2.9).

The decomposed matrices can be obtained by minimizing a cost function of NMF. A simple cost function defines as Euclidean Distance (ED) or Frobenius norm and measures how similarities between original data ($\mathbf{MP}$) and the factors ($\mathbf{W}$, $\mathbf{H}$) are close, and it can be solved by using gradient descent for minimization problem. Chapter 3 explains several objective functions such as ED, β -divergence, and Kullback-Leibler divergence (KLD), and problem solvers including block principal pivoting (BPP), majorization-minimization (MM), and projected gradient (PG).

Figure 2.7: General theme of NMF shows decomposition of $\mathbf{MP}$ into spectral patterns (basis) and temporal activations (coefficient), $\mathbf{W}$ and $\mathbf{H}$, respectively.

2.3 Part C: Feature Extraction

Feature extraction aims to define a set of features with significant information as pre-processing stage of machine learning or classification. For example, characteristics of facial

images contain noses, eyes, ears, colors, *etc.* When they are extracted using NMF, one of the decomposed matrices indicates locations of some characteristics and use of this information is kept into feature vectors for machine learning. The meaningful features exhibit to easily interpret and classify this set of data.

In this thesis, after obtaining the decomposed matrices in Eq.(2.7), (w_k, h_k) were employed to feature extraction and calculate mean frequency of w_k , geometries, which include area, height, and width, of the decomposed factors, and spike locations of h_k . More details of feature extraction are provided in Chapter 3 including how to calculate these features.

2.4 Part D: Classification and Performance

A significant task of classification is to separate data into two subsets: training and testing sets. The training set contains various attributes or features and labeled classes. The testing set only consists of attributes. The objective of a classifier is to generate a model from learned or trained set (called supervised learning), and predict labels from the testing set by learned statistical model of the training set (called unsupervised learning). When the data is partitioned into training and testing sets based on a model selection, as explained in the next subsection as preprocessing of the machine learning, it is called semi-supervised learning [5]. In this thesis, we use semi-supervised learning for training and testing our data between two classes: spike and non-spike where they are labeled classes 1 and 0, respectively.

2.4.1 Classifiers

A simple classifier is a linear separator that easily distinguishes between two classes with maximal margin if features of data are well separated. However, many applications are mixed as ill-separated, and their features require transformation (or kernel function) that maps feature vectors from lower dimensional spaces to higher dimensional spaces in order to be well-separated as depicted in Figure 2.9. This issue can be solved by SVM, originated by Vapnik [62]. To begin with two classes, SVM basically performs to project input feature vectors from lower dimensional spaces Figure 2.9 (a) into higher dimensional spaces Figure 2.9 (b) using a kernel function, and find decision boundary which has the largest margin distance between the classes (see the dash lines).

2.4.2 Assessment Methods

In order to fairly assess the results of performance without overfitting data, as generalization, from partitioned sets, assessment methods such as K-Fold and leave-one-out cross validations are widely used for many applications. The following techniques of the two cross validations are shortly explained below:

- **K-Fold:** As illustrated in Figure 2.8(a), given dataset of size P samples, for the first round, this technique begins to partition randomly and uniformly $\frac{P}{K}$ subsets of

the data including $\frac{(K-1)P}{K}$ groups of training set, and the rest of testing set. Once a classifier learns features of the training sets associated with defined labels, the labels of testing set will be predicted by its learned characteristics or statistics and resulting performance, as explained in the next subsection, will be calculated. For the second round and *etc.*, the technique operates different groups of training set and testing set. All result of the performance will be averaged over K rounds.

- **Leave-One-Out:** This method is very similar to K-fold, but it leaves one sample to be tested and the rest is trained as shown in Figure 2.8(b).

A limitation of the latter method is it requires more expensive computation than the former does if data is very large. This thesis focuses on K-fold technique to evaluate number of spikes and non-spikes of all patients.

Figure 2.8: Assessment methods including (a) K-Fold and (b) leave-one-out cross validations. Exp_k , where $k = 1, \dots, K$, denotes k^{th} experiment.

2.4.3 Performance Measures

Typically, performance can be calculated based on evaluation of models including a confusion matrix, and receiver operator characteristic (ROC) curves. Components in confusion matrix contain true and false classified examples (spikes and non-spikes) as shown in Table

Figure 2.9: Example of two classes ($\circ$ and $\bullet$). (a) Feature vectors of the classes are not well separated in low dimensional space. (b) Using transformation ϕ maps the feature vectors into higher dimensional space in order to be well-separated with linear hyperplane [67].

2.1, and they are described in terms of true positive (TP), true negative (TN), false positive (FP) and false negative (FN). Interpretation of those measurements are explained below:

- TP indicates that the proposed methods detect spikes correctly corresponding to the true spikes which are marked by an epileptologist.
- FP indicates that the methods detect spikes correctly, but they are incorrect and the epileptologist has not confirmed true spikes yet.
- TN indicates that both non-spikes detected by the methods and unmarked spikes are correct.
- FN indicates that the methods does not detect spikes, but they are true spikes.

Relationships between “sensitivity” and “1-specificity” is called receiver operating characteristic (ROC) curve and area under curve (AUC) which provides significant interpretation of accuracy. ROC curve is used to quantify how well the tests or diagnostic systems distinguish the testing sets or patients being diagnosed with unknown states such as normal or HYPS, into one of the groups. Chapters 3 and 4 are provided with the details of performance calculation and evaluation.

Table 2.1: A confusion matrix applied for spike detection algorithms

		Predicted spikes (detected)	
		Correct	Incorrect
True spikes (marked)	Correct	TP	FN
	Incorrect	FP	TN

Chapter 3

Methods

Figure 3.1: General scheme of the proposed methods.

This Chapter provides information on the proposed algorithms to improve spike detection in EEG in the presence of HYPS is based on Matching Pursuit time-frequency domain (MP-TFD) and non-negative matrix factorizations (NMFs) as shown in Figure 3.1.

3.1 Part A: Data Acquisition

Dataset and Preprocessing

Five EEG recordings from infants (4-9 months old) diagnosed with ISS and obtained from the Infantile Spasms Registry and Genetic Studies via a protocol approved by the University of Rochester's Research Subject Review Board were used to evaluate the proposed spike detection algorithm in the presence of HYPS. Each subject of awake EEG is length of time duration in five minutes. EEGs were recorded based on the international standard 10-20 system with sampling rates of 256 {2}, 500 {1}, and 512 {2} samples per second where {.}s indicates the number of individuals involved. The recording EEGs were imported to Persyst EEG software (Persyst, San Diego, CA) for artifact reduction and then were imported into MATLAB and bandpass filtered (0.5-30 Hz) for further analysis.

Construct Time-Frequency Domain

Time-frequency domain (TFD) analysis provides a two-dimensional energy representation of a signal in terms of its temporal and spectral content. The selection of an appropriate TFD method determines the resolution in both time and frequency necessary to successfully represent signal features such as the spike morphologies of interest in this instance. As seen in Figure 2.6, for example, STFT is unable to preserve the transient energy associated with these spikes due to the method's inherent time and frequency resolution trade-off. Therefore, in this thesis, a suitable tool for EEG analysis is a MP-TFD method, which is a high resolution TFD method that lends itself to the analysis of the non-stationary and chaotic EEG associated with HYPS.

The implementation of MP mainly consists of two stages. In stage one, it is used to decompose an input signal, $x(t)$, over an overcomplete dictionary of atoms according to Eq.(3.1):

$$x(t) = \sum_{i=0}^{I-1} \langle R_x^i, g_{\gamma_i} \rangle g_{\gamma_i}(t) + R_x^I, \quad (3.1)$$

where $\langle R_x^i, g_{\gamma_i} \rangle$ is the inner product of expansion coefficient on $g_{\gamma_i}(t)$, R_x^I is a residue signal after I iterations, and g_{γ_i} is a TF Gabor atom obtained using Eq.(3.2):

$$g_{\gamma_i} = \frac{1}{\sqrt{s_i}} g\left(\frac{t - p_i}{s_i}\right) e^{j(2(\pi f_i t + \phi_i))}, \quad (3.2)$$

where the notation γ_i represents the TF decomposition parameters (s_i, p_i, f_i, ϕ_i) denoted as scale factor, translation, frequency modulation and phase, respectively, at each iteration i , and g is the Gabor function:

$$g = 2^{(1/4)} e^{-\pi t^2}. \quad (3.3)$$

In Eq.(3.1), signal $x(t)$ is matching is attempted with all the possible atoms from a redundant dictionary, $D = \{g_{\gamma_1}, g_{\gamma_2}, \dots, g_{\gamma_r}\}$. At each iteration, the best correlated Gabor

Table 3.1: Proposed methods for this thesis.

Methods	NMFs					Feature extractions	Classification			
	Cost function	Solver	Parameters				Thresholdings			SVM
			Model order	Number of iteration	Tolerance		Th_{Aw}	Th_w	Th_h	
I	Euclidean Distance	Block principle pivoting (BPP)	Initial $K = 5$	$I = 200$	$tol = 10^{-6}$	MF(w_k)	-	6-10 Hz 10 Hz	0.015 0.01	-
II	Kullback-Leibler Divergence ($\beta = 1$)	Majorization - minimization (MM)	Initial $K = 30$	$I = 3500$	$tol = 10^{-6}$	MF(w_k), A(w_k)	12	8-19 Hz	0.3	-
III	Kullback-Leibler Divergence	Projected Gradient (PG)	Average optimal K ($K_{\text{eff}} = 10$)	$I = 3500$	$tol = 10^{-15}$	MF(w_k), A(w_k), A(h_k), B(h_k), C(h_k)	-	-	-	x

function is selected from the Gabor dictionary. If large enough numbers of iterations are used, then the residue term can be ignored as most of the coherent energy of the signal has been modeled by the Gabor atoms.

The second stage of constructing the result of the MP-TFD consists of summing the TFD of each decomposed Gabor atom as shown in Eq.(3.4):

$$\mathbf{MP} = \sum_{i=0}^{I-1} |\langle R_x^i, g_{\gamma i} \rangle|^2 W_{g_{\gamma i}}(t, f), \quad (3.4)$$

where $W_{g_{\gamma i}}(t, f)$ is the WVD [7], [14] of each Gabor atom $g_{\gamma i}(t)$, and $\mathbf{MP}$ is the matrix resulting from MP-TFD of $x(t)$.

3.2 Part B: Proposed Methods

While the previous part was concentrated on identifying an appropriate TFD exhibiting high resolution in both time and frequency domains of EEG signals, this section focuses on quantifying feature extraction of decomposed TF matrices using various NMFs (*i.e.*, NMF_1 , NMF_2 , and NMF_3) of proposed methods as shown in Table 3.1. In addition, feature extraction, classification, and performance are later provided. In Eq.(2.9), it measures quantification of similarities between original data ($\mathbf{MP}$ or $\mathbf{V}$) and approximation ($\mathbf{W}$, $\mathbf{H}$). The decomposed matrices can be obtained by minimizing a given cost function:

$$\min_{\mathbf{W}, \mathbf{H}} D(\mathbf{V}|\mathbf{WH}) \quad \text{subject to } \mathbf{W} \geq 0, \mathbf{H} \geq 0, \quad (3.5)$$

where $D(\mathbf{V}|\mathbf{WH})$ is a separable objective function:

$$D(\mathbf{V}|\mathbf{WH}) = \sum_{f=1}^F \sum_{n=1}^N d([v]_{fn} | [wh]_{fn}), \quad (3.6)$$

where $d(x|y)$ denotes a scalar objective function of approximation $y \in \mathbb{R}_+$ given original data $x \in \mathbb{R}_+$.

Various cost functions of NMF such as Euclidean Distance (ED) or Least squares error (LSE), Kullback-Leibler divergence (KLD), and Itakura-Saito (IS) can be formulated in term β -divergence (β -NMF), which parameterizes one variable (β):

$$d_\beta(x|y) = \begin{cases} \frac{x^\beta}{\beta(\beta-1)} + \frac{y^\beta}{\beta} - \frac{xy^{\beta-1}}{\beta-1}, & \beta \in \{0, 1\} \text{ (generalized)}, \\ \frac{1}{2}(x-y)^2, & \beta = 2 \text{ (ED)}, \\ x \log \frac{x}{y} - x + y, & \beta = 1 \text{ (KLD)}, \\ \frac{x}{y} - \log \frac{x}{y} - 1, & \beta = 0 \text{ (IS)}. \end{cases} \quad (3.7)$$

Choice of β parameter depends on applications. For example, $\beta = 1$ is often applied to sound source separation [65], $\beta = 0$ is employed to artifact rejection of single channel EEG [15] and music analysis [17], and $\beta = 2$ is used for facial images [36].

The NMF algorithm begins randomly initial non-negative values of $\mathbf{W}$ and $\mathbf{H}$, and then performs to find iteratively optimal factors obtained by minimizing a given cost function using gradient descent with respect to $\mathbf{W}$ and $\mathbf{H}$. For example, a simple minimization is multiplicative update, proposed by Lee and Seung [39], based on ED and KLD as follows:

$$\text{For ED,} \quad \mathbf{W} \leftarrow \mathbf{W} \cdot \frac{\mathbf{V}\mathbf{H}^T}{\mathbf{W}(\mathbf{H}\mathbf{H}^T) + \epsilon}, \quad \mathbf{H} \leftarrow \mathbf{H} \cdot \frac{\mathbf{W}^T\mathbf{V}}{(\mathbf{W}^T\mathbf{W})\mathbf{H} + \epsilon}. \quad (3.8a)$$

$$\text{For KLD,} \quad \mathbf{W} \leftarrow \mathbf{W} \cdot \frac{(\frac{\mathbf{V}}{\mathbf{W}\mathbf{H}})\mathbf{H}^T}{\mathbf{1} \cdot \mathbf{H}}, \quad \mathbf{H} \leftarrow \mathbf{H} \cdot \frac{\mathbf{W}^T(\frac{\mathbf{V}}{\mathbf{W}\mathbf{H}})}{\mathbf{W} \cdot \mathbf{1}}. \quad (3.8b)$$

Above Eq.(3.8), $(A \cdot B)$ and $(\frac{A}{B})$ are denoted element-wise of multiplication and division, respectively. $\mathbf{1}$ denote a matrix of ones, and adding ϵ (a small constant, 10^{-9}) avoids division by zero.

3.2.1 Method I

The proposed algorithm depicted in Figure 3.2 aims first to identify and quantify the variations of spikes of HYPS in TF representation, second, to extract features associated with spikes from decomposed vectors using NMF based on the paper [37], and last, to evaluate performance of the proposed method.

NMF₁ : Euclidean Distance Solved by Block Principle Pivoting

As seen in Eq.(3.8a), the algorithm is simply to implement and iteratively update $\mathbf{W}$ while fixing $\mathbf{H}$, and then update $\mathbf{H}$ while fixing $\mathbf{W}$. However, it does not guarantee convergence to local minimum (non convex optimization or NP-hard problem) [22]. Kim *et al.* [37] divided Euclidean Distance cost function into two exclusive subproblems which are convex in $\mathbf{W}$ or $\mathbf{H}$, but cannot be convex for both. This technique called alternating nonnegative least squares (ANLS) is summarized in Algorithm 1.

- If fixed $\mathbf{H}$ and updated $\mathbf{W}$,

$$\min_{\mathbf{W} \geq 0} D(\mathbf{V}|\mathbf{WH}) = \min_{\mathbf{W} \geq 0} \|\mathbf{H}^T \mathbf{W}^T - \mathbf{V}^T\|_F^2. \quad (3.9)$$

- If fixed $\mathbf{W}$ and updated $\mathbf{H}$,

$$\min_{\mathbf{H} \geq 0} D(\mathbf{V}|\mathbf{WH}) = \min_{\mathbf{H} \geq 0} \|\mathbf{WH} - \mathbf{V}\|_F^2. \quad (3.10)$$

The subproblems are solved for $\mathbf{W}$ and $\mathbf{H}$ by block principle pivoting (BPP).

Algorithm 1 Alternating Nonnegative Least Squares (ANLS) based on BPP [37]

- 1: **Inputs:** Matrix $\mathbf{V}_{F \times N}$, tolerance $0 < tol \ll 1$, maximum iteration I , number of factor K
 - 2: **Initialize** $\mathbf{W}_{F \times K}$
 - 3: **repeat**
 - 4: Minimize the subproblem in Eq.(3.9) to obtain $\mathbf{H}$ with fixed $\mathbf{W}$
 - 5: Minimize the subproblem in Eq.(3.10) to obtain $\mathbf{W}$ with fixed $\mathbf{H}$
 - 6: **until** A stopping (convergence) criterion is satisfied or number of iteration meets I .
 - 7: **Outputs:** $\mathbf{W}, \mathbf{H}$
-

Feature Extraction

• Mean Frequency

After the TF spectral and temporal vectors, (w_k, h_k) , are calculated in the Eqs.(3.9)-(3.10), the TF spectral vectors that are associated with the spectral structure of the epileptic spikes in HYPS are identified. To find the mean frequency (MF) of the spectral vectors, the following calculation, found in Eq.(3.11), is performed:

$$\text{MF}(w_k) = \frac{\sum_{f=1}^F f w_k(f)}{\sum_{f=1}^F w_k(f)} f_s, \quad (3.11)$$

Figure 3.2: Block diagram of **Method I**.

where $\text{MF}(w_k)$ is the mean frequency value of w_k , $k \in [1, 2, \dots, K]$, F is the number of frequency samples in vector w_k , and f_s is the sampling frequency of the EEG recording. The MF values provide a measure of the frequency content of the decomposed spectral vectors. Since HYPS spikes are generally associated with high frequency energy, the spectral vector with the highest MF value greater than a given threshold, Th_w , is selected as the spectral vector (w_k^*) corresponding to the spikes as shown in Figure 3.2.

- **Spike Localization**

If any w_k^* are not found for a given recording, it indicates that there is not a spectral vector with a MF value greater than the threshold (Th_w) and that there are not spikes in the given recording $x(t)$. However, if a w_k^* for a given EEG segment is found, then the index, k , of w_k^* is used to identify the location of the corresponding spikes. As shown in the Eqs.(3.9)-(3.10), each decomposed spectral vector, w_k , is associated with a temporal vector, h_k , that indicates the temporal locations where the spectral energy of w_k exists in the MP-TFD matrix of the EEG recording of $x(t)$. Once the spectral vector w_k^* that is associated with the spikes in EEG is identified, the corresponding temporal vector h_k^* represents to locate the spikes in $x(t)$.

Classification and Performance Evaluation

The peak locations of h_k^* that are greater than threshold Th_h are identified, and annotated on the EEG recording as possible location of spikes in order to compare true spikes marked by the epileptologists. To evaluate the performance, percentages of correct spikes (true positive: TP), non-spikes (false positive: FP), and missed spikes (false negative: FN) can be calculated:

$$\text{Correct spikes (\%}TP\text{)} = \frac{TP}{TP + FN}, \quad (3.12)$$

$$\text{Incorrect spikes (\%}FP\text{)} = \frac{FP}{TP + FP}, \quad (3.13)$$

$$\text{Missed spikes (\%}FN\text{)} = \frac{FN}{TP + FP}. \quad (3.14)$$

Chapter 4 explains how to calculate the performance.

3.2.2 Method II

While **Method I** selected arbitrary K , **Method II** aims to determine an effective number of factor (or K_{eff}) for TF matrix decomposition using NMF based on automatic relevance determination (ARD) [59].

NMF₂ : β – Divergence Solved by Majorization-Minimization

Determining K is very important for learning data to avoid overfitting and underfitting. For example, if K is too large, the data learning is overfitting. If K is too small, the data learning is underfitting. Vincent *et al.* [59] apply Bayesian NMF model to find K_{eff} in order to obtain better decomposition. They also explain concepts how to determine K_{eff} .

First of all, Vincent *et al.* [59] describe a probabilistic model for ARD, which relates to the index k^{th} of row vector of $\mathbf{W}$ and column vector of $\mathbf{H}$ that ties together through a common scale parameter or relevance weight (λ_k). The parameter λ_k controls those vectors. If λ_k is small or lower bound (τ), those vectors are driven to zero or decoupled, leading to a more sparse model. The authors apply $\mathbf{W}$ and $\mathbf{H}$ to exponential and half-normal priors, denoted $\mathcal{E}$ and $\mathcal{HN}$ for l_1 –ARD and l_2 –ARD, respectively as showed in Table 3.2. λ_k can be obtained from inverse-Gamma priors:

$$p(\lambda_k|a, b) = \frac{b^a}{\Gamma(a)} \lambda_k^{-(a+1)} \exp(-\frac{b}{\lambda_k}), \quad (3.15)$$

where a and b denote shape and scale hyperparameters and are fixed for all k . To set these parameters for this thesis, given a constant a is to obtain b using method of moments [58] as shown in Table 3.2.

Second, the following cost function is derived from Eq.(3.15) and Table 3.2:

$$\begin{aligned} C(\mathbf{W}, \mathbf{H}, \boldsymbol{\lambda}) &= -\log p(\mathbf{W}, \mathbf{H}, \boldsymbol{\lambda}|\mathbf{V}) \\ &= \frac{1}{\phi} D_{\beta}(\mathbf{V}|\mathbf{WH}) + c \sum_{k=1}^K \frac{1}{\lambda_k} (f(w_k) + f(h_k) + b) + \\ &\quad c \log \lambda_k + \text{const}, \end{aligned} \quad (3.16)$$

where $f(w_k)$ and $f(h_k)$ are model functions such as half-normal and exponential models: for l_1 –ARD, $f(u) = \|u\|_1$, and for l_2 –ARD, $f(u) = \frac{1}{2}\|u\|_2^2$, where $u = \{w_k, h_k\}$. ϕ is dispersion parameter, and c is defined in Table 3.2, and $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \dots, \lambda_K)$. This thesis

Table 3.2: Two probabilistic models based on automatic determination relevance for β –NMF [59]. μ_V denotes mean of matrix $\mathbf{V}$.

	Models	
	Exponential (l_1 –ARD)	Half normal (l_2 –ARD)
Temporal vector	$p(h_{kn} \lambda_k) = \mathcal{E}(h_{kn} \lambda_k)$	$p(h_{kn} \lambda_k) = \mathcal{HN}(h_{kn} \lambda_k)$
Spectral vector	$p(w_{fk} \lambda_k) = \mathcal{E}(w_{fk} \lambda_k)$	$p(w_{fk} \lambda_k) = \mathcal{HN}(w_{fk} \lambda_k)$
where, given $u = \{w, h\}$,	$u \geq 0, \mathcal{E}(u \lambda) = \frac{1}{\lambda} \exp(-\frac{u}{\lambda})$ $u < 0, \mathcal{E}(u \lambda) = 0$	$u \geq 0, \mathcal{HN}(u \lambda) = \sqrt{\frac{2}{\pi\lambda}}$ $u < 0, \mathcal{HN}(u \lambda) = 0$
Parameter b for Eq.(3.15)	$b = \sqrt{\frac{(a-1)(a-2)\mu_V}{K}}$	$b = \frac{\pi(a-1)\mu_V}{2K}$
Parameter c for Eqs.(3.16), (3.17)	$c = F + N + a + 1$	$c = \frac{(F+N)}{2} + a + 1$

Algorithm 2 l_1 -Automatic Relevant Determination for β -NMF [59]

- 1: **Inputs:** Matrix $\mathbf{V}_{F \times N}$, tolerance $0 < tol << 1$, maximum iteration I , number of factor K
 - 2: **Initialize:** $\mathbf{W}_{F \times K}$, $\mathbf{H}_{K \times N}$
 - 3: **Calculate:** $c = F + N + a + 1$ and $\gamma(\beta)$ in Eq.(3.18)
 - 4: **Calculate:** $b = \sqrt{\frac{(a-1)(a-2)\mu_v}{K}}$
 - 5: **while** ($\tau < tol$)
 - 6: $\mathbf{H} = \mathbf{H}(\frac{\mathbf{W}^T[(\mathbf{WH})^{(\beta-2)}\mathbf{V}]}{\mathbf{W}^T(\mathbf{WH})^{(\beta-1)} + \phi/repmat(\lambda, 1, N)})^{\gamma(\beta)}$
 - 7: $\mathbf{W} = \mathbf{W}(\frac{[(\mathbf{WH})^{(\beta-2)}\mathbf{V}]\mathbf{H}^T}{(\mathbf{WH})^{(\beta-1)}\mathbf{H}^T + \phi/repmat(\lambda, F, 1)})^{\gamma(\beta)}$
 - 8: $\lambda_k = \frac{\sum_{f=1}^F w_{fk} + \sum_{n=1}^N h_{kn} + b}{c}$ for all k
 - 9: $\tau = \max_{k=1, \dots, K} |(\lambda_k - \hat{\lambda}_k)/\hat{\lambda}_k|$
 - 10: **end while**
 - 11: **Calculate:** K_{eff} in Eq.(3.19)
 - 12: **Outputs:** $\mathbf{W}$, $\mathbf{H}$, λ , and K_{eff} (model order)
-

sets $\phi = 1$ according to the work done by Vincent *et al.* [59] and applies l_1 -ARD model.

As seen in the above equation, observe that the first term represents data fitting, the second and third terms are regularization having a common parameter, λ_k . If λ_k is large, the second term is suppressed while the other term is increased. This inverse proportion aims to prune irrelevant components out of the model, causing it to become sparse. Eq.(3.16) can be reduced parameter λ by optimizing that cost function with respect to λ and keeping λ to control the couples of columns $\mathbf{W}$ and rows $\mathbf{H}$:

$$C(\mathbf{W}, \mathbf{H}) = \frac{1}{\phi} D_\beta(\mathbf{V}|\mathbf{WH}) + c \sum_{k=1}^K \log(f(w_k) + f(h_k) + b) + \text{const}, \quad (3.17)$$

where $\text{const} = Kc(1 - \log c)$.

Finally, the cost function in Eq.(3.17) is solved by majorization-minimization (MM) problem which guarantees convergence at every iteration [26]. Overall, algorithm for l_1 -ARD is summarized in Algorithm 2.

In the loop ‘**while**’ in that Algorithm, iteratively updating the matrices $\mathbf{W}$ and $\mathbf{H}$ employs multiplicative update based on MM. Parameter $\gamma(\beta)$ is given:

$$\gamma(\beta) = \begin{cases} \frac{1}{2-\beta} & , \beta < 1 \\ 1 & , 1 \leq \beta \leq 2 \\ \frac{1}{\beta-1} & , \beta > 2. \end{cases} \quad (3.18)$$

λ_k is derived by partial derivative of Eq.(3.16), and tolerance (tol) is designed for termination if $(\frac{\lambda_k - \hat{\lambda}_k}{\hat{\lambda}_k})$, where λ_k and $\hat{\lambda}_k$ are current and previous updates, falls over a given

threshold, τ . After convergence, K_{eff} can be calculated:

$$K_{\text{eff}} = |\{k \in \{1, \dots, K\} : \frac{\lambda_k - \frac{b}{c}}{\frac{b}{c}} > \tau\}|. \quad (3.19)$$

Feature Extraction

After obtaining the decomposed matrices, $\text{MF}(w_k)$ in the Eq.(3.11) is calculated, decomposed vectors having all values of zero across the vectors are removed due to irrelevance, and the remains are reordered from maximum to minimum MFs. In **Method I**, features of spikes were investigated by only MFs. This method considers features with area under the curve of spectral vector $A(w_k)$. To calculate $A(w_k)$ using Matlab's command, `trapz`, the vector has to be normalized in order to keep the feature because of the magnitude-invariance among the entire data. Figure 3.3 (a) demonstrates this method while Figures 3.3 (b)-(d) are explained later in **Method III**.

Classification and Performance

Once the spectral vector having $\text{MF}(w_k) > Th_w$ and $A(w_k) > Th_{Aw}$ is found, finding spike locations in the corresponding temporal vector is identified and annotated on the EEG to compare the true spike marked by the epileptologists if peaks of energy values of that vector are greater than Th_h .

While **Method I** was considered only one of temporal and spectral vector (w_k^*, h_k^*) , **Method II** employs one or two coupled vectors. The feature extraction algorithm is brief:

1. Find spectral vectors having $\text{MF}(w_k)$, $A(w_k)$ in the conditions of given thresholds (Th_w, Th_{Aw}) .

$$\text{MF}(w_k) > Th_w, A(w_k) > Th_{Aw}, \quad (3.20)$$

where $k = \{A, B, \dots\}$. For example, if (w_A, w_B) are found within the conditions, (h_A, h_B) are possible to indicate spikes.

2. Find common time indexes of temporal vectors:

$$Idx_{\text{com}} = Idx_A \cap Idx_B, \quad (3.21)$$

where $Idx_A, Idx_B, Idx_{\text{com}}$ are time indexes of temporal vectors A, B and common time index with operation of intersection ($\cap$), respectively.

3. Add temporal vectors with corresponding to the common time index, and find peak locations where they are greater than Th_h to locate spikes:

$$h_{\text{com}}(Idx_{\text{com}}) = h_A + h_B > Th_h. \quad (3.22)$$

Performance evaluation is similar to **Method II**.

Figure 3.3: The illustration of feature extraction for **Methods II** and **III** shows how to calculate and tabulate the features. (a) Features of spectral vector (w_k) contains area (A_w) and mean frequency (MF_w) values. To calculate area of w_k , the vector has to be normalized in order to keep the feature because of magnitude-invariance among the entire data. (b) The vector with two distributions has multiple features denoted as A_d , B_d , C_d , where d is a number of distribution (in this example, $d = 2$). The thick-dashed vertical line relied on the first distribution and associated with spike marked by the epileptologist is labeled as class “1”. Otherwise, class “0” indicates non-spike. (c) Table contains features of spectral and temporal vectors, including class labels.

3.2.3 Method III

After finding the appropriate number of factor (K_{eff}) from **Method II**, the average K_{eff} is employed to NMF based on the paper [40]. This method mainly focuses on classification between spikes and non-spikes using classifier, support vector machine (SVM), and K-fold cross validations in order to evaluate performance.

NMF₃ : Kullback-Leibler Divergence Solved by Projected Gradient

The algorithm in **Method II** was slow convergence and exhibited bad reconstruction error. The result will be explained in Chapter 4. However, NMF algorithm using projected gradient (PG) proposed by Lin [40] robusts to stationary point, and also shows better reconstruction between $\mathbf{V}$ and $\mathbf{WH}$. The algorithm is summarized as follows:

- If fixed $\mathbf{H}$ and updated $\mathbf{W}$,

$$\mathbf{W}^{i+1} = \max\{(0, \mathbf{W}^i - \alpha^i \nabla f(\mathbf{W}^i))\}, \quad (3.23)$$

where i is index of iteration: $\{i = 1, 2, \dots, I\}$, $f(\mathbf{W})$ denotes differentiable function with respect to $\mathbf{W}$, $f = \mathbf{V} - \mathbf{WH}$, every iteration, and α is a step size. At the first iteration, initial step size value of the paper [40] is 1. Then, updated α per iteration can be checked in Eq.(3.24):

$$f(\mathbf{W}^{i+1}) - f(\mathbf{W}^i) \leq \sigma \nabla [f(\mathbf{W}^i)]^T (\mathbf{W}^{i+1} - \mathbf{W}^i), \quad (3.24)$$

where σ is a constant ($\sigma = 0.01$). In Eq.(3.23), if updated $\mathbf{W}$ with corresponding to α satisfies with Eq.(3.24), then α is repeatedly decreased by multiplying $\theta = 0.1$ until Eq.(3.24) is unsatisfied. Otherwise, if updated $\mathbf{W}$ does not satisfy with Eq.(3.24), then α is repeatedly increased by dividing θ until Eq.(3.24) is satisfied, see Algorithm 3.

When α is found, the updated $\mathbf{W}$ is closed to a stationary point with condition of termination (or stopping criterion):

$$\|\nabla^P f(\mathbf{W}^i)\| \leq \epsilon \|\nabla f(\mathbf{W}^1)\|, \quad (3.25)$$

where ϵ is a small constant or tol. ∇^P is defined:

$$\nabla^P f(\mathbf{W}) = \begin{cases} \nabla f(\mathbf{W}) & , w_{fk} > 0, \\ \min[0, \nabla f(\mathbf{W})] & , w_{fk} = 0, \end{cases} \quad (3.26)$$

where $f = [1, 2, \dots, F]$, $k = [1, 2, \dots, K]$, and $w_{fk} \in \mathbf{W}$.

- If fixed $\mathbf{W}$ and updated $\mathbf{H}$: variable $\mathbf{W}$ is replaced to $\mathbf{H}$ in Eqs.(3.23)-(3.26).

Above description of the algorithm is summarized in Algorithm 3.

Algorithm 3 Projected Gradient algorithm for NMF [40]

- 1: **Inputs:** Matrix $\mathbf{V}_{F \times N}$, tolerance $0 < tol << 1$, maximum iteration I , number of factor K
 - 2: **Initialize:** $\mathbf{W}_{F \times K}$, $\mathbf{H}_{K \times N}$, given $0 < \sigma < 1$, $0 < \theta < 1$: This thesis defines $(\sigma, \theta) = (0.01, 0.1)$ and $\sigma^1 = 1$ according to [40].
 - 3: **repeat** $i = 1, \dots, I$
 - 4: Update $\alpha^i \leftarrow \alpha^{i-1}$
 - 5: **if** α^i satisfies Eq.(3.24),
 - 6: then (increase repeatedly the value) $\alpha^i \leftarrow \frac{\alpha^i}{\theta}$ until it does not satisfy Eq.(3.24)
 - 7: **else** (decrease repeatedly the value) $\alpha^i \leftarrow \alpha^i \theta$ until it satisfies Eq.(3.24)
 - 8: **end if**
 - 9: update $\mathbf{W}$ in Eq.(3.23)
 - 10: update $\mathbf{H}$ in Eq.(3.23) by changing $\mathbf{W}$ to $\mathbf{H}$
 - 11: **until** A stopping criterion is satisfied in Eq.(3.25) for $\mathbf{W}$ and $\mathbf{H}$ or number of iteration meets I
 - 12: **Outputs:** $\mathbf{W}$, $\mathbf{H}$
-

Feature Extraction

Figures 3.3(a), (b) demonstrate how to calculate mean frequency value, area of each vector w_k , and measure width, height, and area of each distribution of vector h_k . One spectral vector can have several numbers of distribution in temporal vector of k^{th} factor. Example of Figure 3.3(b) shows two distributions where one of them indicates true spike marked by the epileptologists while the other indicates non-spike. All features are calculated for feeding to the table (see Figure 3.3(c)).

Classification and Performance

As seen in the table in Figure 3.3(c), several features associated with classes (spikes and non-spikes) are conveyed to the learner, SVM.

Support Vector Machine

In general, SVMs integrate two main concepts. The first concept determines maximization of an optimum hyperplane, which measures distance from it to the closest data point on each side. The other concept applies a kernel function, which operates the dot product of two training vectors, to map the feature vectors from lower to higher dimensional spaces.

SVM constructs optimal hyperplane for linearly separable and non-separable patterns. In separable patterns case, given training samples (x_i, y_i) , where $i = 1, 2, \dots, N$ samples of data, x_i is input vectors of i^{th} sample corresponding to labels y_i : binary outputs $\{-1, 1\}$, following equation of a decision boundary in the form of a hyperplane is:

$$h(x_i, w, b) = w^T x_i + b, \quad (3.27)$$

where w is an adjustable weight vector, b is a bias, and $h(\cdot)$ is hyperplane function.

Hyperplane function indications with unknown class of vector x_i would belong to a class of y_i . For example, suppose value of $h(x_i, w, b)$ is greater than or equal to zero, vector of x_i belongs to the class, $y_i = 1$. Otherwise, it belongs to the other class, $y_i = -1$. In order to find an optimal hyperplane, a quadratic optimization can solve the problem by minimizing $\|w\|$ from $\rho = \frac{2}{\|w\|}$, where $\|w\|$ is the distance from origin to perpendicular to the line of hyperplane, and ρ is distance between positive and negative examples. Therefore, the norm of w has to be minimized for maximizing the margin of ρ .

In the case of nonlinearly separable patterns, which are more complicated than the previous case, a linear classifier may not be suitable. It cannot be constructed by separating a linear hyperplane without encountering classification errors. An optimal hyperplane can be found by minimizing the probability of classification error, averaged over training samples. Using the kernel function can transform non-separable feature vectors in low dimensional spaces to separable feature vectors in higher dimensional spaces. The kernel function operates as inner dot product of vector: $K(x_i, x_j) = \phi(x_i)^T \phi(x_j)$.

Given x_i, x_j are feature vectors of i^{th} and j^{th} samples, popular kernels are categorized as follows:

- linear: $K(x_i, x_j) = x_i^T x_j$.
- polynomial: $K(x_i, x_j) = (\gamma x_i^T x_j + r)^d$, where $\gamma > 0$ and d is specified by user.
- radial basis function (RBF): $K(x_i, x_j) = \exp(-\frac{\|x_i - x_j\|_2^2}{2\sigma})$, where σ controls width.

This thesis uses SVM based on RBF kernel to classify the characteristics of spikes and non-spikes. In RBF kernel, parameter σ controls width of how far the influence of a single training sample reaches: a small value of σ indicates the sample is far and a large value of σ indicates the sample is close. Parameter C controls the relative importance of:

$$\min_{w, \epsilon, C} \frac{1}{2} \|w\|_2^2 + C \sum_i \epsilon_i, \quad (3.28)$$

where ϵ_i is slack variable and C is a box constraint, which controls the trade-off between achieving a low error on the training dataset and minimizing the norm of the weight. However, Sathiya *et al.* [33] and Chang *et al.* [11] mentioned that changing parameter C does not affect the error on test set when using cross-validations. Therefore, the parameter C is not analyzed for this thesis. Chapter 4 provides results of performance by changing parameter σ using grid search.

Cross Validations

The data is partitioned into training and testing sets based on K-fold cross validations. Using this technique divides the data of all patients into 80% for training and 20% for testing of each label as illustrated in Figure 3.4. SVM learns the training sets in order to find parameters of non-linear hyperplane. Then, it predicts labels being tested using known parameters, or statistic variables. Performance of SVM can be evaluated by plotting

relationships between “sensitivity” and “1-specificity” or called receiver operating characteristic (ROC) curve and calculating its area under curve (AUC), which provides significant interpretation of accuracy. ROC curve is used to quantify how well the tests or diagnostic systems distinguish the testing sets or patients being diagnosed with unknown states such as normal or HYPS, into one of the groups. Figure 3.5 shows characteristics of ROC curve.

Spikes		Non-Spikes	
80% Train	20% Test	80% Train	20% Test

Figure 3.4: Illustration of classification models as a preprocessing step of classifier. Each of spikes and non-spikes having multiple features are split into two subsets for training (80%) and testing (20%) using K-Fold model.

Figure 3.5: ROC curve represents different levels of performance indicated by direction of the arrow (from the worst to better performance).

Chapter 4

Results and Discussion

This Chapter provides results and explains discussion of the three proposed methods.

Preprocessing

For each individual, two EEG sites, P_4 and O_2 , were selected to apply the algorithms. First, each five minute EEG recording was divided into 30 10-second EEG segments, $x(t)$, for computational efficiency.

4.1 Method I

Figure 4.1 illustrates the step-by-step process of the proposed algorithm for a given EEG segment. Figure 4.1(a) shows a 10-second duration of an EEG recording from 60 to 70 seconds. The MP-TFD of the signal is shown in Figure 4.1(b). Analysis showed that $I = 100$ iterations in the matching-pursuit decomposition stage was sufficient to successfully model the EEG signal in the TF domain. The resolution of the MP-TFD is selected to be 0.15 Hz in frequency and $1/f_s$ in time. Since there is no spike related information beyond 30 Hz, the frequency domain was limited to that value. The F and N values for each MP-TFD are 200 and $10 \times 60 \times f_s$. NMF is applied to the MP-TFD (*i.e.*, **MP**) to obtain the spectral and temporal decompositions, **W** in Figure 4.1(c) and **H** in Figure 4.1(e), respectively. The number of NMF decomposition factor (*i.e.*, K) was selected to be 5 based on empirical analysis. The MF values were calculated for each vectors in **W** as shown in the table of Figure 4.1.

Th_h was selected to be 0.015 for the threshold if the MF of w_k^* was between 6 Hz and 10 Hz (*i.e.*, Th_w) and 0.010 if MF was greater than 10 Hz. These threshold values were selected empirically based on the visual inspection of the epileptology experts. In the example given in Figure 4.1, it can be observed that the vector w_3 shows its energy spreads over relatively high frequencies (*i.e.*, 5-20 Hz). When the MF values were calculated, it was found that the third spectral vector presented the highest MF with the value of 10.12 Hz (*i.e.*, w_3^*); therefore, the third temporal vector in **H** (*i.e.*, h_3^*) was used to locate spikes. Vector h_3^* is demonstrated in Figure 4.1(d) where the asterisks indicate the peak values greater than Th_h and the corresponding time locations shown by the dashed lines locate the

Figure 4.1: The step by step illustration of the first proposed algorithm. (a) 10-second EEG recording. (b) The MP-TFD of the EEG recording. (c) and (e) The decomposed spectral and temporal matrices, respectively. The table shows the MF value of each spectral vector and w_3^* which has the highest MF value. (d) The temporal vector corresponding to h_3^* represented by the dashed lines indicate the spike locations.

Figure 4.2: The top displays EEG signal ($x(t)$), the bottom represents h_k^* . Spike locations detected by the h_k^* peaks that are greater than threshold Th_h . The dashed lines indicate the proposed locations of the spikes and arrows on top of the $x(t)$ indicate the spikes, which were marked by the epileptologist.

spikes in the EEG recording. In this case, all the detected spikes matched the epileptiform discharges that were marked by the epileptologist.

The developed algorithm was performed on the five-minute recordings of all five individuals, and the located spikes were compared with the ones that were marked by the epileptologist. The total number of the located spikes that matched the marks is reported as *True Positive: (TP)* in Table 4.1, the total number of spikes that were missed by the algorithm is reported as *False Negative: (FN)*, and the number of spikes that were not confirmed by the epileptologist is reported as *False Positive: (FP)*. For example, in Figure 4.1, *TP* is five, *FP* is zero, and there are no missed spikes (*i.e.*, $FN = 0$). Another example is shown in Figure 4.2 where the algorithm located seven spikes (see the dashed lines) and the epileptologist marked six spikes (the arrows on top of Figure 4.2). There are five instances where the dashed lines and arrows line up, indicating that the algorithm and the interpretation of epileptologist correspond and *TP* is five. At time 2.5 second, a discharge is missed (*i.e.*, $FN = 1$), and the algorithm located two non-spikes at times 3.95 and 9.5 second; therefore, *FP* is two. The analysis of the five individuals is contained in Table 4.1.

The analysis of the results obtained with the algorithm with the spike detection results obtained with the Persyst EEG software, an EEG monitoring software that is used

Table 4.1: Results of detection of spike locations are calculated in Eqs.(3.12)-(3.14) in Chapter 3. True Positive (TP). False Negative (FN). False Positive (FP). The number in each parenthesis indicates the number of spikes.

Patient	Proposed Algorithm			Persyst		
	TP	FN	FP	TP	FN	FP
1 (36)	86% (31)	14% (5)	45% (25)	0% (0)	100% (36)	-
2 (66)	100% (66)	0% (0)	55% (82)	5% (3)	95% (63)	55% (6)
3 (30)	70% (21)	30% (9)	42% (15)	17% (5)	83% (25)	-
4 (54)	96% (52)	4% (2)	68% (110)	0% (0)	100% (54)	100% (1)
5 (47)	79% (37)	21% (10)	57% (50)	0% (0)	100% (47)	-
Average	86%	14%	53%	4%	96%	-

by epileptologists for detection of spikes and seizures in Epilepsy Monitoring Units. The results of the two methods are displayed in Table 4.1, where the number in the parentheses shows the number of spikes under each category.

From the results in Table 4.1, it can be seen that on average 86% of the spikes are successfully detected using the developed algorithm, while the Persyst software detected only 4% of the spikes, leaving 96% of the spikes undetected. Persyst employs the mimetic method, which is a temporal-based method to locate spikes and the accuracy of the method depends on the pre-specified spikes' parameters such as amplitude and duration. Since the HYPs spikes are highly chaotic and not well-characterized, the method has a significant limitation for spike identification. The proposed method considers both time and frequency structures to locate spike waveforms. The MP-TFD that used in this analysis provides a high TF resolution that can significantly indicate the presence of spike activity and NMF clusters, the spectral energy of the spikes in the spectral vector w_k^* and the temporal location h_k^* ; thereby, enabling the proposed algorithm to locate a significantly high percentage of the spikes. The relatively high FP of the algorithm (*i.e.*, 53%) can be improved by using multiple montages of EEG; for example, epileptologists normally rely on both Referential and Longitudinal Bipolar montages for detection of epileptiform discharges while this thesis only used the Longitudinal Bipolar montage. By extending the algorithm to incorporate signals from multiple montages of EEG, improvement in performance is expected in the future work.

4.2 Method II

This method was applied to normalized **MP** or **V** in Figure 4.3(a), which contains small mean values of the data ($0 \leq \mu_v \ll 1$). This provided an unexpected result of reconstruction error as shown in Figure 4.3(b). If the data has small μ_v , this case can suppress the parameters b and λ_k ; as a consequence, the small λ_k obliterated the couples (w_k, h_k) (see Algorithm 2). To alleviate the problem for further experiment, MPs are multiplied by 1000 and the result showed improved reconstruction error as shown in Figure 4.3(c).

This method aims to estimate the number of factor (K_{eff}) with the five patients. Each ten second per patient has a number of spikes (from zero to six spikes according to the epileptologist's marks). Parameter, a , based on l_1 -ARD was set $\{0.5, 100, 500, 1000\}$, and K was initialized 30. The algorithm was generated with three initializations of **W** and **H**, and plotted the results for all patients as depicted in Figure 4.4. The number of iteration was 3500, $\text{tol} = 10^{-6}$, and $\beta = 1$.

From the results in Figure 4.4, when a increases, K_{eff} increases. However, computation of a large a is extensive than small a . Average K_{eff} is 9.24 or 10 for **Method III**.

In Figure 4.5, $a = 0.5$ was selected, and the thresholds were set $Th_h = 0.3$, $Th_w = 8-19$ Hz, and $Th_{Aw} = 12$. Figure 4.5(a) shows original **MP** with dashed lines indicated spikes marked by the epileptologists. The dark color represents high energy while the white color represents low energy. NMF decomposes **MP** into **W**, **H** as displayed in Figures 4.5(b), (c), respectively. Notice that $K_{\text{eff}} = 10$ decreases from initial $K = 30$. The irrelevant decomposed vectors having zero energy values were rejected, the remaining vectors were sorted by decrement $\text{MF}(w_k)$ values, and temporal vectors were reordered associated with the index of w .

In Figure 4.5(d), the table contains index k , $\text{MF}(w_k)$, and area of spectral vectors $A(w_k)$. The first-second rows are in the conditions of thresholds Th_w , and Th_{Aw} . Otherwise, the remaining rows are non-spikes. For example, w_4 and w_9 spread over high frequency with large area. Therefore, h_4 and h_9 are possibly indicated as spikes. Once having those indexes, observe that, at times 184.2, 185.4, 187, 188, and 189 seconds, h_4 and h_9 having high energy values are matched to the dashed lines. However, energy of h_4 at time 180 seconds is smaller than the energy of h_9 , as redisplayed in Figures 4.7(a), (b), respectively. Those feature vectors were employed to the feature algorithm in Eqs.(3.20)-(3.22) to find spikes shown in Figure 4.7(c), which are six correct spikes and three non-spikes. Observe that there are no FP of (c) at time between 180-183 seconds when compared between (a), (b). Using the algorithm of the feature extraction can pull small energy values to high values by combining two vectors.

Another example is depicted in Figure 4.8. Consider the table in Figure 4.8(c) that the second-third rows are in the conditions while the others are irrelevant. Having a mean frequency that is too high such as w_{21} exhibits a missed spike in h_{21} at the time 239.5 seconds. By inspecting vectors $(h_4$ and $h_{17})$ sanctimoniously, this point may indicate a spike. Therefore, Th_w was appropriately set 8-19 Hz, but may not be optimal. In Figure 4.8(c), spike locations were found using the feature extraction algorithm in Eqs.(3.20)-(3.22), there are three correct spikes, three non-spikes at times 230.2, 230.8 and 237.2

Figure 4.3: Differences between original MP and approximation of decomposed factors. (a) Normalized MP is decomposed and each element values are very small as shown in color bar. (b) The results of using β -NMF based on l_1 -ARD provides ill-reconstruction error of $\mathbf{W}$, $\mathbf{H}$ when initial $K = 30$ and the resulting $K_{\text{eff}} = 1$. For further analysis of **Method II**, all MPs are multiplied by 1000, and (c) the results show better reconstruction error when using initial $K = 30$ and resulting $K_{\text{eff}} = 10$. The parameters were set $\{a, \text{tol}, I\} = \{0.5, 10^{-6}, 3500\}$. (d) using NMF based on projected gradient (PG), decomposition of normalized MP shows improved reconstruction error for **Method III**. The parameters were set $\{K_{\text{eff}}, \text{tol}, I\} = \{10, 10^{-15}, 3500\}$

Table 4.2: Comparison results of performance between **Methods I** and **II**. **Method II** uses parameters $\beta = 1$, $a = 0.5$ and thresholds $Th_w = 8-19$ Hz, and $Th_{Aw} = 12$. Parentheses indicate number of spikes. Shadow areas on the table represent significant performance.

Patients	Method I			Method II		
	TP	FP	FN	TP	FP	FN
1	86% (31)	14% (25)	45% (5)	71% (26)	52% (28)	29% (10)
2	100% (66)	55% (82)	0% (0)	86% (57)	38% (35)	14% (9)
3	70% (21)	42% (15)	30% (9)	83% (25)	56% (35)	17% (5)
4	96% (52)	68% (110)	4% (2)	88% (47)	59% (69)	12% (7)
5	79% (37)	57% (50)	21% (10)	70% (33)	55% (41)	30% (14)
Total	86% (207)	53% (282)	14% (26)	80% (188)	52% (205)	20% (45)

seconds, and one missed spike at time 237.7 seconds. The missed spike occurs with the algorithm that uses time indexes between two vectors such that h_4 has the correct spike while h_{17} shows the missed spike.

Table 4.2 shows comparison of performance between **Method I**, and **II**, and the percentages of correct spikes and missed spikes of **Method I** show better results than **Method II** while the percentage of non-spikes of this method shows better reduction of FP from 282 to 205 or percentage of reduced FP is 27.3%.

The feature extraction algorithm has pros and cons. The advantages are the reduction of the number of FP while increasing energy values of the temporal vector. The connections between the energy values and spikes can be seen by the addition of the temporal vectors, as shown in Figure 4.8(c). Drawbacks include an increase in FN values and a reduction of the results for TP . The algorithm will be improved in future work in order to improve the results.

NMF based on l_1 -ARD provides extensive computation and may not be suitable for EEG of ISS in the presence of HYPs because some discharges presented in **MP** are discontinuously distributed around high frequency. For example, at time 180 seconds in Figure 4.5(a), vertical portion distributed around frequencies 5-15 Hz are correct spike associated with the dash-line. In Figures 4.5(b), (c), w_4 representing a small portion corresponds to h_4 having low energy while w_9 representing center part of the portion corresponds to h_9 having high energy. Having small low energy may cause unexpected results of FN and TP .

Figure 4.4: Results of number of K_{eff} versus hyper parameter, $a = \{0.5, 100, 500, 1000\}$. Vertical and horizontal axes of each subplot represent K_{eff} and a , respectively. Each row indicates each patient, and columns from left to right indicate number of spikes $\{0, 1, 2, \dots, 6\}$. The title of each sub-figure shows the number of samples per ten seconds.

Figure 4.5: Result of using NMF based on l_1 -ARD with parameters $\beta = 1$, $a = 0.5$. (a) Original MP with spike locations indicated as dashed lines marked by the epileptologist. MP is decomposed to $\mathbf{W}, \mathbf{H}$ as shown (b), (c), respectively. (b) Spectral vectors sorted decrement values of MF, and the ranges of threshold (Th_w) are indicated by two horizontal dashed lines. Horizontal and vertical axes represent index k^{th} of w , and frequency [Hz]. (c) Temporal vectors are reordered corresponding to the index number of w_k and show spike locations as dashed lines. Horizontal and vertical axes represent time [sec], and index k^{th} of h . (d) Table contains index number, $MF(w_k)$, and area, $A(w_k)$. Notice that, after initial K was 30, K_{eff} is 10 and the remains are removed due to irrelevance.

Figure 4.6: Result of another example based on l_1 -ARD with parameters $\beta = 1$, $a = 0.5$. Following descriptions are similar to Figure 4.5. (a) Original MP. (b) Spectral vectors. (c) Temporal vectors. (d) Table contains index number, $MF(w_k)$, and area, $A(w_k)$. As seen in the table, w_4 and w_7 are in the conditions of the thresholds, and h_4 and h_7 are possibly indicated as spikes.

Figure 4.7: The result of common temporal vectors between h_4 and h_9 , depicted from Figure 4.5(c), as shown in (a), (b). (c) The result of the common h of the two vectors using Eqs.(3.20)-(3.22) in algorithm of feature extraction, and shows six correct spikes, three non spikes.

Figure 4.8: The result of another example with common temporal vectors between h_4 and h_{17} , depicted from Figure 4.6(c), as shown in (a), (b). (c) The result of common h displays three correct spikes, three non spikes and one missed spike.

4.3 Method III

In **Method I**, K was randomly chosen, and **Method II** provided K_{eff} but its reconstruction error was very poor. For further analysis with all patients, **Method III** uses average $K = 10$ from NMF based on l_1 -ARD to NMF based on projected gradient proposed by Lin [40]. Figure 4.3(d) shows improved reconstruction error. Parameters were set: $\text{tol} = 10^{-15}$, number of iteration $I = 3500$.

After decomposing MP, mean frequency ($\text{MF}(w_k)$), area of w_k and h_k ($A(w_k)$, $A(h_k)$), width of h_k (B_h), and height of h_k (C_h) were calculated and tabulated in Figure 3.3(c). Two classes (class 1 = spike, and 0 = non-spike) along with those features were manually specified in that table. Then, these attributes of the data were fed to K-fold cross validations in order to evaluate performance using SVM and compare the results of two features sets:

- Feature set 1: **F1** contains features $\{\text{MF}(w_k), A(w_k), A(h_k), B_h, C_h\}$.
- Feature set 2: **F2** contains features $\{\text{MF}(w_k), A(w_k), C_h\}$.

This method examines 10 folds (see Figure 3.4) based on SVM classifier to evaluate the performance of **F1** and **F2** by plotting receiver operating characteristic (ROC-curve) with different $\sigma = \{0.5, 1, 5, 10\}$. The results are shown in Table 4.3.

From results on Table 4.3, bold numbers show the highest values of accuracy (Acc), precision (Prec), sensitivity (Sens), specificity (Spec) and AUC per feature set. The highest average AUC of each **F1** and **F2** are $98.15 \pm 0.3\%$, $98.09 \pm 0.31\%$, respectively, at $\sigma = 10$. To compare the results in Figures 4.9(a), (b) with the same σ , feature set **F2** provides AUCs better than **F1** when the best and worst performances of each feature set were picked.

4.4 Results Summary

Overall, in **Method I**, the number of factor K was randomly picked to employ NMF and the Euclidean distance cost function was solved by block principal pivoting. The decomposition vectors, **W** and **H**, were analyzed to identify and quantify characteristics of spikes as HYPS if their mean frequency and peak values are greater than the given thresholds (Th_w, Th_h). The performance was evaluated by calculating percentages of correct spikes, non-spikes, and missed spikes as denoted true positive, false positive, false negative, and by comparing between the algorithm and commercial program (Persyst). The results of TP and FN showed that the algorithm achieved to locate spikes better than the other. However, FP will be reduced for future work.

In **Method II**, average effective K (or K_{eff}) can be found by using probabilistic Bayesian model for β -NMF based on automatic relevance determination (ARD) in order to avoid over-fitting, and under-fitting of learning data. A concept of this method is that a relevant weight (λ_k) ties k^{th} column vector of **W** and row vector of **H**. If λ_k is smaller than a threshold (τ), (w_k, h_k) are decoupled as making the decomposition sparse. The results of

Table 4.3: Results of performance based on K-Fold with different σ and comparison between **F1**, **F2**. Acc, Prec, Sens, Spec, AUC are denoted accuracy, precision, sensitivity, specificity and the area under the curve of ROC, respectively.

Feature sets	σ	Acc	Prec	Sens	Spec	AUC
F1	0.5	75.31 $\pm$ 2.01	96.04 $\pm$ 0.27	52.79 $\pm$ 4.11	97.83$\pm$ 0.17	93.42 $\pm$ 1.76
	1	89.59 $\pm$ 1.48	97.77$\pm$ 0.28	83.81 $\pm$ 3.15	95.37 $\pm$ 0.33	96.13 $\pm$ 0.44
	5	93.03 $\pm$ 5.13	94.28 $\pm$ 1.09	91.77 $\pm$ 11.72	94.30 $\pm$ 1.51	97.90 $\pm$ 0.23
	10	94.30$\pm$ 1.72	93.93 $\pm$ 0.83	94.76$\pm$ 4.44	93.83 $\pm$ 1.05	98.15$\pm$ 0.30
F2	0.5	87.45 $\pm$ 2.87	94.43$\pm$ 0.25	79.59 $\pm$ 6.08	95.30$\pm$ 0.41	95.00 $\pm$ 1.11
	1	93.52 $\pm$ 2.21	94.17 $\pm$ 0.23	92.79 $\pm$ 4.78	94.26 $\pm$ 0.44	96.29 $\pm$ 0.51
	5	95.30$\pm$ 0.65	93.93 $\pm$ 0.58	96.87$\pm$ 1.16	93.73 $\pm$ 0.64	97.95 $\pm$ 0.23
	10	94.09 $\pm$ 1.86	93.83 $\pm$ 0.85	94.42 $\pm$ 4.74	93.75 $\pm$ 1.09	98.09$\pm$ 0.31

K_{eff} using different values of hyperparameter (a) based on l_1 -ARD showed that K_{eff} increased when a increased. To compare the results of performance, the feature extraction algorithm of **Method II** had significantly reduced the number of non-spikes while **Method I** provided the numbers of correct spikes and missed spikes better than **Method II**. In fact, the algorithm of l_1 -ARD has drawback of extensive computation and provide results of ill-reconstruction error.

Finally, in **Method III**, an average K_{eff} of 10 was applied in NMF and Kullback-Leibler Divergence cost function was solved by projected gradient. This method mainly focused on classification using SVM based on K-Fold cross validation, and was assessed by plotting ROC-curve. The results of performance using K-fold method displayed AUCs increased when σ increased. At $\sigma = 10$, feature set **F2**, which contains $\{\text{MF}(w), \text{A}(w), \text{C}(h)\}$ showed more significant results than feature set **F1**, which contains $\{\text{MF}(w), \text{A}(w), \text{A}(h), \text{b}(h), \text{C}(h)\}$ by visualizing at AUCs of the ROC curves.

Figure 4.9: Results of ROC curves using K-Fold cross validation with parameter $\sigma = 10$ per feature set, **F1** and **F2**. From the worst to the best performance, AUCs of **F1** (a) and **F2** (b) are (90.38%-98.51%), and (93.14%-98.56%), respectively.

Chapter 5

Conclusions

There are limited available tools for analyzing discharges in the case of ISS. In this thesis, a novel TF feature extraction algorithm based on MP-TFD and NMF for identification of spikes in the presence of HYPS was proposed. The developed method in this thesis applies signal analysis, dimension reduction, and feature extraction in order to identify and quantify the EEG waveforms between spikes and non-spikes in the presence of HYPS.

To analyze the waveforms, MP-TFD represented epileptic discharges high resolution in TF domain was able to capture transient characteristics in short time frame. To extract from MP, three algorithms were investigated and the results of the performance based on two classification techniques, thresholdings and support vector machine (SVM), were discussed.

The first algorithm using NMF based on Euclidean Distance cost function with an arbitrary model order and thresholdings for classification aimed to determine relationships between characteristics of the waveforms from the decomposed vectors. The second algorithm, which used β -NMF based on Kullback-Leibler Divergence (KLD: $\beta = 1$) cost function solved by majorization-minimization (MM) and thresholdings for classification, aimed to determine an optimal model selection (K_{eff}) using Bayesian approach and provided an algorithm of feature extraction to reduce. The third algorithm, which used NMF based on Kullback-Leibler Divergence cost function with an average K_{eff} of 10 solved by projected gradient, focused on classification using SVM.

These methods were performed on a dataset collected from five ISS individuals. The first and second methods resulted in a significantly higher success rate than the existing temporal spike detection method, such as the clinical software (Persyst). The results showed average true positive (TP) and false negative (FN) percentages of 86% and 14%, respectively, for the first method, and 80% and 20%, respectively, for the second method while Persyst showed the results of TP and FN percentages of 4% and 96%, respectively. Using SVM with 10-Fold cross validations, the third method exhibited a remarkably improved area under curve (AUC) of the receiver operating characteristic (ROC), up to 98.56%.

This type of automatic spike detection for EEG can assist epileptologists by identifying epileptiform discharges of interest from long-term EEG recordings; however, there are limited tools available for discharge detection in the case of ISS, which is characterized by HYPS on EEG. Hence, the developed algorithm provides a novel tool to identify and quantify the presence of spikes in the presence of HYPS. The quantitative assessment of

spike detection, as well as other features of HYPS is expected to allow a more accurate, automatic assessment of the relevance of EEG to clinical outcome, which is significantly important in therapy management of ISS.

Bibliography

- [1] R. M. S. S. Abeysekera, R. J. Bolton, L. C. Westphal, and B. Boashash. Patterns in Hilbert transforms and Wigner-Ville distributions of electrocardiogram data. In *Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP'86.*, volume 11, pages 1793–1796. IEEE, 1986.
- [2] T. Z. Baram. Models for infantile spasms: An arduous journey to the holy grail. *Annals of neurology*, 61(2):89–91, 2007.
- [3] E. Battenberg and D. Wessel. Accelerating Non-Negative Matrix Factorization for Audio Source Separation on Multi-Core and Many-Core Architectures. In *ISMIR*, pages 501–506, 2009.
- [4] J. M. Becker, C. Sohn, and C. Rohlfling. NMF with spectral and temporal continuity criteria for monaural sound source separation. In *Signal Processing Conference (EUSIPCO), 2014 Proceedings of the 22nd European*, pages 316–320. IEEE, 2014.
- [5] C. M. Bishop. *Pattern recognition and machine learning*. springer, 2006.
- [6] B. Boashash. Note on the use of the Wigner distribution for time-frequency signal analysis. 1988.
- [7] B. Boashash. *Time frequency analysis*. Gulf Professional Publishing, 2003.
- [8] B. Boashash, G. Azemi, and J. M. O'Toole. Time-frequency processing of nonstationary signals: Advanced TFD design to aid diagnosis with highlights from medical applications. *Signal Processing Magazine, IEEE*, 30(6):108–119, 2013.
- [9] B. Boashash, N. A. Khan, and T. Ben-Jabeur. Time-frequency features for pattern recognition using high-resolution TFDs: A tutorial review. *Digital Signal Processing*, 40:1–30, 2015.
- [10] B. Bouachache and P. Flandrin. Wigner-Ville analysis of time-varying signals. In *Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP'82.*, volume 7, pages 1329–1332. IEEE, 1982.

- [11] Q. Chang, Q. Chen, and X. Wang. Scaling Gaussian RBF kernel width to improve SVM classification. In *Neural Networks and Brain, 2005. ICNN&B'05. International Conference on*, volume 1, pages 19–22. IEEE, 2005.
- [12] H. I. Choi and W. J. Williams. Improved time-frequency representation of multicomponent signals using exponential kernels. *Acoustics, Speech and Signal Processing, IEEE Transactions on*, 37(6):862–871, 1989.
- [13] A. Cichocki, R. Zdunek, and S. I. Amari. Csiszar’s divergences for non-negative matrix factorization: Family of new algorithms. In *Independent Component Analysis and Blind Signal Separation*, pages 32–39. Springer, 2006.
- [14] L. Cohen. *Time-frequency analysis*, volume 1. Prentice hall, 1995.
- [15] C. Damon, A. Liutkus, A. Gramfort, and S. Essid. Non-negative matrix factorization for single-channel EEG artifact rejection. In *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*, pages 1177–1181. IEEE, 2013.
- [16] L. Debnath. *Wavelet transforms and time-frequency signal analysis*. Springer Science & Business Media, 2012.
- [17] C. Févotte, N. Bertin, and J. L. Durrieu. Nonnegative matrix factorization with the Itakura-Saito divergence: With application to music analysis. *Neural computation*, 21(3):793–830, 2009.
- [18] D. Gabor. Theory of communication. Part 1: The analysis of information. *Journal of the Institution of Electrical Engineers-Part III: Radio and Communication Engineering*, 93(26):429–441, 1946.
- [19] B. Ghoraani and S. Krishnan. Time–frequency matrix feature extraction and classification of environmental audio signals. *Audio, Speech, and Language Processing, IEEE Transactions on*, 19(7):2197–2209, 2011.
- [20] F. A. Gibbs and E. L. Gibbs. Atlas of electroencephalography. 1941.
- [21] E. S. Goldensohn, A. D. Legatt, S. Koszer, and S. M. Wolf. *Goldensohn’s EEG interpretation: problems of overreading and underreading*. Futura publishing, 1999.
- [22] E. F. Gonzalez and Y. Zhang. Accelerating the Lee-Seung algorithm for non-negative matrix factorization. *Dept. Comput. & Appl. Math., Rice Univ., Houston, TX, Tech. Rep. TR-05-02*, 2005.

- [23] F. Gu and A. Ball. Use of the Smoothed Pseudo-Wigner-Ville Distribution. *Maintenance & Asset Management Journal*, 10(2):16–23, 1995.
- [24] H. Hassanpour, M. Mesbah, and B. Boashash. Time-frequency feature extraction of newborn EEG seizure using SVD-based techniques. *EURASIP Journal on Applied Signal Processing*, 2004:2544–2554, 2004.
- [25] R. A. Hrachovy, J. D. Frost, and P. Kellaway. Hypsarrhythmia: variations on the theme. *Epilepsia*, 25(3):317–325, 1984.
- [26] D. R. Hunter and K. Lange. A tutorial on MM algorithms. *The American Statistician*, 58(1):30–37, 2004.
- [27] S. A. Hussain, G. Kwong, J. J. Millichap, J. R. Mytinger, N. Ryan, J. H. Matsumoto, J. Y. Wu, J. T. Lerner, and R. Sankar. Hypsarrhythmia assessment exhibits poor inter-rater reliability: A threat to clinical trial validity. *Epilepsia*, 56(1):77–81, 2015.
- [28] W. J. Hwang, S. H. Wang, and Y. T. Hsu. Spike Detection Based on Normalized Correlation with Automatic Template Generation. *Sensors*, 14(6):11049–11069, 2014.
- [29] P. M. Jeavons and B. D. Bower. The natural history of infantile spasms. *Archives of disease in childhood*, 36(185):17, 1961.
- [30] I. Jolliffe. *Principal component analysis*. Wiley Online Library, 2002.
- [31] R. D. Jones, A. Dingle, G. J. Carroll, R. D. Green, M. Black, I. M. Donaldson, P. J. Parkin, P. J. Bones, K. L. Burgess, et al. A system for detecting epileptiform discharges in the EEG: real-time operation and clinical trial. In *Engineering in Medicine and Biology Society, 1996. Bridging Disciplines for Biomedicine. Proceedings of the 18th Annual International Conference of the IEEE*, volume 3, pages 948–949. IEEE, 1996.
- [32] H. Kameoka, N. Ono, K. Kashino, and S. Sagayama. Complex NMF: A new sparse representation for acoustic signals. In *Acoustics, Speech and Signal Processing, 2009. ICASSP 2009. IEEE International Conference on*, pages 3437–3440. IEEE, 2009.
- [33] S. S. Keerthi and C. J. Lin. Asymptotic behaviors of support vector machines with Gaussian kernel. *Neural computation*, 15(7):1667–1689, 2003.
- [34] H. Kim and H. Park. Sparse non-negative matrix factorizations via alternating non-negativity-constrained least squares for microarray data analysis. *Bioinformatics*, 23(12):1495–1502, 2007.

- [35] H. Kim and H. Park. Nonnegative matrix factorization based on alternating non-negativity constrained least squares and active set method. *SIAM Journal on Matrix Analysis and Applications*, 30(2):713–730, 2008.
- [36] J. Kim, Y. He, and H. Park. Algorithms for nonnegative matrix and tensor factorizations: A unified view based on block coordinate descent framework. *Journal of Global Optimization*, 58(2):285–319, 2014.
- [37] J. Kim and H. Park. Toward faster nonnegative matrix factorization: A new algorithm and comparisons. In *Data Mining, 2008. ICDM'08. Eighth IEEE International Conference on*, pages 353–362. IEEE, 2008.
- [38] S. Kim and J. McNames. Automatic spike detection based on adaptive template matching for extracellular neural recordings. *Journal of neuroscience methods*, 165(2):165–174, 2007.
- [39] D. D. Lee and H. S. Seung. Algorithms for non-negative matrix factorization. In *Advances in neural information processing systems*, pages 556–562, 2001.
- [40] C. B. Lin. Projected gradient methods for nonnegative matrix factorization. *Neural computation*, 19(10):2756–2779, 2007.
- [41] C. J. Lin. On the convergence of multiplicative update algorithms for nonnegative matrix factorization. *Neural Networks, IEEE Transactions on*, 18(6):1589–1596, 2007.
- [42] Y. C. Liu, C. C. K. Lin, J. J. Tsai, and Y. N. Sun. Model-based spike detection of epileptic EEG data. *Sensors*, 13(9):12536–12547, 2013.
- [43] S. S. Lodder, J. Askamp, and M. J. van Putten. Inter-ictal spike detection using a database of smart templates. *Clinical neurophysiology*, 124(12):2328–2335, 2013.
- [44] F. Lotte, M. Congedo, A. Lécuyer, and F. Lamarche. A review of classification algorithms for EEG-based brain–computer interfaces. *Journal of neural engineering*, 4, 2007.
- [45] P. LúAvígsson, E. Ólafsson, S. Sigurardóttir, and W. A. Hauser. Epidemiologic features of infantile spasms in Iceland. *Epilepsia*, 35(4):802–805, 1994.
- [46] M. B. Malarvili and M. Mesbah. Newborn seizure detection based on heart rate variability. *Biomedical Engineering, IEEE Transactions on*, 56(11):2594–2603, 2009.
- [47] S. G. Mallat and Z. Zhang. Matching pursuits with time-frequency dictionaries. *Signal Processing, IEEE Transactions on*, 41(12):3397–3415, 1993.

- [48] K. E. Misulis. *Atlas of EEG, Seizure Semiology, and Management*. Oxford University Press, 2013.
- [49] Z. Nenadic and J. W. Burdick. Spike detection using the continuous wavelet transform. *Biomedical Engineering, IEEE Transactions on*, 52(1):74–87, 2005.
- [50] E. Niedermeyer and F. L. da Silva. *Electroencephalography: basic principles, clinical applications, and related fields*. Lippincott Williams & Wilkins, 2005.
- [51] A. V. Oppenheim, R. W. Schaffer, J. R. Buck, et al. *Discrete-time signal processing*, volume 2. Prentice-hall Englewood Cliffs, 1989.
- [52] P. Paatero and U. Tapper. Positive matrix factorization: A non-negative factor model with optimal utilization of error estimates of data values. *Environmetrics*, 5(2):111–126, 1994.
- [53] J. M. Pellock, R. Hrachovy, S. Shinnar, T. Z. Baram, D. Bettis, D. J. Dlugos, W. D. Gaillard, P. A. Gibson, G. L. Holmes, D. R. Nordli, et al. Infantile spasms: a US consensus report. *Epilepsia*, 51(10):2175–2189, 2010.
- [54] T. Peng, Y. Chen, and Z. Liu. A Time–Frequency Domain Blind Source Separation Method for Underdetermined Instantaneous Mixtures. *Circuits, Systems, and Signal Processing*, pages 1–13, 2015.
- [55] L. F. Portugal, J. J. Júdice, and L. N. Vicente. A comparison of block pivoting and interior-point algorithms for linear least squares problems with nonnegative variables. *Mathematics of Computation*, 63(208):625–643, 1994.
- [56] R. Riikonen. Epidemiological data of West syndrome in Finland. *Brain and Development*, 23(7):539–541, 2001.
- [57] M. E. Tagluk and M. J. English. Time-frequency analysis of ECG signals using an image processing technique. *Journal of medical engineering & technology*, 23(5):190–195, 1999.
- [58] V. Y. F. Tan and C. Févotte. Supplementary material to Automatic Relevance Determination in Nonnegative Matrix Factorization with the β -Divergence, 2012.
- [59] V. Y. F. Tan and C. Févotte. Automatic Relevance Determination in Nonnegative Matrix Factorization with the/spl beta/-Divergence. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 35(7):1592–1605, 2013.

- [60] S. K. Tjoa and K.J. R. Liu. Multiplicative update rules for nonnegative matrix factorization with co-occurrence constraints. In *Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on*, pages 449–452. IEEE, 2010.
- [61] A. T. Tzallas, M. G. Tsipouras, D. Fotiadis, et al. Epileptic seizure detection in EEGs using time–frequency analysis. *Information Technology in Biomedicine, IEEE Transactions on*, 13(5):703–710, 2009.
- [62] V. Vapnik. *The nature of statistical learning theory*. Springer Science & Business Media, 2013.
- [63] S. A. Vavasis. On the complexity of nonnegative matrix factorization. *SIAM Journal on Optimization*, 20(3):1364–1377, 2009.
- [64] J. de. Ville et al. Théorie et applications de la notion de signal analytique. *Cables et transmission*, 2(1):61–74, 1948.
- [65] T. Virtanen. Monaural sound source separation by nonnegative matrix factorization with temporal continuity and sparseness criteria. *Audio, Speech, and Language Processing, IEEE Transactions on*, 15(3):1066–1074, 2007.
- [66] W. J. West. On a peculiar form of infantile convulsions. *The Lancet*, 35(911):724–725, 1841.
- [67] Wikipedia. Support vector machine. https://en.wikipedia.org/wiki/Support_vector_machine. Accessed: 2015-09-8.
- [68] H. P. Zaveri, R. B. Duckrow, and S. S. Spencer. On the use of bipolar montages for time-series analysis of intracranial electroencephalograms. *Clinical neurophysiology*, 117(9):2102–2108, 2006.