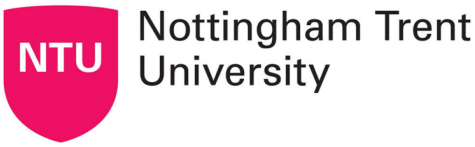
# IDENTIFICATION OF BRAIN EPILEPTIFORM DISCHARGES FROM ELECTROENCEPHALOGRAMS

BAHMAN ABDI-SARGEZEH

A thesis submitted to Nottingham Trent University in candidature of the Degree of Doctor of Philosophy




School of Science and Technology
Depertemant of Computer Science
Nottingham Trent University

October 2022

Bahman Abdi-Sargezeh: *IDENTIFICATION OF BRAIN EPILEPTIFORM DISCHARGES FROM ELECTROENCEPHALOGRAMS* , A thesis submitted to Nottingham Trent University in candidature of the Degree of Doctor of Philosophy, © October 2022

SUPERVISORS:
Prof. Saeid Sanei
Dr. James Lewis

# DECLARATION

I hereby declare that the this thesis is my original work and it has been written by me in its entirety. I have duly acknowledged all the sources of information which have been used in the thesis. This thesis has also not been submitted for any degree in any university previously.

*Nottingham, UK, October 2022*

Bahman Abdi-Sargezeh

# ACKNOWLEDGEMENTS

I would like to thank my wonderful supervisor, Professor Saeid Sanei, for his continuing support over these years. His encouragement, enthusiasm, friendship, and invaluable advices have been very useful to accomplish this thesis. His guidance and patience helped me gain independence in my research. I would also like to thank Neurologists Dr. Antonio Valentin and Professor Gonzalo Alarcon from King's College London for providing the necessary dataset and for their great clinical advices and guidance.
At the end, I would like to thank my friends and family for their love and support. Most importantly, to my beloved wife, best friend, and my rock, Maryam: Your love, support, and unwavering kindness made this possible. This thesis is dedicated to you.

# ABSTRACT

Brain interictal epileptiform discharges (IEDs), as the fundamental indicators of seizure, are transient events occurring between two or before seizure onsets, captured using electroencephalogram (EEG). For epilepsy diagnosis and localization of seizure sources, both interictal and ictal recordings are extremely informative. Accurate detection of IEDs from over the scalp helps faster diagnosis of epilepsy. The scalp EEG (sEEG) suffers from a low signal-to-noise ratio and high attenuation of IEDs due to the high skull electrical impedance. On the other hand, the intracranial EEG (iEEG) recorded using implanted electrodes enjoys high temporal-spatial resolution and enables capturing most IEDs. Therefore, in this thesis, the focus is on the identification of IEDs from the concurrent scalp and intracranial EEGs.

Multi-way analysis provides an opportunity to jointly analyse the data in different domains. IEDs may share some features within and between the segments. We have developed methods based on multi-way analysis and tensor factorization to detect the IEDs from the concurrent sEEG in both segmented and real-time approaches.

The diversities in IED morphology, strength, and source location within the brain cause a great deal of uncertainty in their labeling by clinicians. We have exploited and incorporated this uncertainty (the probability of the waveform being an IED) in an IED detection system. Furthermore, IEDs are naturally sparse. We have benefited from the sparsity of IED waveforms in developing an algorithm to exploit sparse common features among the IED segments, referred to as sparse common feature analysis.

By mapping sEEG to iEEG, the sEEG quality is improved. In this thesis, the proposed tensor factorization maps the time-frequency features of sEEG to those of iEEG to detect the IEDs from over the scalp with high sensitivity. We have concatenated time, frequency, and channel modes of iEEG recordings into a tensor. After decomposing the tensor into temporal, spectral, and spatial components, the EEG time-frequency features have been extracted and projected onto the temporal components. Furthermore, we have developed two novel algorithms based on generative adversarial networks to map the raw sEEG to iEEG.

As a result of this work, the visibility of IEDs from sEEG has over 4-fold improvement. Additionally, the outcome paves the path for future research in epilepsy prediction, seizure source localisation, and modeling the brain seizure pathways.

# PUBLICATIONS

***Journals:***

**Abdi-Sargezeh, B.**, Valentin, A., Alarcon, G., & Sanei, S. (2022). "EEG-to-EEG: Scalp-to-Intracranial EEG Translation using a Combination of Variational Autoencoders and Generative Adversarial Networks." *Submitted to IEEE Transactions on Neural Networks and Learning Systems.*

**Abdi-Sargezeh, B.**, Valentin, A., Alarcon, G., & Sanei, S. (2022). "Sparse Common Feature Analysis for Detection of Interictal Epileptiform Discharges from Concurrent Scalp EEG." *IEEE Access,* 10, pp. 49892-49904.

**Abdi-Sargezeh, B.**, Valentin, A., Alarcon, G., & Sanei, S. (2021). "Incorporating uncertainty in data labeling into automatic detection of interictal epileptiform discharges from concurrent scalp-EEG via multi-way analysis." *International Journal of Neural Systems,* 31(08), pp. 2150019.

**Abdi-Sargezeh, B.**, Valentin, A., Alarcon, G., Martin-Lopez, D., & Sanei, S. (2021). "Higher-order tensor decomposition based scalp-to-intracranial EEG projection for detection of interictal epileptiform discharges." *Journal of Neural Engineering*, 18(6), pp. 066039.

**Abdi-Sargezeh, B.**, & Sanei, S. (2021). "Advances in epilepsy monitoring by detection and analysis of brain epileptiform discharges." *Psychology & Neuroscience.* DOI: 10.1037/pne0000275

***Conferences:***

**Abdi-Sargezeh, B.**, Valentin, A., Alarcon, G., & Sanei, S. (2022, May). "Online Detection of Scalp-Invisible Mesial-Temporal Brain Interictal Epileptiform Discharges from EEG." *In IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* pp. 1416-1420.

**Abdi-Sargezeh, B.**, Valentin, A., Alarcon, G., & Sanei, S. (2021). "Incorporating uncertainty in data labeling into detection of brain interictal epileptiform discharges from EEG using weighted optimization." *In IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP),* pp. 1000-1004.

**Abdi-Sargezeh, B.**, Valentin, A., Alarcon, G., & Sanei, S. (2021). "Detection of brain interictal epileptiform discharges from intracranial EEG by exploiting their morphology in the tensor structure." *In 2021 29th European Signal Processing Conference (EUSIPCO),* pp. 1167-1171.

CONTENTS

# LIST OF FIGURES

# LIST OF TABLES

# ALGORITHMS

# LIST OF ABBREVIATIONS

| | |
|---|---|
| ACC | Accuracy |
| AE | Autoencoder |
| ALS | Alternating Least Squares |
| ASAE | Asymmetric-Symmetric Autoencoder |
| AUC | Area Under the Receiver Operating Characteristic Curve |
| BSS | Blind Signal Separation |
| CAR | Common Average Reference |
| CFA | Common Feature Analysis |
| cGAN | conditional Generative Adversarial Network |
| CNN | Convolutional Neural Networks |
| CNN-Bin | Binary Convolutional Neural Network |
| CNN-Multi | Multiclass Convolutional Neural Network |
| COBE | Common Orthogonal Basis Extraction |
| CP-OPT | CANDECOMP/PARAFAC OPTimization |
| CP-WOPT | Weighted CANDECOMP/PARAFAC Optimization |
| CPD | CANDECOMP/PARAFAC Decomposition |
| DLDA | Diagonal Linear Discriminant Analysis |
| DNN | Deep Neural Network |
| DTE | Decision Tree Ensemble |
| EEG | Electroencephalogram |
| EEGer | Electroencephalographers |
| F1-S | F1-score |
| FC | Fully Connected |
| FCD | Focal Cortical Dysplasia |

| | |
|---|---|
| FN | False Negative |
| FO | Foramen Ovale |
| FP | False Positive |
| GAN | Generative Adversarial Network |
| GS | Gold Standard Spike |
| HOOI | Higher Order Orthogonal Iteration |
| HOSVD | Higher Order Singular Value Decomposition |
| HS | Hippocampal Sclerosis |
| IED | Interictal Epileptiform Discharges |
| iEEG | Intracranial Electroencephalogram |
| KNN | k-Nearest Neighbors |
| LReLU | Leaky Rectified Linear Unit |
| LSR | Least-Squares Regression |
| LSTM | Long Short-Term Memory |
| MRMR | Minimum Redundancy — Maximum Relevance |
| MStI | Mapping Scalp to Intracranial |
| NB | Naïve Bayes |
| NCPD | Nonnegative CANDECOMP/PARAFAC Decomposition |
| NTD | Nonnegative Tucker Decomposition |
| OMP | Orthogonal Matching Pursuit |
| PCA | Principal Component Analysis |
| ReLU | Rectified Linear Unit |
| RNN | Recurrent Neural Network |
| ROC | Receiver Operating Characteristic |
| SCFA | Sparse Common Feature Analysis |
| SCOBE | Sparse Common Orthogonal Basis Extraction |

| | |
|---|---|
| SE | Standard Error |
| sEEG | Scalp Electroencephalogram |
| SEN | Sensitivity |
| SNN | Shallow Neural Network |
| SPC | Specificity |
| SPADE | SPatially-Adaptive DEnormalization |
| STD | Standard Deviation |
| STFT | Short-Term Fourier Transform |
| SVD | Singular Value Decomposition |
| SVM | Support Vector Machines |
| Tanh | Hyperbolic Tangent |
| TD | Tucker Decomposition |
| TF | Time-Frequency |
| TN | True Negative |
| TP | True Positive |
| U-net | U Network |
| VAE | Variational Autoencoder |
| WF | Wavelet Feature |

# 1

# INTRODUCTION

The human brain is a complex network of neurons that controls the information flow and human interactions with the environment. The study of how neurons activate is being investigated in order to diagnose and treat many neurological diseases, such as epilepsy, Parkinson's, dementia, and Alzheimer's. The neural activity of the human brain is measured using different methods. Electroencephalogram (EEG) and functional magnetic resonance imaging are the most popular techniques. EEG is a brain recording modality that captures the electric activity of the brain. It can be recorded either from the scalp (here called scalp EEG (sEEG)) or from the cerebral cortex or deeper brain (called intracranial EEG (iEEG)) [2]. The sEEG is recorded in a non-invasive manner and has no side effects for individuals. However, it suffers from low resolution, signal-to-noise ratio, and the crucial internal brain sources being mixed and blurred at each electrode. On the other hand, the iEEG recorded in an invasive approach enjoys high resolution of the neuron activities as the background cortical activities and attenuation are less prominent. Here, we aim to analyse the sEEG and iEEG recorded simultaneously from patients suffering from mesial-temporal lope epilepsy.

According to the World Health Organization [3], epilepsy is the most common neurological disease affecting 50 million people worldwide. It is estimated that up to 70% of people with epilepsy could live seizure-free if properly diagnosed and treated [3]. Therefore, the study of epilepsy and the effort at early diagnosis have always been of paramount importance in the biomedical field of research.

Epilepsy is a chronic brain disease characterised by epileptic seizures occurring due to excessive discharges of a group (or groups) of neurons in the cerebral cortex or hippocampus. The mainstay of diagnosis remains the detection of interictal (or pre-ictal) epileptiform discharges (IEDs) which occur between two seizure onsets.

IED identification can establish a guideline for pre-ictal state monitoring, seizure prediction, treatment, and surgical planning. Seizure prediction mitigates taking regular anticonvulsants, preliminary surgical interventions for the identification of ED generators, and the hazards of fall injury. In the traditional seizure prediction methods, the increase in synchrony and decrease in the chaotic behavior of cortical activities have been the main quantification parameters, regrettably, with low accuracy and insufficient consistency in terms of seizure types and across the subjects. Identifying the IEDs from sEEG up to sufficient accuracy has raised hopes for developing a new direction for more concise seizure prediction.

In terms of their morphology, IEDs can fall into five groups: (1) spike, lasting for 20-70 ms and detectable from the background activity; (2) sharp wave, the same as spikes but lasting for 70-200 ms; (3) sharp-and-slow-wave complex, comprised of a sharp wave followed by a slow wave; (4) spike-and-slow-wave complex, consisting of a spike followed by a slow wave; and (5) multiple spike-and-slow-wave complexes, the same as spike-and-slow-wave complex but with two or more spikes associated with one or more slow waves [4]. In addition, depending on the location of seizure sources, the spatial distribution of IEDs varies.

sEEG suffers from low resolution and consequently fails in capturing all IEDs. As a result, around 30% to 40% of patients considered for epilepsy surgery require iEEG recording [5]. However, the iEEG is recorded using invasive techniques with many side effects. Therefore, improving the sensitivity of sEEG in detecting IEDs is crucial in clinical practice.

## 1.1 CHALLENGES IN IED DETECTION

There are several challenges in using scalp-recorded data to detect IEDs automatically. The first challenge is that studies conducted on concurrent sEEG and iEEG recordings have revealed that standard electrodes record only a relatively small proportion of spikes detectable from over the scalp or at the cortical surface. Nayak et al. recorded simultaneously sEEG (using 20 standard scalp electrodes) and iEEG (using 12 FO electrodes) from patients suffering from temporal lobe epilepsy from 1990 to 1998 [6]. Their study has shown that only 9% of IEDs are observable over the scalp without averaging across IED segments or concurrent iEEG recordings. Less than thirteen percent (12.8%) of IEDs were detectable over the scalp as a small transient by referring to the concurrent iEEG as ground truth, and 59.7% of IEDs were identified over the scalp by referring to the simultaneous iEEG and averaging across IED segments. Finally, 18.7% of IEDs showed no signal or signature on the scalp. Later, Yamazaki *el al.* compared dense EEG and iEEG for interictal spike detection [7]. The dense EEG was recorded using 256 standard electrodes and, for recording the iEEG, from 48 to 102 subdural strip and grid electrodes were implanted over the mesial and lateral temporal lobes. They showed that 45% of spikes can be detected over the dense EEG. However, by reducing the number of channels to 19 according to 10-20 system, they detected only 22% of the spikes over the scalp. Nayak et al. [6] and Yamazaki et al. [7] obtained different results in terms of the percentage of visible IEDs from over the scalp. This can be due to using old recording systems [6] compared to [7]. However, this challenge has been addressed in very few studies [8, 9]. One way to solve this problem is to record the sEEG and iEEG signals simultaneously, then use the iEEG as ground truth for the IED annotation

and detect the IEDs from sEEG recordings [8]. Another way is to derive the iEEG from sEEG signals through applying signal processing and machine learning methods [9].

The second challenge in automatically detecting IEDs is that they are similar to many normal brain activities and artefacts, such as extracerebral potentials from muscles, eyes, heart, and electrodes. Eyeblink and heartbeat artefacts also resemble spike waves. However, they differ from IEDs in terms of source location and frequency range. Therefore, analysing EEG signals in the frequency domain and involving space diversity can suppress such artefacts. Through multi-way analysis, spatial and frequency components were analyzed to detect epileptic spikes [10].

The final challenge in IED detection is uncertainty in annotating a waveform as IED. There is a huge disagreement among expert epileptologists as to whether a wave is an IED or not. Webber et al. conducted research to find out how electroencephalographers (EEGers) interpret EEG signals [11]. Eight EEGers scored IEDs in twelve short EEG records. Only 18% were marked as IEDs by all the EEGers and 38% by only one EEGer. Quite recently, Halford et al. performed research to evaluate the neurologist performances in IED detection [12]. Thirty-five EEGers participated in the research and were supposed to annotate the IEDs in 200 EEG segments, each with a length of 300 seconds. The number of events marked as IEDs was different among EEGers from 6 to 212 (with a mean of 67.7).

It is also worth mentioning that the IEDs referred to in this thesis are the most common types. There are rare IED types, such as strong spike trains resulting from idiopathic generalized epilepsy, which can be observed over the scalp and easily detected.

## 1.2 AIMS AND OBJECTIVES

Human knowledge about the brain function is still insufficient to understand the properties of an epileptic brain. Automatically detecting IEDs from over the scalp is indeed a big challenge. Models provided for the detection of IEDs scored by sEEG are inadequate since they are unable to detect the scalp-invisible IEDs, thereby being biased to detect only a subset of IEDs that are visible over the scalp. Only very few studies have been carried out to investigate scalp IEDs from simultaneous sEEG and iEEG recordings [8, 9].

The techniques used in IED detection involve many advanced signal processing and machine learning techniques as well as their associated mathematics. The aim of this project is therefore to design and develop algorithms to gain new insights into trends and patterns of the EEG signals of epileptics. In summary, the objectives of this thesis are as follows:

1. Review state-of-the-art research in IED detection.

2. Develop and apply algorithms to the concurrent sEEG and iEEG to detect both scalp-visible and scalp-invisible IEDs.

3. Improve the sensitivity of sEEG in IED detection by extracting and analysing common components.

4. Analyse the ongoing (online) sEEG and iEEG in multi-dimensions using tensor factorisation.

5. Incorporate the morphology of IEDs into a tensor structure to boost the performance of an IED detection system.

6. Incorporate the level of uncertainty in labeling the IEDs given by the clinicians into the IED detection system.

7. Map time-frequency (TF) features of the sEEG into that of the iEEG.

8. Develop a data-driven approach and estimate the iEEG from the given sEEG using a generative adversarial network (GAN).

## 1.3 ORGANIZATION OF THE THESIS

In the first chapter, we have explained why detecting the IEDs with high accuracy is crucial. Afterward, the challenges of IED detection systems have been addressed. Finally, we have described our aims and objectives.

Chapter 2 is devoted to reviewing the publications in the IED detection area. The research that developed or employed machine learning techniques to detect or analyse the IEDs are reviewed.

In Chapter 3, the concept of tensor factorisation is first described. Then, the developed tensor-based technique proposed for detecting IEDs from the ongoing sEEG and iEEG is described. This chapter introduces two new methods, namely temporal components analysis and spatial component analysis, for IED identification.

Chapter 4 introduces a novel method namely spatial and morphological component analysis for IED detection. This method exploits the morphology of IEDs in the tensor structure. Here, a 4-way tensor with time, channel, segment, and morphology dimensions is constructed. After decomposing the tensor, the IED and non-IED data segments are projected onto the spatial and morphological factors.

The diversity in IED morphologies, strengths, and source locations within the brain cause a great deal of uncertainty in their labeling by clinicians. Chapter 5 proposes a method to exploit and incorporate this uncertainty (the probability of the waveform being an IED) in the IED detection system, which combines spatial component analysis with the IED probabilities.

In Chapter 6, we first present a novel method, called common feature analysis, for IED detection. In the second approach, we benefit from the sparsity of IED waveforms in developing another new algorithm for IED detection, namely sparse common feature analysis.

Chapter 7 is devoted to introducing a model based on tensor factorisation to map the TF features of sEEG to that of iEEG. Continuous wavelet transform (CWT) is employed to extract the TF features. Time, frequency, and channel modes of IED segments from intracranial recordings are concatenated into a four-way tensor. After decomposing the tensor into temporal, spectral, spatial, and segmental factors, the TF features of both IED and non-IED segments from scalp recordings are projected onto the temporal components for classification.

We propose a couple of EEG-to-EEG translation models based on the GAN structure in Chapter 8. In the first one, a GAN based on a U-net network is employed to map the sEEG to iEEG. In the second one, a conditional GAN is combined with a variational autoencoder to estimate the iEEG from the given sEEG.

Finally, in Chapter 9, the presented work in the thesis is summarised and concluded, and some future works are suggested.

# 2

# LITERATURE REVIEW

In the diagnosis of epilepsy and localization of seizure sources, both interictal and ictal recordings are extremely informative. For this purpose, computerized intelligent spike and seizure detection techniques have been researched and are constantly improving. This is to detect more IEDs from over the scalp and classify epileptic and non-epileptic discharges. Here, the EEG recording techniques in IED detection studies are explained first. Then, we introduce the most used private and publicly available IED datasets. Afterward, the methods used for IED detection are reviewed, which is the main contribution of this chapter. These techniques have opened a new window to the epilepsy diagnosis and management spheres. Finally, the evaluation criteria used as performance metrics are presented.

## 2.1 EEG RECORDING TECHNIQUES IN IED DETECTION STUDIES

For capturing IEDs, the EEG has been recorded from both scalp (sEEG) and sub-cortex, and hippocampus (iEEG) capturing brain mesial-temporal activities. Here, we explain the most used techniques for recording the EEG from epileptics suffering from temporal lobe epilepsy.

For recording the sEEG, standard scalp electrodes are usually placed on the scalp according to the 10-20 system [13, 14] or Maudsley electrode placement system [8, 15]. Maudsley electrode placement system is essentially similar to 10–20 system, but mid-temporal, posterior-temporal, and occipital electrodes in the Maudsley system are approximately 20 mm lower than those in the 10–20 system. Therefore, it provides more extensive coverage of the lower part of the cerebral convexity and adapts itself to cranial asymmetries [16], improving the recordings from temporal lobes [6].

For recording the iEEG, different types of electrodes, from subdural electrodes [1] to Stereo-EEG electrodes [17] and Foramen ovale (FO) electrodes, have been employed. Among them, FO electrodes [18] have been extensively used in recording iEEG from epileptic patients suffering from temporal lobe epilepsy [19, 20]. This technique counts as a semi-invasive method. The FO electrodes are inserted through FO holes in the skull and directly placed on the exposed mesial temporal structures. Therefore, it provides an opportunity to simultaneously record the sEEG and iEEG without disrupting brain coverings [21, 22].

## 2.2 IED DATASETS

Most studies used their private datasets, meaning they recorded EEGs from epileptic patients, annotated the IEDs, and finally classified the IEDs and non-IEDs. However, there are also public datasets for sEEG [23] and iEEG [1].

### 2.2.1 *Public Scalp IED Dataset*

There is only a public scalp IED dataset [23]. TUH-EEG Corpus is the most extensive publicly available corpus of clinical EEG recordings worldwide, available to download (www.nedcdata.org). The data was recorded at the Department of Neurology at Temple University Hospital at various sampling frequencies from 250 to 1024 with different electrode configurations. In addition, a team of neurologists and undergraduate annotators annotated a subset of TUH-EEG in which EEGs were recorded using the averaged reference electrode configuration, called TUH EEG Events Corpus and freely available (https://isip.piconepress.com/projects/tuh_eeg/html/downloads.shtml).

The annotations fell into six groups:

- spike and/or sharp waves: patterns of EEGs observed during epileptic seizures,
- periodic lateralized epileptiform discharges: a pattern of repetitive periodic, focal, or hemispheric epileptiform discharges like sharp waves or spikes,
- generalized periodic epileptiform discharges: a pattern of periodic short-interval diffuse discharges, periodic long-interval diffuse discharges, and suppression-burst patterns,
- eye movement: slow and occasionally sharp signals that occur during eye movement,
- artifacts: a recorded electrical activity that is not of cerebral origin, including physiologic artifacts,
- background: all other data that do not fall in the five classes above.

Overall, 518 sessions were annotated. No more than one session was included from a subject. 113453 segments were extracted. However, around 65% of the segments were annotated as background activities.

### 2.2.2 *Public Intracranial IED Dataset*

The public intracranial IED dataset (https://isarg.fel.cvut.cz/downloads/spike-detector/) consists of recordings from seven patients with refractory epilepsy who underwent invasive exploration as part of presurgical examination [1]. The number of electrodes is

different among subjects (the median number of electrodes is 65). However, to avoid human labelling bias, only 15 electrodes with the highest spike rate were selected from each dataset. Five-minute iEEG recordings of each subject were analysed and made publicly available. Three neurophysiologists with at least 10-year experience reviewed the data and labelled the spikes. Each spike is labeled as either obvious or ambiguous. Waveforms associated with hesitation and/or doubts are classified as ambiguous. Obvious and ambiguous spikes on which two or more readers agreed are selected as gold standard spikes (GS). Table 2.1 presents the patient characteristics (sex, age, number of GS, epilepsy location, seizure duration, and pathology) for each of the seven patients.

Table 2.1: Summary of patients' information used in the public IED dataset, adopted from [1].

| Patient | Age (sex) | Num. of GS | Epilepsy location | Seizure duration (years) | Pathology |
|---|---|---|---|---|---|
| 1 | 24 (Female) | 613 | Right frontal | 6 | No abnormality |
| 2 | 37 (Female) | 453 | Left frontal | 28 | No abnormality |
| 3 | 17 (Female) | 21 | Left temporal | 2 | FCD Ib |
| 4 | 8 (Female) | 3 | Right frontal | 3 | FCD IIb |
| 5 | 14 (Female) | 319 | Right multilobar | 8 | HS, FCD Ia |
| 6 | 31 (Male) | 335 | Right frontal | 27 | FCD IIb |
| 7 | 10 (Male) | 212 | Left multilobar | 3 | FCD IIa |

***GS:Gold Standard Spike; FCD: Focal Cortical Dysplasia; HS: Hippocampal Sclerosis.

### 2.2.3 *Private IED Dataset*

Table 2.2 summarizes the private datasets used in some previous studies. A research group led by Professor Saeid Sanei has been working on the concurrent scalp and intracranial recordings [8, 9, 15, 20, 24, 25]. We use the same dataset in our research. The dataset is described in Section 3.1.

Another most widely used IED dataset was recorded from over the scalp from 554 subjects (84 epileptic patients with annotated IEDs and 461 non-epileptic EEGs) at the Massachusetts General Hospital (MGH), Boston, USA [26, 27]. Nineteen (19) channels, according to the 10-20 international electrode system, were employed for recording.

In [13, 28], the researchers used an sEEG dataset recorded from 50 patients (24 males and 26 females) with childhood epilepsy with centro-temporal spikes at the Department of Pediatrics, Juntendo University Nerima Hospital. For recording the sEEG, 16 electrodes were used in accordance with the 10–20 international system.

Table 2.2: Some Datasets Used in IED Detection Studies.

| Study source | Recording modality | Num. of channels | Num. of subjects | Num. of IEDs |
|---|---|---|---|---|
| [29] | sEEG | 18 | 22 | 684 |
| [30] | sEEG | 32 | 19 | 1491 |
| [31, 32] | sEEG | - | 23 | 723 |
| [33] | sEEG | 21 | 7 | 4850 |
| [10] | sEEG | 19 | 7 | 1442 |
| [34] | sEEG | 19 | 5 | 7500 |
| [26, 27, 35] | sEEG | 19 | 50/545 | 8929/14170 |
| [13, 28] | sEEG | - | 50 | 15899/16008 |
| [19] | iEEG | 4 | 46 | 13959 |
| [8, 9, 15, 20, 24, 25] | Concurrent sEEG and iEEG | 20 scalp and 12 FO electrodes | 25/24/18 | 7831/6609/2776 |

***sEEG: Scalp EEG; iEEG: Intracranial EEG; FO: Foramen Ovale.

## 2.3 IED DETECTION METHODS

Different types of IED detection techniques have been introduced in the literature. They generally fall into six groups: (1) template matching, (2) feature representation (mimetic, TF, and nonlinear features), (3) matrix decomposition, (4) tensor factorisation, (5) neural networks, and (6) mapping sEEG to iEEG.

### 2.3.1 *Template Matching*

In the template matching method, a single or a number of templates is manually extracted from the training dataset. Then, the waveforms whose similarity to the template database exceeds the pre-defined threshold value are detected as the IED segments. Template matching was first used for seizure detection in 1972 [36]. Afterward, this method has been broadly employed for IED detection [31, 32, 35, 37]. To measure the similarity between the waveform and the template, correlation [31, 32, 37, 38] and Euclidean distance [35, 39, 40] were measured in most studies. However, other similarity measures such as mean square error [41] can be used.

### 2.3.2 *Feature Representation*

Feature representation has been used for IED detection for the past three decades. In this approach, different types of features are extracted from the IED segments and fed into a classifier or ensembled classifiers to classify the IEDs and non-IEDs. We

divide this group into three sub-groups: mimetic features, frequency or TF features, and nonlinear features.

#### 2.3.2.1 *Mimetic features*

As it is evident from its name, in this feature representation method, the aim is to mimic the visual analysis process. A number of features based on some characteristics used by expert neurologists to distinguish IEDs from non-IEDs are extracted and used for automatic IED detection. These features can be based on morphological characteristics and/or context information. This technique first divides the waveforms into two half-waves using peak detection algorithms. Then, parameters, such as amplitude, sharpness, slope, and duration, of the wave and its constituent half-waves are calculated and used for IED detection [11, 42–46]. For distinguishing an IED segment from background activity, the calculated parameters are compared against a threshold in the same way as those measured from the background activity. The threshold is selected based on the visual analysis and needs to be acceptable physiologically.

In the past decade, the mimetic technique has been used in multi-stages approaches [47, 48]. In this manner, all possible IED candidates are first detected. Then, the mimetic features are extracted from the detected candidates and fed into a classifier to be classified as the IED or non-IED segments. In [47], the k-point nonlinear energy operator was employed to detect all possible spike candidates. In [48], the authors applied template matching as the candidate spike detection algorithm. In both studies, morphological characteristics such as amplitude, duration, and slop of half-waves of spikes were extracted and classified.

This method has some limitations. The first limitation of this method is that it is not able to distinguish IEDs from transients or artefacts with similar morphology, such as extracerebral potentials from muscles, eyes, heart, and electrodes. The second limitation is that it ignores the IED spike shape variations across subjects, ages, and even trials. The morphological characteristics of IEDs change with age [49]. IEDs have higher amplitudes, sharper peaks, larger slopes, shorter durations, larger slow-wave areas, and wider distributions in children. On the other hand, these morphological characteristics dwindle, and the IEDs become more lateralized with increasing age.

#### 2.3.2.2 *TF features*

TF representation separates the main components of IEDs from the background activity. TF features have been extensively used in EEG signal processing and IED detection. Wavelet transform [50–54] is the most popular technique for TF feature extraction. However, other techniques such as Walsh transform [55], fast Fourier transform [56], spectrogram [8], and Hilbert transforms [57] have been employed. Spectrogram is a way of representing the signal strength over time at various frequencies. It can be generated

by Fourier transform or wavelet transform. Spyrou et al. applied short-term Fourier transform (STFT) and obtained its magnitude squared as the spectrogram features [8].

#### 2.3.2.3 *Nonlinear features*

A few studies have extracted nonlinear features for detecting IEDs/spikes. Among nonlinear features, the nonlinear energy operator is the most popular technique for identifying IEDs/spikes [54, 58]. In [59], the authors proposed a method based on the scale-dependent Lyapunov exponent for IED detection. In [54], apart from applying a nonlinear energy operator, the authors extracted waveform morphological features, TF features obtained using wavelet, and spectrogram features.

### 2.3.3 *Matrix Decomposition*

Matrix decomposition is referred to any method that decomposes the main signal matrix into sub-matrices. Among matrix decomposition techniques, singular value decomposition (SVD), sparse representation, independent component analysis (ICA), and principal component analysis (PCA) have been used in IED detection studies. In [60], the authors proposed a two-step method based on SVD and dipole source analysis. In the first step, SVD was employed to decompose the signals into uncorrelated sources. Then, a single dipole model was applied for dipole source analysis.

In sparse representation, the signal is considered to be sparse. The aim is to decompose a signal into a dictionary matrix and a sparse vector with a few non-zero components. The train of spikes emitted from individual neurons in the brain can be considered sparse in some domains, such as time and space. One of the interesting characteristics of IEDs is their sparsity in the time domain. Quite recently, Jiang et al. proposed a multi-component dictionary-based sparse representation for detecting EEG epileptic spikes [61].

ICA and PCA are generally performed for blind signal separation. Blind source separation aims to separate source signals from their mixtures. PCA, exploiting the lack of correlation between the components, was employed to detect epileptiform activities [62]. After applying PCA for dimensionality reduction and the decorrelation of components, wavelet transform was combined with approximate entropy to distinguish epileptic EEG from the normal one. As a powerful signal processing tool, ICA has been employed for IED detection [63–66]. In [65], the authors proposed an ICA-based method for tracking and detecting epileptiform activities. In [66], a model based on ICA was proposed to discriminate the epileptiform activities from eye blink artefacts.

### 2.3.4 *Tensor Decomposition*

In contrast to a matrix that has two dimensions, a tensor can have many dimensions. This provides an opportunity to analyse different domains of data such as time, frequency, channel, and trial together. Different methods have been provided for decomposing a tensor. More details about tensor decomposition methods are given in Chapter 3. However, among tensor decomposition methods, CANDECOMP/PARAFAC decomposition (CPD), Tucker decomposition (TD), and their nonnegative versions have been employed for IED detection.

For the first time, tensor decomposition was employed for IED detection from iEEG in 2015 [24]. They used the spectrogram method to obtain TF features. Then, they constructed a four-way tensor with channel, time, frequency, and segment dimensions and applied TD to decompose the tensor into spatial, temporal, and spectral factors. Finally, spatial factors were used as features for IED detection. Later, the same authors employed TD to detect IEDs from sEEG [15]. Quite recently, nonnegative TD (NTD) and nonnegative CPD (NCPD) with a novel tensor formulation were introduced to detect epileptic spikes [10]. At first, the authors decomposed the EEG channels using CWT. Then, they concatenated the channel, time, frequency, and segment of only epileptic spikes to construct a four-way tensor. The tensor was decomposed using NTD and NCPD to obtain the factor matrices and the core tensor. Finally, both epileptic and non-epileptic spikes from the training and test datasets were projected onto the factor matrices to extract the features for classification. The main differences between [15, 24] and [10] are: (1) NTD and NCPD were used in [10] while TD in [15, 24] and (2) only epileptic spikes were concatenated into the forth order of a tensor in [10] while all IED and non-IED segments in [15, 24]. Since different datasets were used in the mentioned papers, comparing them is difficult. In [10], only the sEEG was recorded and analysed, meaning that only scalp-visible spikes were detected. Meanwhile, sEEG and iEEG were simultaneously recorded in [15], and all scalp-visible and scalp-invisible IEDs were detected from the concurrent sEEG.

### 2.3.5 *Neural Networks*

Neural networks can be shallow or deep. Here, shallow neural networks (SNNs) refer to networks with a few dense layers. On the other hand, deep neural networks(DNNs) consist of numerous layers with complex structures and architectures for various functionality.

SNNs were used for IED detention for the first time in 1989 [67]. In most studies, SNN has been used as a classifier. This means some features are extracted from the EEG trials and applied to an SNN for classification. In [68–70], wavelet transform was em-

ployed for feature extraction and then an SNN for classification. Mimetic features were extracted and applied to an SNN for epileptiform spike detection as well [44, 71, 72]. After feature extraction from IEDs/spikes and classifying them using an SNN, some studies [70, 73–75] utilised a knowledge-based system using the expert's knowledge and context information to determine which of the detected candidates are epileptiform spikes and which ones not.

In a few studies, SNNs were applied to the raw EEG (without applying any feature extraction method) to detect epileptiform spikes [44, 70, 76, 77]. Feature-based SNNs outperform the raw EEG-based SNNs [44, 70]. Because of this limitation of SNNs, numerous studies have a feature extraction step before feeding the trials to an SNN.

SNNs are inefficient when the data are noisy and their size is large, whereas the DNNs come forward as a state-of-the-art method in artificial intelligence. DNNs are more complex than SNNs, and numerous (millions of) trainable parameters exist in a DNN. They are usually employed in an end-to-end approach. This means that the raw EEG signals are given to a DNN. The DNN learns its feature space representation and classifies the data samples. DNNs, as a popular approach in EEG signal processing, have been extensively used for IED detection.

DNNs were introduced to detect the IEDs from the sEEG [25] and iEEG [34] in 2016. The convolutional neural network (CNN) is the most popular DNN technique used for IED detection. The CNN enables capturing both temporal and spatial information. As a result, it is a powerful tool for processing two-dimensional datasets. Antoniades et al. developed a multiclass IED detection model [20]. They divided IEDs into different groups (classes) based on their morphology and spatial information captured through multichannel recording. Then, a multi-class CNN was employed to detect the IEDs from non-IEDs. They compared their proposed approach with the case of having all the IEDs as a single group of IEDs (binary CNN). They showed that the CNN distinguishes the IEDs from non-IEDs with higher accuracy even when there are IEDs with different morphology.

Abou Jaoude et al. developed a CNN-based method to detect the IEDs from the iEEG recorded using a single bipolar channel [19]. The authors averaged the CNN outputs of multiple bipolar channels to determine whether the test segment was related to an IED event or not. Thomas et al. employed one-dimensional (1D) CNN to detect IEDs from single-channel sEEG [26]. The IEDs were labeled separately for each channel, and the 1D CNN was trained to detect IEDs from a single channel. Finally, the channel outputs were combined to produce a single value used as an index to identify an IED segment. To detect epileptic spikes, Fukumori et al. developed a CNN model based on a bank of linear-phase finite impulse response filters, performing as bandpass filters that extract the biomarkers of IEDs without destroying the waveforms as the result of linear-phase condition [78].

The EEG may be presented as or converted to an image [17, 79]. Quon et al. employed STFT to obtain the spectrogram of IED and non-IED segments recorded using iEEG [17]. The spectrogram was considered as an image and applied to an image augmentation method. Then, the augmented data are fed to a CNN designed based on a residual neural network for IED detection.

Recurrent neural networks (RNNs) rank second among the most popular DNN techniques employed for IED detection. They are suitable for the analysis of time series and are able to capture only temporal information. Long short-term memory (LSTM) and gated recurrent units (GRU), the modified models of RNNs, have been used for IED detection. In [80], the authors employed LSTM to detect epileptiform spikes and high-frequency oscillations from the iEEG [80]. Their model was not an end-to-end framework, and they applied the power spectrogram of different frequency bands as features to the network. Quite recently, an LSTM-based model with a self-attention mechanism was proposed to detect epileptic spikes that were not temporally aligned [13].

In a study, researchers compared a discrete wavelet transform-based feature extraction approach with an end-to-end framework approach (fully data-driven approach) [81]. In the first approach, the extracted features are given to both traditional classifiers and DNNs for IED detection. On the other hand, in the second approach, a convolutional layer is employed as a bandpass filter to estimate the frequency bands of interest. Then, the feature space of the convolutional layer is passed through DNNs (LSTM, GRU, and CNN) for IED detection. The obtained results using both approaches are comparable.

CNN and RNN-based networks have been combined for IED detection [14, 82]. In [82], five different architectures, namely 1D CNN, 2D CNN, LSTM, combined 1D CNN and LSTM, and combined 2D CNN and LSTM, were implemented. In 1D CNN, the filters were defined for each channel separately involving temporal information only, while, in 2D CNN, the filters had two dimensions involving temporal and spatial information. 2D CNN outperforms 1D ones. As a result, considering both temporal and spatial information in the convolutional layers boosts the model's performance. However, the model did not achieve high sensitivity, although the authors trained 346 different DNNs with various architectures to achieve the best result. The length of segments given to DNNs was 2 seconds which is unnecessarily long. The length of IED segments in most IED detection studies [8, 10, 83] is less than or around 500 ms since the length of an IED waveform is not longer than 200 ms [84]. Therefore, using too long IED segments leads the model to fit over non-ED segments.

Apart from standard CNNs, GRU, LSTM, other DNN techniques such as deep belief network [30], region-CNN [85], and generative adversarial network (GAN) [86] have also been proposed for IED detection. The developed deep belief network [30] was compared with a deep autoencoder and traditional classifiers, which significantly out-

performed them. The developed GAN-based technique [86] was compared with only traditional classifiers and outperformed them. This comparison in [86] is unfair since DNN-based techniques are powerful and complex, while traditional classifiers are not.

### 2.3.6 *Mapping sEEG to iEEG*

sEEG channels are far away from IED sources and hence fail in capturing a large proportion of IEDs [6, 7]. To improve the performance of sEEG in IED detection, some studies mapped sEEG to iEEG [9, 83, 87].

Spyrou and Sanei proposed a method based on a learning coupled dictionary between multimodal datasets [83]. The algorithm generally aims to reconstruct the data of one mode by using the data of another mode. They employed their method to reconstruct iEEG from the concurrent sEEG.

In an effective approach, Antoniades et al. analysed the concurrent sEEG and iEEG recordings and estimated the intracranial samples using scalp recordings via DNNs [9]. The authors developed an autoencoder (AE) to map the sEEG to iEEG. Then, a CNN was applied to the estimated iEEG to detect the IEDs. They compared the performance of their proposed method with those using sEEG only. Their mapping model significantly outperformed other popular methods.

Recently, Took et al. proposed a method based on transfer learning [87]. In transfer learning, the first layers of a trained DNN are frozen (not trained), and its last layers are adapted to the problem at hand. In [87], the authors used the same mapping model proposed in [9] for estimating the iEEG. Then, transfer learning was applied. The estimated iEEG passed through the frozen convolutional layers trained using the actual iEEG in [20]. Finally, an LSTM layer was employed to exploit inter-trial correlations.

## 2.4 PERFORMANCE METRICS

Different metrics have been utilised to evaluate the performance of an IED detection system. These are accuracy (ACC), sensitivity (SEN), specificity (SPC), F1-score (F1-S), false positive per minute (FP/min), and area under the receiver operating characteristic curve (AUC). Accuracy presents how accurately the IEDs and non-IEDs are detected. SEN shows the ability of a system to detect IEDs correctly. SPC shows the ability of a system to detect non-IEDs correctly. FP/min illustrates the number of non-IED segments recognized as IED segments in a minute. AUC measures the entire two-dimensional area underneath the entire receiver operating characteristic (ROC) curve, which is a probability curve showing the performance of a classifier for different thresh-

olds. The ROC curve is a plot of SEN versus 1-SPC. ACC, SEN, and SPC are defined as follows:

$$\begin{aligned} \text{ACC} &= \frac{TP+TN}{TP+FP+TN+FN} \times 100\%, \\ \text{SEN} &= \frac{TP}{TP+FN} \times 100\%, \\ \text{SPC} &= \frac{TN}{TN+FP} \times 100\%, \\ \text{F1-S} &= \frac{2TP}{2TP+FP+FN}, \end{aligned} \tag{2.1}$$

where TP is the number of IED samples classified correctly as IED, TN represents the number of non-IED samples recognized accurately as non-IED samples, FP indicates the number of non-IED samples detected incorrectly as IED samples, and FN indicates the number of IED samples categorized wrongly in the non-IED class.

However, various authors have applied their proposed methods to their own datasets. No public IED dataset can be used as a benchmark. Therefore, here we do not intend to compare the performance of those methods.

## 2.5 CONCLUSION

Detection of IEDs is of great importance in managing and monitoring epilepsy. Therefore, automatic detection of IEDs is clinically and scientifically important. First, this chapter describes public and private iEEG and sEEG datasets. Then, a review of the literature describing IED analysis and detection is provided. Different methods have been developed for IED detection. The proposed methods are grouped into six groups: (1) template matching, (2) feature representation (mimetic, TF, and nonlinear features), (3) matrix decomposition, (4) tensor factorisation, (5) neural networks, and (6) estimating the iEEG from the sEEG. Finally, the performance metrics used in IED detection studies are explained.

# 3

# REAL-TIME DETECTION OF SCALP-INVISIBLE IEDS FROM EEG

Multiway analysis (e.g., tensor decomposition) provides an opportunity to simultaneously analyze multi-aspect data (e.g., time, space, frequency, segment, subject, and morphology). Tensor factorisation and multiview classification have recently attracted the attention of researchers in biomedical signal processing [88, 89] and IED detection [10, 15]. Several electrodes may capture the IED signatures, meaning the EEG signals are spatially and temporally correlated. Therefore, multi-channel and multi-trial (multi-segment) EEG processing using tensor decomposition methods is expected to be effective in IED detection.

Most of the previous studies are based on aligning and averaging IEDs. The main drawback of these methods is that the IED positions are aligned in all segments. However, in the real world, the aim is to detect the IEDs from ongoing EEG recordings in which there is no alignment among the IED segments. In addition, in places where only the scalp-EEG is recorded and analyzed [53, 54], the scalp-invisible IEDs are not considered. Models provided for the detection of IEDs scored by scalp-EEG are inadequate since they are unable to detect the scalp-invisible IEDs, thereby being biased to detect only a subset of IEDs that are visible over the scalp. Therefore, exploiting simultaneous iEEG and sEEG recordings in designing an algorithm to detect the scalp-invisible IEDs from sEEG is of great interest. Only very few studies have been carried out to investigate scalp IEDs from simultaneous sEEG and iEEG recordings [8, 9, 15].

This chapter aims to detect the IEDs from ongoing concurrent sEEG and iEEG signals. It should be noted that here both scalp-visible and scalp-invisible IEDs are included. Here, we develop two tensor-based methods: temporal component analysis (TCA) and spatial component analysis (SCA). The proposed methods are employed to detect the IEDs from the ongoing EEG data. Furthermore, both SCA and TCA are combined, referred to as TCA-SCA, to improve performance.

First, this chapter describes the dataset used in this thesis. Then, popular tensor decomposition methods are explained. Afterward, the proposed methods, TCA, SCA, and TCA-SCA, are presented. Next, the results of applying the proposed methods to both sEEG and iEEG recordings are shown. Finally, the last section concludes the chapter.

## 3.1 DATASET

The details about patient population, electrode placement, recording session, and IED scoring are given below.

### 3.1.1 *Patient Population*

The sEEG and iEEG signals of 18 epileptic subjects (11 males, seven females, average age 25.2 years, and age range 13–37 years) were simultaneously recorded at King's College Hospital London. This retrospective study is based on standard medical recordings from patients at King's College Hospital by following routine clinical ethical procedures. It did not need approval by the ethics committee.

Patients suffering from seizures arising from mesial temporal structures were submitted for telemetry recording with FO when history, interictal scalp EEG, neuroimaging, and neuropsychological studies could not confidently determine the side of seizure onset or there were doubts about a lateral temporal or extra-temporal seizure onset. In 10 patients, the seizure onset was identified within mesial temporal structures preceding in at least 2 ms the scalp changes, while in eight patients, it was located outside the mesial temporal region (lateral temporal).

### 3.1.2 *Electrode Placement and Recording Session*

Twenty (20) standard silver chloride electrodes were used for recording sEEG, placed on the scalp according to the "Maudsley" electrode placement system. The iEEG was recorded using 12 intracranial multi-contact FO electrodes consisting of a couple of 6-electrode bundles. The FO electrodes were inserted through the patients' FOs under general anaesthesia and fluoroscopic control. Each individual electrode was made up of a 0.1 mm fully insulated stainless steel wire. The recording contacts of the three deepest electrodes were 3 mm long, and those of the most superficial electrodes were 5 mm long. The distance between contiguous electrodes was 10 mm except for the two most superficial electrodes, whose interelectrode distance was 15 mm. For each electrode bundle, the two deepest electrodes were placed next to medial temporal structures and are generically called 'deep FO electrodes'. The two most superficial electrodes were laid at or close to the FO of the sphenoid bone and are generically called 'superficial' FO electrodes. The position of the FO bundles was confirmed with post-insertion radiography, shown in Figure 3.1.

The data were recorded at the sampling rate of 200 Hz and filtered by a bandpass filter with cutoff frequencies of 0.3 and 70 Hz. The system input range was 2 mV, and the data were digitised with a 12-bit analog-to-digital converter (an amplitude

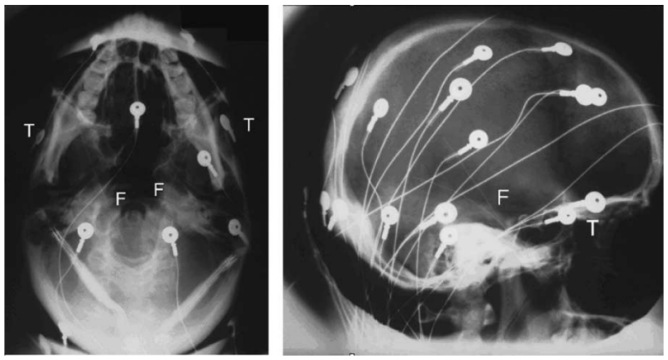



Figure 3.1: Lateral and basal X-rays showing scalp and intracranial FO electrodes. T, scalp anterior temporal electrodes; F, intracranial FO bundles.

resolution of 0.488 mV). Both sEEG and iEEG were recorded with respect to Pz as a common reference. A period of 20 min of simultaneous sEEG and iEEG was transcribed onto a digital file. Such interictal recordings showed interictal discharges and no signs of post-ictal slowing or flattening.

### 3.1.3 *IED Scoring*

An expert epileptologist scored the IEDs from the iEEG based on the morphology and spatial distribution of the observed waveforms. Interictal events location and morphology were evaluated in the background context. The IEDs were evaluated in terms of morphology and distribution, following the standard definitions for epileptiform pattern, spike and sharp wave of the International Federation of Clinical Neurophysiology (A glossary of terms most commonly used by clinical EEGers and proposal for the report form for the EEG findings) [4]. In addition, each IED was given a certainty score (1–9). Briefly, each IED is classified into one of the following groups: (I) scalp-invisible IED, (II) scalp-visible IED by considering the concurrent iEEG, and (III) scalp-visible IED without considering the concurrent iEEG. Examples of IED and non-IED segments are shown in Figure 3.2. In the scalp-invisible IED segment, there is no sign of epileptiform discharges in the scalp channels, while the FO channels capture the IED waveforms. In the "scalp-visible IED by considering the concurrent iEEG," a weak IED waveform can be detected in the scalp channels by referencing to the concurrent iEEG. In the "scalp-visible IED without considering the concurrent iEEG," the epileptiform discharges are individually observable from the scalp channels.

However, in essence, the IEDs were scored by considering the followings:

1. Nonphysiological artefact: instrumental, electrode, environmental, and quantization artefacts,

2. Physiological artefact: eye movement, electrocardiogram, muscle (lateral rectus, frontalis, temporalis, occipital), and blink artefacts,

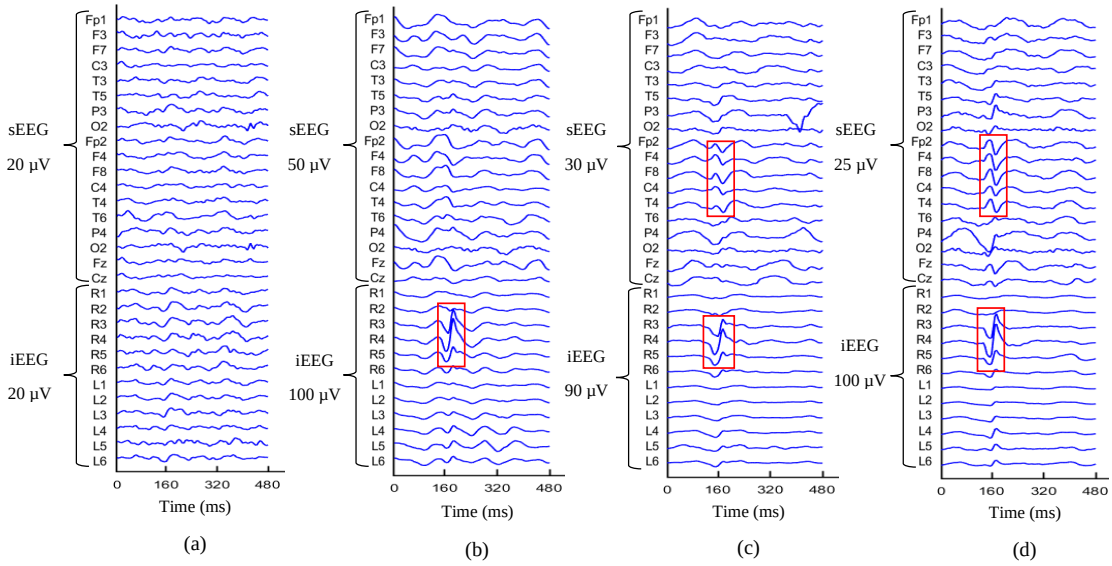



Figure 3.2: Samples of non-IED and IED segments; (a) non-IED, (b) scalp-invisible IED, (c) scalp-visible IED by referencing to the concurrent iEEG, and (d) scalp-visible IED without considering the concurrent iEEG. The interest areas are pointed by red rectangles. The IEDs start at 160 ms. Channels R1 to R6 and L1 to L6 correspond respectively to the FO channels of right and left hemispheres.

3. Low amplitude irregularity/physiologic "sharpened/spiky" activities (vertex waves, K-complexes),
4. Irregularity: IEDs barely distinguishable from the background activity and restricted to 1–2 channels,
5. Sharp wave (restricted to at least three channels),
6. Broad distribution sharp wave (> three channels),
7. Spike (restricted to at least three channels),
8. Broad distribution spike (> three channels),
9. Spike or sharp wave.

## 3.2 MULTI-WAY ANALYSIS

Tensor decomposition was introduced by Hitchcock in 1927 [90, 91] and the multi-way model by Cattell in 1944 [92, 93]. These concepts did not receive much attention until the work of Tucker in the 1960s [94–96] and Harshman [97] and Carroll and Chang [98] in 1970. All of these were introduced in psychometrics literature. However, now, it has found applications in different spheres, from chemometrics [99–101] to computational biology [102, 103], from image analysis [104, 105] to signal processing [106, 107].

A tensor is a multidimensional or multi-way array. The order of a tensor is equivalent to the number of its dimensions, also known as ways or modes. For instance, a tensor with only one way is a first-order tensor (vector), and with two dimensions is a two-way tensor (matrix). Tensors of order three or higher are called higher-order tensors. When a subset of indices in a tensor is fixed, sub-tensors or subarrays are formed. Like matrices with rows and columns, tensors have fibres and slices. A fibre is defined by fixing every index but one. Slices are two-dimensional sections of a tensor, defined by fixing all but two indices.

### 3.2.1 *Tensor Notation and Production*

The notation used in this thesis has been adopted from Ref. [108]. Lowercase letters, e.g., $a$, denote scalars. Boldface lowercase letters, e.g., $\mathbf{a}$, represent vectors. Boldface capital letters, e.g., $\mathbf{A}$, denote matrices. Boldface Euler script letters, e.g., $\boldsymbol{\mathcal{X}}$, denote higher-order tensors. $\mathbf{a}_j$ denote the $j$-th column of matrix $\mathbf{A}$. The symbols '$\circ$' and '$*$' represent respectively the vector outer product and the Hadamard product. $\odot$ and '$\otimes$' denote respectively the Khatri-Rao and Kronecker products. Here, important operations and notations used in tensor factorisation are explained.

#### 3.2.1.1 *Matricization*

Matricization, also known as unfolding or flattening, means transforming a tensor into a matrix. It is the process of reordering the elements of a $V$-way array into a matrix. The mode-v matricization of a tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_V}$ is shown by $\mathbf{X}_{(v)}$. More information about how to maricize a tensor is given in Ref. [109].

#### 3.2.1.2 *Matrix Hadamard product*

The Hadamard product is an elementwise matrix product. The Hadamard product of $\mathbf{A}$ and $\mathbf{B}$ of size $I \times J$ shown by $\mathbf{A} * \mathbf{B}$ is calculated as follows:

$$\mathbf{A} * \mathbf{B} = \begin{bmatrix} a_{11}b_{11} & a_{12}b_{12} & \cdots & a_{1J}b_{1J} \\ a_{21}b_{21} & a_{22}b_{22} & \cdots & a_{2J}b_{2J} \\ \vdots & \vdots & \ddots & \vdots \\ a_{I1}b_{I1} & a_{I2}b_{I2} & \cdots & a_{IJ}b_{IJ} \end{bmatrix} \tag{3.1}$$

#### 3.2.1.3 *Matrix Kronecker product*

The Kronecker product of matrices $\mathbf{A} \in \mathbb{R}^{I \times J}$ and $\mathbf{B} \in \mathbb{R}^{K \times L}$ shown by $\mathbf{A} \otimes \mathbf{B}$ is defined as follows:

$$\mathbf{A} \otimes \mathbf{B} = \begin{bmatrix} a_{11}\mathbf{B} & a_{12}\mathbf{B} & \cdots & a_{1J}\mathbf{B} \\ a_{21}\mathbf{B} & a_{22}\mathbf{B} & \cdots & a_{2J}\mathbf{B} \\ \vdots & \vdots & \ddots & \vdots \\ a_{I1}\mathbf{B} & a_{I2}\mathbf{B} & \cdots & a_{IJ}\mathbf{B} \end{bmatrix} \tag{3.2}$$

#### 3.2.1.4 *Matrix Khatri-Rao product*

The Khatri–Rao product [200] is the "matching columnwise" Kronecker product. The Khatri–Rao product of matrices $\mathbf{A} \in \mathbb{R}^{I \times K}$ and $\mathbf{B} \in \mathbb{R}^{J \times K}$ shown by $\mathbf{A} \odot \mathbf{B}$ is defined as follows:

$$\mathbf{A} \odot \mathbf{B} = \begin{bmatrix} a_1 \otimes b_1 & a_2 \otimes b_2 & \cdots & a_K \otimes b_K \end{bmatrix} \tag{3.3}$$

The dimension of resulted matrix is $(IJ) \times K$.

#### 3.2.1.5 *Rank-one tensors*

A V-way tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_V}$ is rank one if it can be written as the outer product of $V$ vectors as follows:

$$\mathcal{X} = a^{(1)} \circ a^{(2)} \circ \cdots \circ a^{(V)}. \tag{3.4}$$

#### 3.2.1.6 *Tensor multiplication: The v-mode product*

Tensors can be multiplied together. Here, only v-mode multiplication is explained. The v-mode product means multiplying a tensor by a matrix (or a vector) in mode v.

The v-mode product of a tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_V}$ with a matrix $\mathbf{U} \in \mathbb{R}^{J \times I_v}$ is shown by $\mathcal{X} \times_v \mathbf{U}$. The size of multiplication result is $I_1 \times \cdots \times I_{v-1} \times J \times I_{v+1} \times \cdots \times I_V$. Each mode-v fiber is multiplied by the matrix $\mathbf{U}$. Elementwise, we have

$$(\mathcal{X} \times_v \mathbf{U})_{i_1 \cdots i_{v-1} j i_{v+1} \cdots i_V} = \sum_{i_v=1}^{I_v} x_{i_1 i_2 \cdots i_V} u_{j i_v}. \tag{3.5}$$

#### 3.2.1.7 *Tensor rank*

The rank of a tensor $\mathcal{X}$ is defined as the smallest number of rank-one tensors generating $\mathcal{X}$ as their sum [90]. However, the estimation of tensor rank is under question; in fact, the problem is NP-hard [110]. Generally, the tensor rank is higher than the ranks of slab

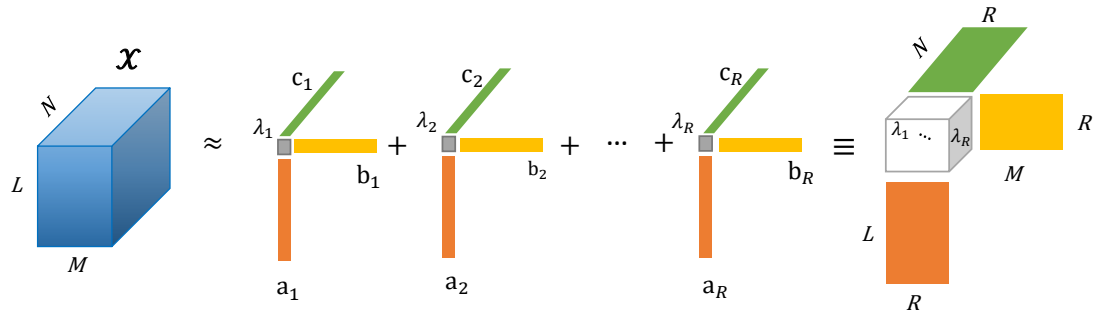



Figure 3.3: CANDECOMP/PARAFAC decomposition of a three-way array.

matrices. Therefore, tensors have found many applications in source decomposition, especially where the corresponding systems are underdetermined, i.e., the number of sources exceeds the number of sensors.

### 3.2.2 *Common Tensor Decomposition Techniques*

There are different methods for decomposing a tensor [108]. The most popular methods which have been employed for IED detection are CPD and TD. Here, CPD and TD are explained.

#### 3.2.2.1 *CANDECOMP/PARAFAC decomposition*

CPD decomposes a tensor into the sum of rank-one components. Suppose we are given a three-way tensor $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{L \times M \times N}$. According to CPD problem, the tensor $\boldsymbol{\mathcal{X}}$ can be formulated as:

$$\boldsymbol{\mathcal{X}} \approx \sum_{r=1}^{R} \mathbf{a}_r \circ \mathbf{b}_r \circ \mathbf{c}_r, \tag{3.6}$$

where the symbol '$\circ$' represents the vector outer product, $R$ is a positive integer and $a_r \in \mathbb{R}^L$, $b_r \in \mathbb{R}^M$, and $c_r \in \mathbb{R}^N$ for $r = 1, \ldots, R$. It is often useful to assume that the components are normalized to length one with the weights put into the vector $\boldsymbol{\lambda} \in \mathbb{R}^R$. Therefore, (3.6) is reformulated to

$$\boldsymbol{\mathcal{X}} \approx \sum_{r=1}^{R} \lambda_r \, \mathbf{a}_r \circ \mathbf{b}_r \circ \mathbf{c}_r. \tag{3.7}$$

The factor matrices are constructed from the combination of the rank-one tensors, i.e., $\mathbf{A} = [\mathbf{a}_1 \ldots \mathbf{a}_R]$. Following the "Kruskal operator" [109], (3.7) can be modified to

$$\boldsymbol{\mathcal{X}} \approx [\![\boldsymbol{\lambda}; \mathbf{A}, \mathbf{B}, \mathbf{C}]\!] \equiv \sum_{r=1}^{R} \lambda_r \, \mathbf{a}_r \circ \mathbf{b}_r \circ \mathbf{c}_r. \tag{3.8}$$

where $\mathbf{A} \in \mathbb{R}^{L \times R}$ and $\mathbf{B} \in \mathbb{R}^{M \times R}$, and $\mathbf{C} \in \mathbb{R}^{N \times R}$ are factor matrices. Figure 3.3 shows a diagram of CPD of a three-way array.

To compute factor matrices with $R$ components that best approximates $\boldsymbol{\mathcal{X}}$, (3.8) can be formulated as an optimization problem:

$$\min_{\hat{\boldsymbol{\mathcal{X}}}} \left\| \boldsymbol{\mathcal{X}} - \hat{\boldsymbol{\mathcal{X}}} \right\| \quad \text{where} \quad \hat{\boldsymbol{\mathcal{X}}} = \sum_{r=1}^{R} \lambda_r \, \mathbf{a}_r \circ \mathbf{b}_r \circ \mathbf{c}_r = [\![ \boldsymbol{\lambda}; \mathbf{A}, \mathbf{B}, \mathbf{C} ]\!]. \tag{3.9}$$

There are different algorithms to compute CPD of a given tensor. Here, we will take a brief look at the alternating least squares (ALS) algorithm, the most popular one. ALS was proposed in the original papers by Harshman [97] and Carroll and Chang [98]. This algorithm reduces the problem to a linear least-squares problem by fixing all factor matrices but one. This procedure is applied for every matrix repeatedly until some convergence criterion is satisfied.

The matricized form of (3.8) (one per mode) can be written as

$$\begin{aligned} \mathbf{X}_{(1)} &\approx \mathbf{A}(\mathbf{C} \odot \mathbf{B})^T, \\ \mathbf{X}_{(2)} &\approx \mathbf{B}(\mathbf{C} \odot \mathbf{A})^T, \\ \mathbf{X}_{(3)} &\approx \mathbf{C}(\mathbf{B} \odot \mathbf{A})^T. \end{aligned} \tag{3.10}$$

From (3.10), the above minimization problem (3.9) can be rewritten as:

$$\begin{aligned} \hat{\mathbf{A}} &\leftarrow \arg \min_{\hat{\mathbf{A}}} \left\| \mathbf{X}_{(1)} - \hat{\mathbf{A}}(\mathbf{C} \odot \mathbf{B})^T \right\|_F \\ \hat{\mathbf{B}} &\leftarrow \arg \min_{\hat{\mathbf{B}}} \left\| \mathbf{X}_{(2)} - \hat{\mathbf{B}}(\mathbf{C} \odot \mathbf{A})^T \right\|_F \\ \hat{\mathbf{C}} &\leftarrow \arg \min_{\hat{\mathbf{C}}} \left\| \mathbf{X}_{(3)} - \hat{\mathbf{C}}(\mathbf{B} \odot \mathbf{A})^T \right\|_F , \end{aligned} \tag{3.11}$$

where $\hat{\mathbf{A}} = \mathbf{A}.\mathrm{diag}(\boldsymbol{\lambda})$, $\hat{\mathbf{B}} = \mathbf{B}.\mathrm{diag}(\boldsymbol{\lambda})$, $\hat{\mathbf{C}} = \mathbf{C}.\mathrm{diag}(\boldsymbol{\lambda})$, and $\|.\|$ refers to Frobenius norm. Finally, the optimal solution is obtained by

$$\begin{aligned} \hat{\mathbf{A}} &= \mathbf{X}_{(1)}[(\mathbf{C} \odot \mathbf{B})^T]^\dagger = \mathbf{X}_{(1)}(\mathbf{C} \odot \mathbf{B})(\mathbf{C}^T\mathbf{C} * \mathbf{B}^T\mathbf{B})^\dagger, \\ \hat{\mathbf{B}} &= \mathbf{X}_{(2)}[(\mathbf{C} \odot \mathbf{A})^T]^\dagger = \mathbf{X}_{(2)}(\mathbf{C} \odot \mathbf{A})(\mathbf{C}^T\mathbf{C} * \mathbf{A}^T\mathbf{A})^\dagger, \\ \hat{\mathbf{C}} &= \mathbf{X}_{(3)}[(\mathbf{B} \odot \mathbf{A})^T]^\dagger = \mathbf{X}_{(3)}(\mathbf{B} \odot \mathbf{A})(\mathbf{B}^T\mathbf{B} * \mathbf{A}^T\mathbf{A})^\dagger. \end{aligned} \tag{3.12}$$

So far, we have focused on the three-way tensor. For a general $V$-way tensor $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_V}$, (3.8) is generalized to

$$\boldsymbol{\mathcal{X}} \approx [\![ \boldsymbol{\lambda}; \mathbf{A}^{(1)}, \mathbf{A}^{(2)}, \ldots, \mathbf{A}^{(V)} ]\!] \equiv \sum_{r=1}^{R} \lambda_r \, \mathbf{a}_r^{(1)} \circ \mathbf{a}_r^{(2)} \circ \cdots \circ \mathbf{a}_r^{(V)}. \tag{3.13}$$

Algorithm 3.1: ALS algorithm to compute a CPD with R components for a $V$-way tensor $\mathcal{X}$ of size $I_1 \times I_2 \times \cdots \times I_V$.

```
procedure CPD-ALS (𝒳, R)
    initialize A^(v) ∈ ℝ^(I_v × R),  v = 1,2,...,V
    repeat
        for v = 1,2,...,V do
            Z ← A^(1)ᵀA^(1) * ··· * A^(v−1)ᵀA^(v−1) * A^(v+1)ᵀA^(v+1) * ··· * A^(V)ᵀA^(V)
            A^(v) ← X_(v)(A^(V) ⊙ ··· ⊙ A^(v+1) ⊙ A^(v−1) ⊙ ··· ⊙ A^(1))Y†
            normalize columns of A^(v) (storing norms as λ)
        end for
    until the convergence criterion is satisfied or maximum iterations exhausted
    return λ, A^(1), A^(2), ..., A^(V)
end procedure
```

where $\boldsymbol{\lambda} \in \mathbb{R}^R$ and $\mathbf{A}^{(v)} \in \mathbb{R}^{I_v \times R}$ for $v = 1, 2, \ldots, V$. In this case, the mode-$v$ is matricized by

$$\mathbf{X}_{(v)} \approx \mathbf{A}^{(v)} \boldsymbol{\Lambda} (\mathbf{A}^{(V)} \odot \cdots \odot \mathbf{A}^{(v+1)} \odot \mathbf{A}^{(v-1)} \odot \cdots \odot \mathbf{A}^{(1)})^T, \tag{3.14}$$

where $\boldsymbol{\Lambda} = \text{diag}(\boldsymbol{\lambda})$.

The ALS algorithm for CPD of a V-way tensor is shown in Algorithm 3.1. It assumes that the number of components, R, of the CPD is specified. The factor matrices can be initialized in any way, such as randomly or by setting

$$\mathbf{A}^{(v)} = R \text{ leading left singular vectors of } \mathbf{X}_{(v)} \text{ for } n = 1, 2, \ldots, V$$

The reader is referred to [111] to see more discussion about initialization methods. The procedure repeats until convergence or reaching a maximum iteration number.

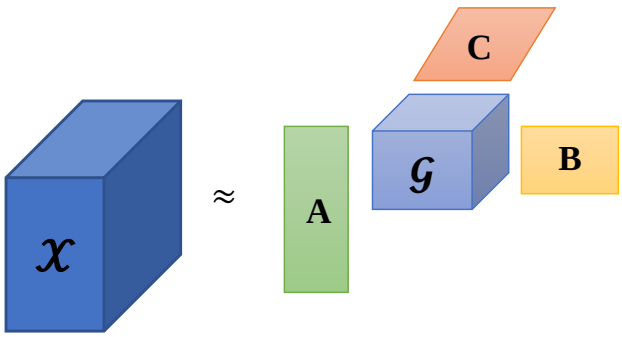



Figure 3.4: Tucker Decomposition of a Three-way Array.

#### 3.2.2.2 *Tucker decomposition*

Tucker [94] introduced TD in 1963, then modified it in the next years [95, 96]. TD decomposes a tensor into a core tensor multiplied by a matrix along each mode. In the three-way case where $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{L \times M \times N}$, the aim is to find a tensor $\hat{\boldsymbol{\mathcal{X}}} \in \mathbb{R}^{L \times M \times N}$, having $\text{rank}_1(\hat{\boldsymbol{\mathcal{X}}}) = P$, $\text{rank}_2(\hat{\boldsymbol{\mathcal{X}}}) = Q$, and $\text{rank}_3(\hat{\boldsymbol{\mathcal{X}}}) = R$, that minimizes the least-squares cost function

$$\begin{aligned} \min_{\hat{\boldsymbol{\mathcal{X}}}} \left\| \boldsymbol{\mathcal{X}} - \hat{\boldsymbol{\mathcal{X}}} \right\| \quad \text{with} \quad \hat{\boldsymbol{\mathcal{X}}} &= \sum_{p=1}^{P} \sum_{q=1}^{Q} \sum_{r=1}^{R} g_{pqr}\, \mathbf{a}_p \circ \mathbf{b}_q \circ \mathbf{c}_r \\ &= \boldsymbol{\mathcal{G}} \times_1 \mathbf{A} \times_2 \mathbf{B} \times_3 \mathbf{C} = [\![ \boldsymbol{\mathcal{G}}; \mathbf{A}, \mathbf{B}, \mathbf{C} ]\!]. \end{aligned} \tag{3.15}$$

where $\mathbf{A} \in \mathbb{R}^{L \times P}$, $\mathbf{B} \in \mathbb{R}^{M \times Q}$, and $\mathbf{C} \in \mathbb{R}^{N \times R}$, all with orthonormal columns, are factor matrices. $\boldsymbol{\mathcal{G}} \in \mathbb{R}^{P \times Q \times R}$ is the core tensor, and its entries present the level of interaction between the factor matrices. Note that $\times_i$ shows the $i$-th mode product. The schematic of TD is shown in Figure 3.4.

(3.15) is matricized as follows:

$$\begin{aligned} \hat{\mathbf{X}}_{(1)} &= \mathbf{A}\mathbf{G}_{(1)}(\mathbf{C} \otimes \mathbf{B})^T \\ \hat{\mathbf{X}}_{(2)} &= \mathbf{B}\mathbf{G}_{(2)}(\mathbf{C} \otimes \mathbf{A})^T \\ \hat{\mathbf{X}}_{(3)} &= \mathbf{C}\mathbf{G}_{(3)}(\mathbf{B} \otimes \mathbf{A})^T. \end{aligned} \tag{3.16}$$

We introduced the three-way tensor. However, TD can be generalized to a $V$-way tensor as

$$\boldsymbol{\mathcal{X}} = \boldsymbol{\mathcal{G}} \times_1 \mathbf{A}^{(1)} \times_2 \mathbf{A}^{(2)} \ldots \times_V \mathbf{A}^{(V)} = [\![ \boldsymbol{\mathcal{G}}; \mathbf{A}^{(1)}, \mathbf{A}^{(2)}, \ldots, \mathbf{A}^{(V)} ]\!], \tag{3.17}$$

and it can be matricized as

$$\hat{\mathbf{X}}_{(v)} = \mathbf{A}^{(v)}\mathbf{G}_{(v)}(\mathbf{A}^{(V)} \otimes \cdots \otimes \mathbf{A}^{(v+1)} \otimes \mathbf{A}^{(v-1)} \otimes \cdots \otimes \mathbf{A}^{(1)})^T \tag{3.18}$$

There are different techniques to perform TD. The most simple and popular ones are Higher Order Singular Value Decomposition (HOSVD) and Higher Order Orthogonal Iteration (HOOI).

HIGHER ORDER SINGULAR VALUE DECOMPOSITION: HOSVD is a convincing generalization of the matrix singular value decomposition [112]. The key idea behind the HOSVD is to find the components that best capture the variation in mode $v$ while not considering the other modes at each time instant [113]. The HOSVD algorithm is given in Algorithm 3.2.

Algorithm 3.2: HOSVD algorithm to compute a rank-$(R_1, R_2, \ldots, R_V)$ TD for a $V$-way tensor $\boldsymbol{\mathcal{X}}$ of size $I_1 \times I_2 \times \cdots \times I_V$.

```
procedure HOSVD (X, R_1, R_2, ..., R_V)
    for v = 1, 2, ..., V do
        A^(v) = R_v leading left singular vectors of X_(v)
    end for
    G ← X ×_1 A^(1)T ×_2 A^(2)T ... ×_V A^(V)T
    return G, A^(1), A^(2), ..., A^(V)
end procedure
```

Algorithm 3.3: ALS algorithm to compute a rank-$(R_1, R_2, \ldots, R_V)$ TD for a $V$-way tensor $\boldsymbol{\mathcal{X}}$ of size $I_1 \times I_2 \times \cdots \times I_V$, also known as the HOOI.

```
procedure HOOI (X, R_1, R_2, ..., R_V)
    initialize A^(v) ∈ R^(I_v × R),  v = 1, 2, ..., V, using HOSVD
    repeat
        for v = 1, 2, ..., V do
            Z ← X ×_1 A^(1)T ··· ×_(v-1) A^(v-1)T ×_(v+1) A^(v+1)T ··· ×_V A^(V)T
            A^(v) = R_v leading left singular vectors of Z_(v)
        end for
    until convergence criterion is satisfied or maximum iterations exhausted
    G ← X ×_1 A^(1)T ×_2 A^(2)T ... ×_V A^(V)T
    return G, A^(1), A^(2), ..., A^(V)
end procedure
```

HIGHER ORDER ORTHOGONAL ITERATION: HOOI proposed in [112] is more efficient than HOSVD for calculating the factor matrices. In fact, the HOOI is an ALS algorithm that uses the outcome of performing HOSVD on a tensor as a starting point for initializing the factor matrices. The HOOI algorithm is shown in Algorithm 3.3.

## 3.3 SPATIAL AND TEMPORAL COMPONENT ANALYSIS FOR IED DETECTION

The IEDs recorded from a subject share in their morphology and locations within the brain. Therefore, they may share some features. In contrast, normal brain activities or artefacts are entirely independent of these discharges. As a result, in our methods, all IEDs are concatenated into a three-way tensor with the dimension of time, channel, and IED segment. It should be noted that the idea of concatenating only IEDs into a tensor has been practiced already, resulting in appropriate results [10].

Suppose our training dataset consists of $N$ IED segments with $L$ time samples and $M$ channels. All IEDs are concatenated into a tensor $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{L \times M \times N}$. Then, CPD is employed to decompose $\boldsymbol{\mathcal{X}}$ to the factor matrices with $R$ components as follows: temporal factors $\mathbf{A} \in \mathbb{R}^{L \times R}$, spatial factors $\mathbf{B} \in \mathbb{R}^{M \times R}$, and segmental factors $\mathbf{C} \in \mathbb{R}^{N \times R}$. Finally,

both IED and non-IED segments are projected onto the temporal factors in TCA and onto the spatial factors in SCA. The details are given below (the overall schematic is shown in Figure 3.5).

### 3.3.1 *Temporal Component Analysis*

Since the IEDs have very similar morphologies, projecting the IED and non-IED segments onto the temporal components can provide discriminative features. In TCA, the Khatri-Rao product is employed for the projection as follows:

$$\mathbf{Y}_k = \mathbf{A}^T \odot \mathbf{X}_k^T, \tag{3.19}$$

where the symbol '$\odot$' denotes the Khatri-Rao product, $\mathbf{X}_k \in \mathbb{R}^{L \times M}$ is an IED or non-IED segment, and $\mathbf{Y}_k \in \mathbb{R}^{(MR) \times L}$ represents the same segment after projection onto the temporal factors.

### 3.3.2 *Spatial Component Analysis*

It is clear that the IEDs originate from the same location. Thus, we can derive the discriminative features by projecting the IED and non-IED segments onto the spatial factors. The projection is performed as follows:

$$\tilde{\mathbf{Y}}_k = \mathbf{X}_k \mathbf{B} \tag{3.20}$$

where $\tilde{\mathbf{Y}}_k \in \mathbb{R}^{L \times R}$ is the projected IED or non-IED segment onto the spatial factors.

### 3.3.3 *TCA-SCA*

To improve the performance of the IED detection system, SCA and TCA are combined. Each segment is given separately to SCA and TCA to be classified in TCA-SCA; then, those segments classified in the IED class by both methods are recognized as IED.

## 3.4 EXPERIMENT

### 3.4.1 *Training and Test Datasets*

The scalp and intracranial EEGs (with a length of 20 minutes) of 7 epileptic subjects were analysed. A bandpass filter of 4-70 Hz bandwidth and a notch filter with a notch frequency of 50 Hz were applied to both sEEG and iEEG signals to increase the SNR

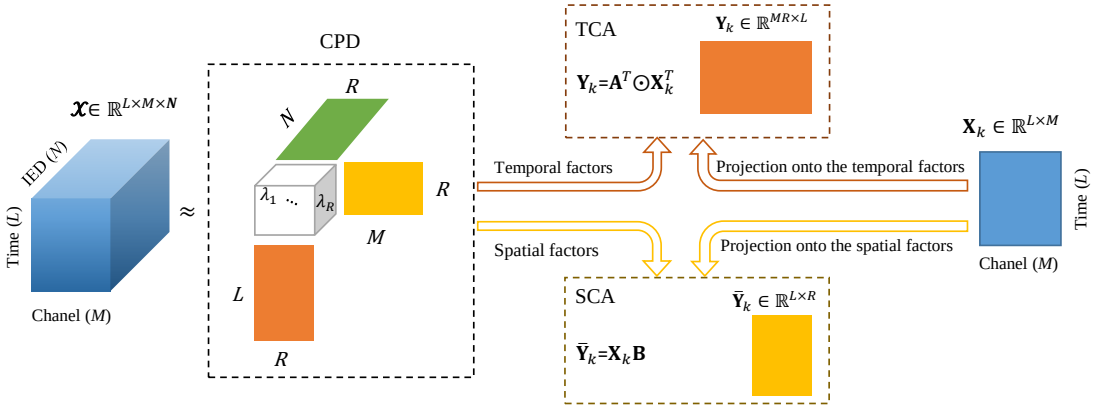



Figure 3.5: The proposed IED detection models. $\mathcal{X}$ includes the IED segments only. CPD is applied to $\mathcal{X}$ to decompose it to temporal, spatial, and segmental factors. $\mathbf{X}_k$ is an IED or non-IED segment from the training or test data, which is projected onto the temporal factors $\mathbf{A}$ in TCA and onto the spatial factors $\mathbf{B}$ in SCA.

and eliminate the power line interference. Details about the dataset are given in Section 3.1.

The first 10-minute recordings are used as the training data. For training the models, the IED segments were selected from the training signals with a length of 480 ms (96 samples). The peaks marked as IEDs were centered between $30^{th}$ to $34^{th}$ samples of IED segments. That is, the maximum difference between the IED peaks was four samples in the training dataset. Non-IED segments with 480 ms length were selected from the time segments in which there were no scored IEDs. We chose the same number of IED and non-IED segments for each subject to have a balanced classification problem.

The number of all IEDs and scalp-visible IEDs are shown in Table 3.1. The most scalp-visible IEDs appear in subject 7 with 31.1%. The percentage of scalp-visible IEDs for subjects 3-6 is less than 2. However, their average across subjects is 8.5% in the training dataset.

The second 10-minute recordings were used as the test data. A window of 96-sample length and 4-sample stride is slid along the test signals. Because of this 4-sample stride, the IEDs were not centered at the same point in the training dataset. However, after sliding over the entire test signals, we have 30000 segments (a few IED segments and numerous non-IED segments) for each subject.

### 3.4.2 *Training the Models and Real-Time IED Detection*

The IED segments in the training dataset, shown by $N$, are concatenated into a three-way tensor $\mathcal{X} \in \mathbb{R}^{96 \times M \times N}$, where 96 is time samples, and $M$ is the number of recorded channels ($M = 18$ for sEEG and $M = 12$ for iEEG). CPD is employed to decompose the

Table 3.1: The number of all IEDs and scalp-visible IEDs for each subject in the training dataset. The percentage of scalp-visible IEDs from all IEDs is illustrated in parentheses. The same number of non-IED segments were selected for each subject.

| Subject | No. of all IEDs | No. of scalp-visible IEDs (their percentage over all IEDs) |
|---|---|---|
| S1 | 182 | 36 (19.7%) |
| S2 | 270 | 16 (5.9%) |
| S3 | 179 | 3 (1.6%) |
| S4 | 482 | 6 (1.2%) |
| S5 | 420 | 3 (0.7%) |
| S6 | 303 | 0 (0.0%) |
| S7 | 135 | 42 (31.1%) |
| Mean | — | — (8.5%) |

tensor into the temporal, $\mathbf{A} \in \mathbb{R}^{96\times R}$, spatial, $\mathbf{B} \in \mathbb{R}^{M\times R}$, and segmental, $\mathbf{C} \in \mathbb{R}^{N\times R}$, factors.

In TCA, both IED and non-IED segments are projected onto the temporal factors by (3.19). Then, the magnitudes of STFT are obtained from the projected segments $\mathbf{Y}_k \in \mathbb{R}^{(MR)\times 96}$ using the spectrogram. In the spectrogram, a Hanning window with a length of 80 ms (16 samples) and 50% overlap is applied. The number of discrete Fourier transform points is set to 16, resulting in 9 frequency features. We obtain $(MR) \times 11 \times 9$ features ($M$ and $R$ respectively correspond to the number of channels and factors, 11 and 9 are respectively the number of time and frequency features) for each IED or non-IED segment.

In SCA, the segments are projected onto the spatial factors by (3.20). Then, TF features, as described above, are obtained from the projected segments $\bar{\mathbf{Y}}_k \in \mathbb{R}^{96\times R}$. This method extracts $R \times 11 \times 9$ features.

The optimized number of components $R$ is determined by the k-fold nested cross-validation technique. The training dataset is divided into five folds. Four folds are employed to train a classifier and the rest for validation. This procedure is repeated for all folds. We found that CPD with three components $(R = 3)$ provides the best performance.

A decision tree classifier is employed for classification. The classifier is trained separately for each subject by all the training trials. A 96-sample window with 4-sample stride slides along the signals to detect the IEDs from the ongoing signals. Each window is given to TCA and SCA to be classified as an IED or a non-IED segment. Recall that TCA-SCA is the combined model of TCA and SCA. If both TCA and SCA detect a segment as an IED, TCA-SCA categorizes it in the IED class. Otherwise, the segment is categorized in the non-IED class.

Table 3.2: The number of all IEDs and scalp-visible IEDs existed in the ongoing test data. The percentage of scalp-visible IEDs from all IEDs is illustrated in parentheses.

| Subject | No. of all IEDs | No. of scalp-visible IEDs (their percentage from all IEDs) |
|---|---|---|
| S1 | 160 | 14 (8.7%) |
| S2 | 202 | 3 (1.4%) |
| S3 | 162 | 17 (10.4%) |
| S4 | 366 | 5 (1.3%) |
| S5 | 408 | 1 (0.2%) |
| S6 | 303 | 0 (0.0%) |
| S7 | 89 | 10 (11.2%) |
| Mean | — | — (4.7%) |

Table 3.3: The performance of classifiers detecting IEDs from the iEEG. SEN is shown in percent (%). The mean of FP/min is adjusted to be approximately 5 by choosing a high threshold for the classifier.

| Subject | TCA | | SCA | | TCA-SCA | |
|---|---|---|---|---|---|---|
| | SEN | FP/min | SEN | FP/min | SEN | FP/min |
| S1 | **95.0** | 5.7 | 91.8 | 3.7 | 93.7 | 3.8 |
| S2 | 54.4 | 6.5 | 54.4 | 5.9 | **59.4** | 7.1 |
| S3 | 75.9 | 6.2 | **82.1** | 6.7 | 78.4 | 6.4 |
| S4 | 94.0 | 4.5 | 94.5 | 3.5 | 94.5 | 3.2 |
| S5 | 71.3 | 5.1 | 75.7 | 5.9 | **80.0** | 7.2 |
| S6 | 74.2 | 4.7 | 74.2 | 4.9 | **75.6** | 4.9 |
| S7 | 83.1 | 3.9 | 80.9 | 5.8 | **86.5** | 4.2 |
| Mean | 78.3 | 5.2 | 79.1 | 5.2 | **81.2** | 5.2 |

## 3.5 EXPERIMENTAL RESULTS AND DISCUSSION

For evaluation of the methods, SEN and FP/min are calculated. Recall that SEN shows the percentage of IEDs detected correctly, and FP/min is the number of non-IED segments recognized as IEDs incorrectly per minute. In all methods, a high threshold value is selected for the classifier to have an average of around 5 FP/min.

In the evaluation procedure, IEDs are not given as a segment to the models. A window slides across the ongoing signals. Therefore, there is no alignment among IED segments, causing a decrease in sensitivity and an increase in the FP rate. However, a segment classified as an IED segment counts as TP if an IED segment exists within 32 samples before or after the detected sample; otherwise, it counts as FP.

The number of IEDs and scalp-visible IEDs for the test signal is shown in Table Table 3.2. The average of scalp-visible IEDs is 4.7%, which is extremely low. Approxi-

Table 3.4: The performance of classifiers detecting IEDs from the sEEG. SEN is shown in percent (%). The mean of FP/min is adjusted to be approximately 5 by choosing a high threshold for the classifier.

| Subject | TCA | | SCA | | TCA-SCA | |
|---|---|---|---|---|---|---|
| | SEN | FP/min | SEN | FP/min | SEN | FP/min |
| S1 | 61.2 | 5.0 | 49.4 | 3.0 | **64.4** | 3.4 |
| S2 | 27.2 | 5.6 | **35.6** | 6.6 | 25.2 | 7.1 |
| S3 | 22.2 | 6.1 | **32.7** | 6.7 | 30.2 | 5.6 |
| S4 | **27.3** | 6.4 | 25.7 | 7.0 | 20.0 | 6.7 |
| S5 | **36.2** | 4.9 | 28.2 | 5.1 | 32.3 | 3.9 |
| S6 | 40.9 | 7.6 | 40.9 | 5.8 | **44.9** | 7.2 |
| S7 | 30.3 | 1.8 | 29.2 | 3.7 | **44.9** | 3.3 |
| Mean | 35.1 | 5.3 | 34.5 | 5.4 | **37.4** | 5.3 |

mately 10% of IEDs are visible over the scalp in subjects 1, 3, and 7, less than 1.5% in subjects 2 and 4, and approximately 0% in subjects 5 and 6.

Table 3.3 illustrates the performance of methods in detecting the IEDs from ongoing iEEG recordings. The methods provide different SEN and FP/min values for different subjects. TCA detects IEDs with 78.3% SEN. SCA provides a higher performance of 79.1% SEN. However, TCA-SCA outperforms both TCA and SCA by achieving 81.2% SEN. All three methods detect IEDs with 5.2 FP/min.

The performance of IED detection methods from sEEG signals is shown in Table 3.4. TCA and SCA respectively detect IEDs with 35.1% and 34.5% SEN. However, the best performance is obtained by TCA-SCA, providing 37.4% SEN. TCA-SCA detects IEDs of subject 1 with 64.4 SEN. For subjects 6 and 7, it achieves the SEN of 44.9 %. However, the SEN values are less than 33% for other subjects.

The IEDs are labeled based on the iEEG. In other words, the iEEG is used as a ground truth for labeling the IEDs. Therefore, a large proportion of IEDs in our data is invisible over the scalp. The percentages of scalp-visible and invisible IEDs in the data used in this study were already investigated [6]. In their study, the IEDs of 20 patients were used. They showed that only 9% of IEDs averaged across all 20 subjects are visible in their sEEG. These scalp-visible IEDs were observed from 13 out of 20 subjects, meaning that some patients may not have scalp-visible IEDs. In our dataset, seven patients are examined. In the test data, only 4.7% of IEDs are observable in our sEEG. Subject 6 does not even have any scalp-visible IED. However, the obtained results from the sEEG show that the proposed method enables the detection of scalp-invisible IEDs from the sEEG by fair sensitivity and low FP/min. The obtained SEN from the sEEG is significantly higher than the percentage of scalp-visible IEDs for all subjects.

## 3.6 CONCLUSION

In this chapter, we first show the importance of multi-way analysis in EEG signal processing. Then, the common tensor factorisation techniques, namely CPD and TD, are explained. Afterward, our developed methods based on tensor factorisation – namely TCA and SCA – to detect the IEDs from ongoing concurrent sEEG and iEEG recordings are presented. At the end, the results of TCA and SCA are combined to improve the sensitivity. TCA-SCA outperforms others by providing 81.2% and 37.4% SEN when the IEDs are detected respectively from the iEEG and sEEG signals. The obtained results are promising since only 4.7% of IEDs are visible from our sEEG dataset. The findings show that multi-way analysis can detect scalp-visible IEDs from ongoing sEEG recordings.

## 3.7 AKNOWLEDGE

Electronic version of an article published as "Online Detection of Scalp-Invisible Mesial-Temporal Brain Interictal Epileptiform Discharges from EEG," ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Singapore, Singapore, 2022, pp. 1416-1420, doi: 10.1109/ICASSP43922.2022.9746807.

# 4

# DETECTION OF IEDS FROM INTRACRANIAL EEG BY EXPLOITING THEIR MORPHOLOGY IN THE TENSOR STRUCTURE

The tensor factorisation approach has successfully exploited the waveform structure in various domains to decompose the data into its constituent components. It provides an opportunity to consider the data diversity. This means that different aspects of data (e.g., time, location, morphology, etc.) can be analyzed together by tensor factorisation. On the other hand, incorporating both IED and non-IED segments that have shape diversity into a slab of the tensor can deteriorate the factors obtained using tensor decomposition. Due to this fact, Thanh *et al.* [10] put only epileptic spikes in the fourth slab of the tensor. However, the epileptic spikes or IEDs have various morphologies and strengths. To the best of our knowledge, no study considers the impact of IED morphology in an automatic IED detection system. Therefore, we propose a model based on tensor factorisation to take the effect of IED morphologies into account.

The remainder of the chapter is structured as follows. First, we propose an IED detection system based on tensor factorisation in which the IEDs with similar morphology are concatenated into the same slice of a tensor to incorporate the IED morphology into account. Then, some detail about the IED dataset is given. Next, the results of applying our proposed method to the iEEG are presented. Finally, the last section concludes the chapter.

## 4.1 EXPLOITING IED MORPHOLOGY IN THE TENSOR STRUCTURE

Here, a tensor-based IED detection model is proposed. The IEDs may share some spatial and morphological information. Nonetheless, the non-IED segments can be non-epileptic spikes or normal brain activities; hence, there is no common information among them. Therefore, the feature space, including only the IED segments, can be more reliable and discriminative. Furthermore, since the IED morphologies can be different, we are interested in separating the IEDs with different scores by exploiting their morphological diversities as described in Chapter 1 of this thesis. In this study, the IEDs are given a score based on their morphology and spatial information by an expert epileptologist. IEDs with similar scores are concatenated into a three-way tensor with the dimensions of time samples, channels, and IED segments of the same scores. Next, all three-way tensors are concatenated into a single four-way tensor with the dimensions of time, channel, IED segment, and morphology. The CPD algorithm is then

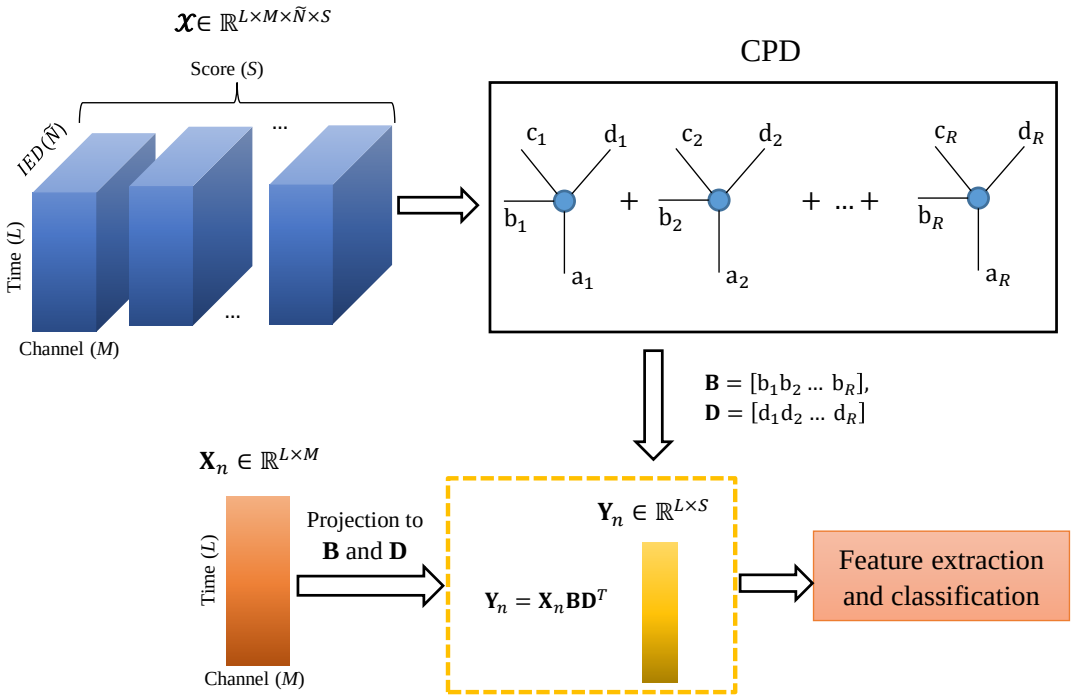



Figure 4.1: The schematic of proposed SMCA model. $\mathcal{X}$ includes the IED segments which are concatenated in four-way tensors according to their scores given by an expert based on the IED morphologies. CPD is applied to $\mathcal{X}$ to decompose it to temporal, spatial, segmental, and morphological factors. $\mathbf{X}_n (n = 1, \dots, N_1 + N_2)$ is an IED or non-IED segment from the training or test data. $\mathbf{Y}_n$ represents the same segment after projection onto the spatial $\mathbf{B}$ and morphological $\mathbf{D}$ components.

employed to extract the factor matrices. It should be noted that other tensor factorisation methods, such as TD or block term decomposition, can also be used instead. Finally, both IED and non-IED segments are projected onto the spatial and morphological components to achieve the most discriminative features. This model is called IED detection based on spatial and morphological component analysis (SMCA).

Suppose we are given $N_1$ IED segments with different morphologies (scores) and $N_2$ non-IED segments. We construct a four-way tensor $\mathcal{X} \in \mathbb{R}^{L \times M \times \tilde{N} \times S}$, where $L$ and $M$ denote respectively time samples and the number of channels, $\tilde{N}$ corresponds to the number of IED segments in each group which needs to be equal, and $S$ is the number of morphological groups.

CPD is employed to decompose the tensor $\mathcal{X}$ into its factor matrices:

$$\min_{\mathbf{A},\mathbf{B},\mathbf{C},\mathbf{D}} \quad f \equiv \frac{1}{2} \left\| \mathcal{X} - [\![\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}]\!] \right\|^2, \tag{4.1}$$

where $\mathbf{A} \in \mathbb{R}^{L \times R}$ and $\mathbf{B} \in \mathbb{R}^{M \times R}$ correspond respectively to the temporal and spatial factors, and $\mathbf{C} \in \mathbb{R}^{\tilde{N} \times R}$ and $\mathbf{D} \in \mathbb{R}^{S \times R}$ are respectively the segmental and morphological factors. Recall that $R$ is the number of components.

As IEDs originate from temporal lope regions and a large proportion of them are captured by the same electrode for the same subject, the spatial components can provide the most discriminative features. Moreover, morphological components are informative due to capturing the IED waveform information. Therefore, both IED and non-IED segments of the training and test datasets are projected onto the spatial and morphological factors as follows:

$$\mathbf{Y}_n = \mathbf{X}_n \mathbf{B} \mathbf{D}^T \tag{4.2}$$

where $\mathbf{X}_n \in \mathbb{R}^{L \times M} (n = 1, \ldots, N_1 + N_2)$ is an IED or non-IED segment from the training or test datasets and $\mathbf{Y}_n \in \mathbb{R}^{L \times S} (n = 1, \ldots, N_1 + N_2)$ is the same segment after projection. Now, features of $\mathbf{Y}_n$ are extracted and used for classification. Here, we extract time-frequency (TF) features using the spectrogram method. The schematic of the proposed SMCA method is illustrated in Figure 4.1.

## 4.2 EXPERIMENTS

### 4.2.1 *Dataset*

The iEEG signals of 18 epileptic subjects are analysed. The dataset details are given in Section 3.1. The iEEG recordings are filtered using a Butterworth filter of order six and cut-off frequencies of 4 and 70 Hz. The highpass frequency of 4 Hz has been selected to eliminate eye blink artefacts. In addition, a 50 Hz notch filter was employed to remove the power line interference.

### 4.2.2 *IEDs Morphology*

An expert epileptologist identified the IEDs and gave a confidence score between 1 to 9 for each spike/sharp wave based on their morphology and spatial distribution of the observed waveforms. This score represents his confidence in labeling the spike as an IED. More details about how the IEDs were scored are given in Section 3.1.3. The spikes/sharp waves scored between 1 and 4 are excluded from the analysis and considered artefacts or undistinguishable IEDs. Therefore, we have five types of IED morphology with a score of 5 to 9. The IEDs scored the same look approximately similar in morphology.

The iEEG signals were segmented into IED and non-IED segments before classification. The lengths of IEDs were selected to be 480 ms (96 time samples) – 160 ms before and 320 ms after the positions of peaks manually marked as IED by an expert. Non-IEDs with the same length as IEDs were selected from time segments where there was no sign of IEDs. Note that non-IED segments included non-epileptic spikes and sharp

Table 4.1: The total number of IED and non-IED segments for each subject. The same number of IED and non-IED segments were chosen for each subject.

| Subject | No. of segments | Subject | No. of segments |
|---|---|---|---|
| S1 | 38 | S10 | 622 |
| S2 | 524 | S11 | 692 |
| S3 | 302 | S12 | 344 |
| S4 | 108 | S13 | 26 |
| S5 | 158 | S14 | 20 |
| S6 | 648 | S15 | 692 |
| S7 | 250 | S16 | 22 |
| S8 | 552 | S17 | 178 |
| S9 | 38 | S18 | 338 |

waves, biological and non-biological artefacts, and normal brain activities. The same number of IEDs and non-IEDs was used to have a balanced classification problem. The number of segments is illustrated in Table 4.1.

### 4.2.3 *IED Detection Based on SMCA*

We construct a four-way tensor with the dimensions of time, channel, IED trial, and morphology, $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{96\times 12\times \widetilde{N}\times 5}$, where 96 is time samples, 12 is the number of channels, $\widetilde{N}$ denotes the number of IED segments with the same score, and 5 corresponds to the number which refers to a particular morphology. In other words, IEDs with the same morphology are put in the third mode, and the morphology types are put in the fourth mode. The number of IEDs for each type of morphology, $\widetilde{N}$, is set to the lowest number of IEDs scored 1, 2, 3, 4, or 5. Then, CPD is employed to decompose the tensor into temporal, spatial, segmental, and morphological factors. In CPD, the number of components, $R$, has to be less than or equal to the lowest number of observations in tensor modes. Thus, it cannot be bigger than 5 in our study. It is set to the maximum value of 5. As a result, the factors are $\mathbf{A} \in \mathbb{R}^{96\times 5}$, $\mathbf{B} \in \mathbb{R}^{12\times 5}$, $\mathbf{C} \in \mathbb{R}^{\widetilde{N}\times 5}$, and $\mathbf{D} \in \mathbb{R}^{5\times 5}$. After decomposition, the IED and non-IED segments ($\mathbf{X}_n$) are projected onto the spatial and morphological components using (4.2). The projected IEDs and non-IEDs ($\mathbf{Y}_n$) have the dimensions of $96 \times 5$.

For classification, TF features are obtained. The TF features of the projected segments are exploited using the spectrogram. A Hanning window of length 80 ms (16 samples) and overlapping of 50% slid over each channel of the projected segments (five channels) to obtain time-frequency features (totally 11 windows). The squared magnitudes of STFT are obtained using the spectrogram and utilized as classification features. The number of discrete Fourier transform points has been set to 16 (the same as the number

of time samples in a window), resulting in 9 frequency features. Finally, $11 \times 9 \times 5$ features (495) were obtained from each IED or non-IED segment, where 11 is the number of time slabs, 9 is the number of frequency slabs, and 5 is the number of components of the projected segments.

### 4.2.4 *Compared Methods*

To show the effect of incorporating IED morphologies into an IED detection system, we compared our proposed SMCA method with spatial component analysis (SCA), in which all IEDs are concatenated into a three-way tensor. The SCA is presented in Chapter 3 for IED detection from sEEG. For comparison purposes, here, it is used for IED detection from the iEEG. For a fair comparison, five components $(R = 5)$ giving the best performance are extracted like SMCA. Finally, both IEDs and non-IEDs are projected onto the spatial components. After projection, the same feature extraction and selection are applied to extract the significant features.

The proposed method is also compared with three state-of-the-art methods in which the authors use the same dataset. [20, 24]. In [20], the authors proposed a binary convolutional neural network (CNN-Bin) and a multiclass CNN (CNN-Multi) to detect IEDs using the same data. In CNN-Bin, the IEDs and non-IEDs are detected in a binary classification approach. In CNN-Multi, the authors detected IEDs based on their scores. Spyrou, Kouchaki, and Sanei [24] proposed a TD-based method. In the TD method, TF features are extracted using the spectrogram method. Then, a three-way tensor with the dimension of channel, time, and frequency is constructed and decomposed using TD. Finally, IED and non-IED segments are projected onto the spatial components.

In addition, we have compared two traditional methods, namely wavelet features (WFs) and TF features. In the TF approach, TF features are extracted from the raw iEEG in the same manner extracted here.

### 4.2.5 *Feature Selection*

The Fisher score is employed to select the most significant features. This is defined as:

$$f_i = \frac{\sum_{c=1}^{c=C} n_c(\mu_{ic} - \mu_i)^2}{\sum_{c=1}^{c=C} n_c \sigma_{ic}^2}, \tag{4.3}$$

where $\mu_{ic}$ and $\sigma_{ic}$ denote respectively the mean and standard deviation of the $i$-th feature in the $c$-th class, $n_c$ is the number of instances in the $c$-th class, and $\mu_i$ is the mean of the $i$-th feature.

Table 4.2: The performance of classifiers with results averaged over all subjects. The classifiers were trained and tested using leave-one-subject-out cross validation. ACC, SEN, and SPC are presented in percent %.

| Classifiers | Method | ACC | SEN | SPC | F1-S | AUC |
|---|---|---|---|---|---|---|
| KNN(k=3) | SCA | 84.8 | 77.4 | 92.2 | 0.81 | 0.91 |
| | SMCA | **86.7** | **78** | **95.4** | **0.84** | **0.93** |
| KNN(k=5) | SCA | 85.4 | 77.8 | 93 | 0.82 | 0.93 |
| | SMCA | **87** | **77.7** | **96.2** | **0.84** | **0.94** |
| NB | SCA | 84.4 | 71 | 97.8 | 0.79 | 0.88 |
| | SMCA | **86.4** | **74.8** | **98** | **0.83** | **0.9** |
| DTE | SCA | 92.6 | **89.7** | 95.6 | **0.92** | 0.99 |
| | SMCA | **92.9** | **89.7** | **96.1** | **0.92** | 0.99 |

### 4.2.6 *Cross Validation and Classification*

Leave-one-subject-out cross validation was employed to validate the models. The IEDs and non-IEDs of a subject were used as the test data and others for training the classifiers. Decision tree ensembles (DTE) with the bagging technique, naïve Bayes (NB), k-nearest neighbors (KNN) with $k = 3$ and $k = 5$ were employed as the classification methods. For evaluation of the methods, ACC, SEN, SPC, F1-S, and AUC defined in (2.1) were obtained.

## 4.3 EXPERIMENTAL RESULTS

First, we compare our proposed SMCA method with SCA, both of which are based on spatial components, and CPD is employed as the decomposition method. The obtained results are shown in Table 4.2. KNN models are based on the first 10 features, and NB is based on the first 30 features obtained using the Fisher score. In DTE, the first 80 features were utilized in the SMCA method, and the first 100 features were employed for the SCA method. These numbers of features gave the best performances for their algorithms.

Using both KNNs, SMCA outperforms the compared methods. Using KNN with $k = 3$, SMCA provides 86.7% accuracy, which is approximately 2% higher than the SCA accuracy value. Using KNN with $k = 5$, SMCA presents the best accuracy of 87%, which is 1.6% higher than that of SCA. In terms of SEN and SPEC, SMCA also outperforms SCA using both KNNs.

Using the NB classifier, the SMCA method detects the IEDs and non-IEDs with 86.4% accuracy, while SCA presents 84.4% accuracy value. In terms of SEN, SMCA signif-

Table 4.3: Comparing the proposed SMCA model with WF, TF, TD, CNN-Bin, and CNN-Multi proposed in [24] and [20]. ACC, SEN, and SPC are presented in percent %.

| Method | ACC | SEN | SPC | F1-S | AUC |
|---|---|---|---|---|---|
| WF | 72.3 | 70 | 72 | 0.73 | 0.73 |
| TF | 85.6 | 78 | 72 | 0.76 | 0.85 |
| TD | 86 | - | - | - | - |
| CNN-Bin | 85.9 | 90 | 87 | 0.88 | 0.88 |
| CNN-Multi | 89 | **94** | 81 | 0.88 | 0.9 |
| SMCA | **92.9** | 89 | **96** | **0.92** | **0.99** |

icantly outperforms the compared methods, though all models present comparable SPC values. SMCA provides the best F1-S and AUC as well.

Using DTE, SMCA provides the best ACC of 92.9% and SPC of 96.1%, followed by SCA with a small difference. In terms of SEN and F1-S, SMCA and SCA achieve comparable values, and both methods obtain 0.99 AUC.

Overall, both methods using all four types of classifiers result in higher SPC than SEN. SPC values are higher than 90% in both models, while SEN values are less than 80% in both methods when KNNs and NB classifications are employed. In DTE, there is an appropriate trade-off between SEN and SPC in both methods.

Furthermore, our proposed method is compared with TD developed in [24] and with WF, TF, CNN-Bin, and CNN-Multi proposed in [20]. The results are illustrated in Table 4.3. The performance of SMCA using the DTE classifier leads to the best performance.

Our proposed SMCA model significantly outperforms the compared methods by providing 92.9% accuracy. Among the compared methods, CNN-Multi presents the highest accuracy of 89%, which is approximately 4% less than that of SMCA. Although CNN-Multi detects IEDs with the highest SEN of 94%, SMCA leads to the best values of SPC, 96%, F1-S, 0.92, and AUC, 0.99.

## 4.4 CONCLUSION

Here, we propose a new method for IED detection based on spatial and morphological components. The proposed SMCA method has been compared with 1) SCA, 2) WF, 3) TF, 4) TD, 5) CNN-Bin, and 6) CNN-Multi. SMCA outperforms all compared methods and detects IEDs with 92.9% accuracy. In addition, the difference between SMCA and SCA is that, in SCA, all IEDs are concatenated into a three-way tensor without considering their scores. The findings show that considering the IED morphologies in an IED detection system can boost the model's performance.

## 4.5 AKNOWLEDGE

Electronic version of an article published as "Detection of Brain Interictal Epileptiform Discharges from Intracranial EEG by Exploiting their Morphology in the Tensor Structure," 2021 29th European Signal Processing Conference (EUSIPCO), Dublin, Ireland, 2021, pp. 1167-1171, doi: 10.23919/EUSIPCO54536.2021.9616233.

# 5

# INCORPORATING LABELING UNCERTAINTY IN AN IED DETECTION SYSTEM

The IEDs have a wide variety of morphologies. They may appear as sharp waves, spikes, or poly-spikes, often followed by an inhibitory damped oscillation. Also, depending on having unilateral or bilateral IED sources, the number of EEG channels with visible IEDs can be different, meaning that the IEDs have various spatial distributions. Furthermore, some IEDs are similar to artefacts (i.e., extracerebral potentials from muscle, eyes, heart, electrodes, etc.) and waves which are part of the brain's normal activity [84]. These properties (having different morphologies and spatial distributions as well as the IEDs' similarity to some artefacts and normal waves) make a great deal of uncertainty in labeling the IEDs and making IED detection difficult. This uncertainty can be mathematically expressed by the probability of the waveform being an IED. To the best of our knowledge, there have not been any studies that incorporate such uncertainty in automatically detecting the IEDs. Therefore, we aim to involve the IED probabilities in the design of a high-dimensional tensor decomposition system.

In this chapter, IEDs are detected in two approaches, within-subject and between-subject classification, from sEEG. In the first approach, we train a classifier for each subject individually. Conversely, in the between-subject classification approach, we train a classifier for all the subjects.

At first, a method based on SCA, which has been proposed in Chapter 3, is employed with a bit change to detect IEDs. In SCA, we construct a three-way tensor of time, sEEG channel, and IED segment in the within-subject classification approach and a four-way tensor of time, sEEG channel, IED segment, and subject in the between-subject one. Then, the tensor is decomposed into temporal, spatial, and segmental modes using the CANDECOMP/PARAFAC optimization (CP-OPT) algorithm proposed by Acar *et al.* [114]. Finally, both IED and non-IED segments are projected onto the spatial factors to derive the discriminative features.

In the second model, apart from the data tensor, a probability tensor is defined according to the probability of the waveform being an IED. Then, a weighted CP-OPT (CP-WOPT) [115] is employed to obtain the factor matrices. Finally, we projected both IED and non-IED segments onto the spatial factors to extract the most significant features. This is called spatial component analysis by considering the IED probabilities (SCA-IEDP)-based method for IED detection.

The chapter is organized as follows. First, a brief review of CP-OPT and CP-WOPT algorithms, our proposed SCA and SCA-IEDP models for IED detection, feature extrac-

tion and selection, and the compared methods are presented. The next section shows the experimental results, followed by an in-depth discussion on the pros and cons of the proposed models compared to competing methods. Finally, the last section concludes the chapter.

## 5.1 SIMPLE AND WEIGHTED CANDECOMP/PARAFAC OPTIMIZATION

### 5.1.1 *CP-OPT*

Suppose we are given a $V$-way tensor $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_V}$. We aim to factorize the tensor into the sum of rank-one tensors as follows:

$$\boldsymbol{\mathcal{X}} \approx \sum_{r=1}^{R} \mathbf{a}_r^{(1)} \circ \cdots \circ \mathbf{a}_r^{(V)}, \tag{5.1}$$

where $\mathbf{a}_r^{(v)} \in \mathbb{R}^{I_v}$ for $v = 1, \ldots, V$ and $r = 1, \ldots, R$ ($R$ is the number of components). The factor matrices are constructed from the combination of vectors from the rank-one components, i.e., $\mathbf{A}^{(v)} = \left[\mathbf{a}_1^{(v)} \cdots \mathbf{a}_R^{(v)}\right]$. Following the "Kruskal operator" [109], we can rewrite (5.1) as

$$\boldsymbol{\mathcal{X}} \approx [\![\mathbf{A}^{(1)}, \ldots, \mathbf{A}^{(V)}]\!] \equiv \sum_{r=1}^{R} \mathbf{a}_r^{(1)} \circ \cdots \circ \mathbf{a}_r^{(V)}. \tag{5.2}$$

The problem of computing CP (5.2) can be formulated as a least-square optimization problem:

$$\min_{\mathbf{A}^{(1)}, \ldots, \mathbf{A}^{(V)}} \quad f \equiv \frac{1}{2} \left\| \boldsymbol{\mathcal{X}} - [\![\mathbf{A}^{(1)}, \ldots, \mathbf{A}^{(V)}]\!] \right\|^2. \tag{5.3}$$

Unlike the alternating least-squares approach [97] solving the factor matrices one by one, the CP-OPT algorithm developed by Acar *et al.* [114] solves all the factor matrices simultaneously using a gradient-based optimization approach. It is straightforward to derive the gradient of objective function $f$ by calculating the partial derivatives with respect to each $\mathbf{A}^{(v)}$, i.e.,

$$\frac{\partial f}{\partial \mathbf{A}^{(v)}} = -\mathbf{X}_{(v)} \mathbf{A}^{(-v)} + \mathbf{A}^{(v)} \boldsymbol{\Gamma}^{(v)}, \tag{5.4}$$

for $v = 1, \ldots, V$, in which $\mathbf{X}_{(v)}$ is the mode-v matricization of the tensor $\boldsymbol{\mathcal{X}}$ and $\mathbf{A}^{(-v)}$ is defined as

$$\mathbf{A}^{(-v)} \equiv \mathbf{A}^{(V)} \odot \cdots \mathbf{A}^{(v+1)} \odot \mathbf{A}^{(v-1)} \odot \cdots \odot \mathbf{A}^{(1)}; \tag{5.5}$$

and $\mathbf{\Gamma}^{(v)}$ is defined as

$$\mathbf{\Gamma}^{(v)} = \mathbf{Y}^{(1)} * \cdots * \mathbf{Y}^{(v-1)} * \mathbf{Y}^{(v+1)} * \cdots * \mathbf{Y}^{(V)}, \tag{5.6}$$

where $\mathbf{Y}^{(v)} = \mathbf{A}^{(v)T}\mathbf{A}^{(v)}$.

To see the proof of (5.4) the reader is referred to [114]. Once the gradients are known, any gradient-based method can be used to solve the optimization problem. In this thesis, the nonlinear conjugate gradient method [116] is used.

### 5.1.2 *CP-WOPT*

Let $\boldsymbol{\mathcal{X}}$ be a $V$-way tensor with the dimension of $I_1 \times I_2 \times \cdots \times I_V$. In CP-WOPT, we define a weight tensor $\boldsymbol{\mathcal{W}}$ of the same size as $\boldsymbol{\mathcal{X}}$ such that $\{0 \leq w_{i_1 i_2 \cdots i_V} \leq 1\}$ for all $i_v \in 1, \ldots, I_v$ and $v \in 1, \cdots, V$.

The optimization for the general $V$-way CP-OPT factorisation in (5.3) changes to

$$\min_{\mathbf{A}^{(1)},\ldots,\mathbf{A}^{(V)}} f_w \equiv \frac{1}{2} \left\| \boldsymbol{\mathcal{W}} * \left( \boldsymbol{\mathcal{X}} - [\![ \mathbf{A}^{(1)}, \ldots, \mathbf{A}^{(V)} ]\!] \right) \right\|^2. \tag{5.7}$$

The above optimization problem may be reformulated to

$$\min_{\mathbf{A}^{(1)},\ldots,\mathbf{A}^{(V)}} f_w \equiv \frac{1}{2} \left\| \boldsymbol{\mathcal{Y}} - \boldsymbol{\mathcal{Z}} \right\|^2, \tag{5.8}$$

where $\boldsymbol{\mathcal{Y}} = \boldsymbol{\mathcal{W}} * \boldsymbol{\mathcal{X}}$ and $\mathcal{Z} = \boldsymbol{\mathcal{W}} * [\![ \mathbf{A}^{(1)}, \ldots, \mathbf{A}^{(V)} ]\!]$. We aim to find the factor matrices $\mathbf{A}^{(v)} \in \mathbb{R}^{I_v \times R}$, for $v = 1, \ldots, V$, which minimize the weighted objective function in (5.8). We derive the gradient in (5.8) by calculating the partial derivatives of $f_w$ with respect to each factor matrix $\mathbf{A}^{(v)}$ as follows:

$$\frac{\partial f_w}{\partial \mathbf{A}^{(v)}} = \left( \mathbf{Z}_{(v)} - \mathbf{Y}_{(v)} \right) \mathbf{A}^{(-v)}, \tag{5.9}$$

for $n = 1, \ldots, V$, where $\mathbf{A}^{(-v)}$ is defined in (5.5).

The detailed proof of (5.9) is expressed in Ref. [115]. After calculating the partial derivatives using (5.9), any gradient-based optimization method can be utilized to solve the optimization problem.

## 5.2 SCA (CP-OPT) AND SCA-IEDP (CP-WOPT) FOR IED DETECTION

The IEDs are detected in two within- and between-subject classification approaches using the proposed SCA and SCA-IEDP methods.

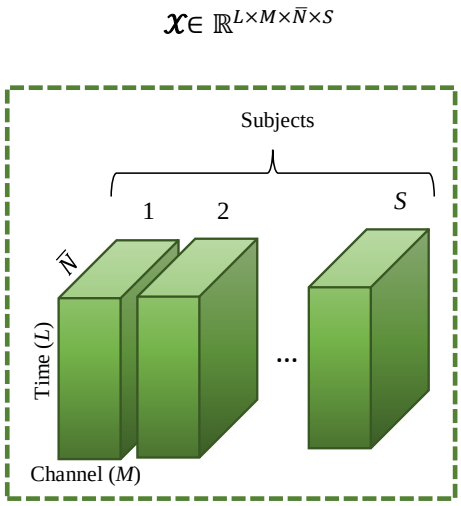



Figure 5.1: The four-way tensor constructed by concatenating the IEDs of $S$ subjects.

### 5.2.1 *SCA for Detecting IEDs in the Within-Subject Approach*

In the within-subject classification approach, a classifier is trained and validated using the data from the same subject. This method is explained in Section 3.3. The only difference is that here the tensor is decomposed by employing CP-OPT [114].

We construct a three-way tensor, $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{L \times M \times N}$ (whose dimensions $L$, $M$, and $N$ correspond respectively to time, channel, and IED segment of sEEG) for each subject by concatenating their IEDs. Then, the tensor is decomposed into the temporal $\mathbf{A} \in \mathbb{R}^{L \times R}$, spatial $\mathbf{B} \in \mathbb{R}^{M \times R}$, and segmental factors $\mathbf{C} \in \mathbb{R}^{N \times R}$ using CP-OPT. Then, both the IED and non-IED segments are projected onto the spatial factor $\mathbf{B}$ as follows:

$$\mathbf{P}_k = \mathbf{X}_k \mathbf{B} \tag{5.10}$$

where $\mathbf{X}_k \in \mathbb{R}^{L \times M} (k = 1, \ldots, K)$ is an IED or non-IED segment from the training or test data and $\mathbf{P}_k \in \mathbb{R}^{L \times R} (k = 1, \ldots, K)$ is the projected IED or non-IED segment.

### 5.2.2 *SCA for Detecting IEDs in the Between-Subject Approach*

In the between-subject classification approach, a number of subjects are used for training a classifier. Then, the classifier is validated using new subjects. The IEDs of subjects are concatenated into a four-way tensor $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{L \times M \times \bar{N} \times S}$, where $\bar{N}$ denotes the number of IED segments of sEEG supposed to be selected equally for all the training subjects, and $S$ is the number of subjects in the training dataset. Figure 5.1 shows the schematic of the four-way tensor.

CP-OPT is employed to decompose tensor $\boldsymbol{\mathcal{X}}$ to its factor matrices:

$$\min_{\mathbf{A},\mathbf{B},\mathbf{C},\mathbf{D}} \quad f \equiv \frac{1}{2}\left\|\boldsymbol{\mathcal{X}} - [\![\mathbf{A},\mathbf{B},\mathbf{C},\mathbf{D}]\!]\right\|^2, \tag{5.11}$$

where $\mathbf{A} \in \mathbb{R}^{L\times R}$ and $\mathbf{B} \in \mathbb{R}^{M\times R}$ represent respectively the temporal and spatial factors, and $\mathbf{C} \in \mathbb{R}^{\bar{N}\times R}$ and $\mathbf{D} \in \mathbb{R}^{S\times R}$ correspond respectively to the segmental and individual (subject) factors.

Finally, both IED and non-IED segments of the training and test subjects' sEEG are projected onto the spatial factor as follows:

$$\mathbf{P}_{\bar{k}} = \mathbf{X}_{\bar{k}}\mathbf{B} \tag{5.12}$$

where $\mathbf{X}_{\bar{k}} \in \mathbb{R}^{L\times M}(\bar{k} = 1,\ldots,\bar{K})$ is an IED or non-IED segment from the training or test subjects and $\mathbf{P}_{\bar{k}} \in \mathbb{R}^{L\times R}(\bar{k} = 1,\ldots,\bar{K})$ is the projected IED or non-IED segment.

### 5.2.3 *SCA-IEDP for Detecting IEDs in the Within-Subject Approach*

IEDs of each subject are concatenated into a three-way tensor $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{L\times M\times N}$. Apart from the data tensor, in SCA-IEDP, we define a probability tensor $\boldsymbol{\mathcal{W}}$ of the same size as $\boldsymbol{\mathcal{X}}$ based on the IED probabilities as follows:

$$\begin{cases} \mathbf{W}_{::n} = c_1 & \text{for case 1} \\ \mathbf{W}_{::n} = c_2 & \text{for case 2} \\ \vdots & \\ \mathbf{W}_{::n} = c_{J-1} & \text{for case } J-1 \\ \mathbf{W}_{::n} = c_J & \text{for case } J, \end{cases} \tag{5.13}$$

for all $n = 1,\ldots,N$, where $\{0 \le c_j \le 1\}$ for $j \in \{1,\ldots,J\}$ and $\mathbf{W}_{::n}$ are the frontal slices of $\boldsymbol{\mathcal{W}}$. The cases are defined based on spatial distribution and the scores given by expert epileptologists. The scores refer to the closeness of a spike to an expected IED morphology.

The CP-WOPT algorithm is applied to the data tensor $\boldsymbol{\mathcal{X}}$ and probability tensor $\boldsymbol{\mathcal{W}}$, as:

$$\min_{\mathbf{A},\mathbf{B},\mathbf{C}} \; f_w \equiv \frac{1}{2}\left\|\boldsymbol{\mathcal{W}} * (\boldsymbol{\mathcal{X}} - [\![\mathbf{A},\mathbf{B},\mathbf{C}]\!])\right\|^2, \tag{5.14}$$

to obtain the factor matrices, where $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$ correspond respectively to the temporal, spatial, and segmental factors. Then, both the IED and non-IED segments are projected onto the spatial factor $\mathbf{B}$ using (5.10). Finally, the projected IEDs and non-IEDs, $\mathbf{P}_k$, are

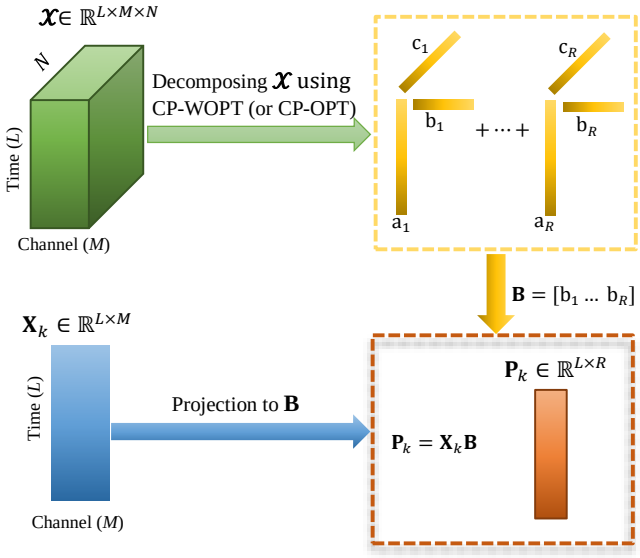



Figure 5.2: The IED detection system proposed for the within-subject classification approach. $\mathcal{X}$ includes the IED segments only, $N$. CP-OPT (or CP-WOPT) is applied to $\mathcal{X}$ to decompose it to temporal, spatial, and segmental factors. $\mathbf{X}_k (k = 1, \dots, K)$ is an IED or non-IED segment from the training or test data, which is projected onto the spatial components $\mathbf{B}$. $\mathbf{P}_k$ represents the same segment after projection.

used for classification. The flow diagram of the proposed models in the within-subject approach is illustrated in Figure 5.2.

### 5.2.4 *SCA-IEDP for Detecting IEDs in the Between-Subject Approach*

Apart from the four-way data tensor $\mathcal{X}$ described in Section 5.2.2, a four-way probability tensor $\mathcal{W}$ is also defined. The three-way probability tensors of the training subjects described in (5.13) are concatenated into a single four-way tensor, called four-way probability tensor $\mathcal{W}$.

The CP-WOPT algorithm is employed to decompose the data tensor $\mathcal{X}$ and probability tensor $\mathcal{W}$, as illustrated below, into the factor matrices:

$$\min_{\mathbf{A},\mathbf{B},\mathbf{C},\mathbf{D}} f_w \equiv \frac{1}{2} \left\| \mathcal{W} * (\mathcal{X} - [\![\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}]\!]) \right\|^2. \tag{5.15}$$

Finally, both the IED and non-IED segments are projected onto the spatial factor $\mathbf{B}$ using (5.12), and the projected IEDs and non-IEDs, $\mathbf{P}_{\bar{k}}$, are used for classification.

In the classification stage, all IEDs with different scores fall within the same IED class. However, we define a probability tensor $\mathcal{W}$ based on these scores, which gives an opportunity to incorporate the probability of a waveform being an IED in our IED detection system.

## 5.3 EXPERIMENTS

### 5.3.1 *Dataset*

The sEEG signals of 18 epileptic subjects are analysed. The dataset details are given in Section 3.1. A 4-48 Hz bandpass filter was applied to the sEEG signals to increase the signal-to-noise ratio. We selected the highpass frequency to be 4 Hz to alleviate the baseline and eye movement artefacts and the lowpass frequency to be 48 Hz to preserve the IEDs' morphology as IEDs are likely to have high-frequency components.

For analysis and classification, the EEG signals were sliced. The length of IED segments was selected to be 480 ms (96 time sample) – 160 ms before and 320 ms after the positions of peaks in the concurrent iEEG marked as IED. Non-IED segments with 480 ms length were selected from time segments where there were no scored IEDs. We chose the same number of IED and non-IED segments for each subject. Both IED and non-IED segments were linearly detrended to remove the undesired drifts. The total number of IED and non-IED segments for different subjects are illustrated in Table 4.1. Subjects 13, 14, and 16 were excluded from the analysis because of having less number of segments.

### 5.3.2 *IED Uncertainty*

The iEEG was used as the ground truth for scoring the IEDs by an expert epileptologist who gave a score between 1 to 9 for each spike/sharp wave based on their morphology and spatial distribution. More details about how the IEDs were scored are given in Section 3.1.3. The spikes/sharp waves scored between 1 and 4 are excluded from the analysis since they cannot be differentiated from artifacts. Therefore, IEDs from score 5 (the lowest certainty) to 9 (the highest certainty) are analysed here. Score 5 refers to the lowest probability of the waveform being an IED, meaning that the epileptologist is not certain whether the activity is an IED or not. Score 9 denotes the highest probability of being an IED, meaning that they are definitely IED activities. Figure 5.3 shows samples of non-IED and IED waveforms with scores 5 to 9. No spike or sharp wave can be seen in the non-IED segment. In the IED segment scored 5, there is a sharp wave in the FO channel. A couple of FO channels contain sharp waves in the IED segment scored 6. However, no spike or sharp wave is observable over the scalp for IEDs scored 5 or 6. There are spikes in a couple of FO channels in the IED segment scored 7, and some broad waves can be seen on a few scalp channels by referencing to FO channels. These segments (with IEDs scored 5, 6, or 7) have been marked as scalp-invisible IEDs. On the other hand, in the IEDs scored 8 or 9, there are spike waves in many FO and scalp channels with higher amplitudes. These have been marked as scalp-visible IEDs.

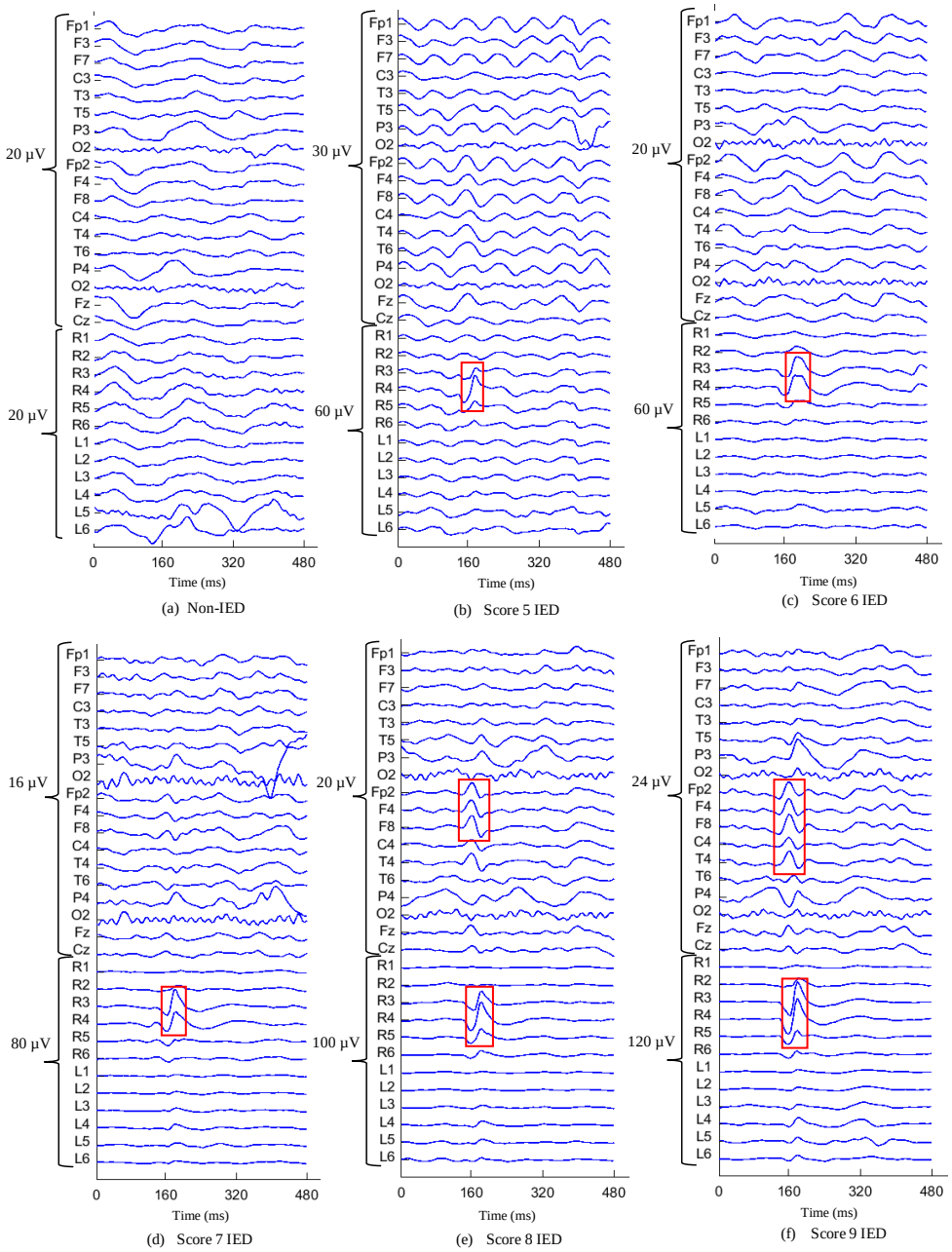



Figure 5.3: Samples of non-IED and IED waveforms with scores 5 to 9. (a) The non-IED segment, and (b)-(f) respectively the IED with score 5 to 9 (score 9 refers to an IED with the highest probability (or the lowest uncertainty)). An expert epileptologist used the iEEG as ground truth in scoring IEDs. The IEDs, pointed by red rectangles, start at 160ms.

### 5.3.3 *Feature Extraction*

#### 5.3.3.1 *Feature extraction in the within-subject approach*

Here, a three-way tensor $\mathcal{X} \in \mathbb{R}^{96\times 18\times N}$ is constructed, where 96 and 18 are respectively the number of time samples and recorded channels, and $N$ denotes the number of IED

segments in the training fold. Apart from the IED tensor, we build a tensor $\mathcal{W}$ with the same dimension as tensor $\mathcal{X}$ in the SCA-IEDP model to which different certainty levels are allocated as follows:

$$\begin{cases} \mathbf{W}_{::n} = 0.2 & \text{If the IED is given score 5} \\ \mathbf{W}_{::n} = 0.6 & \text{If the IED is given score 6} \\ \mathbf{W}_{::n} = 0.8 & \text{If the IED is given score 7} \\ \mathbf{W}_{::n} = 0.9 & \text{If the IED is given score 9} \\ \mathbf{W}_{::n} = 1 & \text{If the IED is given score 9,} \end{cases}$$

for $n = 1, \ldots, N$. $\mathbf{W}_{::n} = 0.2$ means that the waveform is related to an IED activity with 20% possibility and to a non-IED activity with 80% possibility. $\mathbf{W}_{::n} = 0.6$ means that the waveform is an IED activity with 60% possibility and it is a non-IED activity with 40% possibility and so on.

The tensor is decomposed separately using CP-OPT and CP-WOPT to extract the factor matrices:

$$\mathcal{X} \approx \sum_{r=1}^{R} \mathbf{a}_r \circ \mathbf{b}_r \circ \mathbf{c}_r \equiv [\![\mathbf{A}, \mathbf{B}, \mathbf{C}]\!],$$

where $\mathbf{A} \in \mathbb{R}^{96 \times R}$ and $\mathbf{B} \in \mathbb{R}^{18 \times R}$ denote respectively the temporal and spatial factors, and $\mathbf{C} \in \mathbb{R}^{N \times R}$ is the segmental factor. Then, both the IED and non-IED segments are projected onto the spatial factor $\mathbf{B}$ using (5.10), where $\mathbf{X}_k \in \mathbb{R}^{96 \times 18}$ is an IED or non-IED segment from the training or test data and $\mathbf{P}_k \in \mathbb{R}^{96 \times R}$ is the projected IED or non-IED segment.

Time-frequency representations have been broadly and successfully used in IED detection [8] and epilepsy diagnosis [117, 118]. The spectrogram method is applied to the projected IEDs and non-IEDs, $\mathbf{P}_k$, to extract the time-frequency features. The magnitudes of STFT obtained using the spectrogram are measured and used as features. For the spectrogram, we define a Hanning window size of 80 ms (16 samples) and an overlap of 50% (8 samples). Overall, 11 time slabs are captured by sliding the window over each segment. The number of discrete Fourier transform points has been set to 16 (the same as the number of time samples in a window), resulting in 9 frequency features. Totally, we obtained $R \times 11 \times 9$ features, where $R$ is the number of spatial components extracted using the tensor, 11 is the number of time slabs, and 9 is the number of frequency slabs for each IED or non-IED segment.

#### 5.3.3.2 *Feature extraction in the between-subject approach*

After constructing the four-way tensor $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{96\times18\times30\times S}$ and defining $\boldsymbol{\mathcal{W}}$ for CP-WOPT, we decompose $\boldsymbol{\mathcal{X}}$ using CP-OPT and CP-WOPT separately into temporal $\mathbf{A} \in \mathbb{R}^{96\times R}$, spatial $\mathbf{B} \in \mathbb{R}^{18\times R}$, segmental $\mathbf{C} \in \mathbb{R}^{30\times R}$ , and individual (subject) $\mathbf{D} \in \mathbb{R}^{S\times R}$ factors. We randomly select 30 IED segments from each training subject. (Therefore, subjects with at least 30 IED segments and reasonable performance in the within-subject approach are chosen as the training subjects). Then, both the IED and non-IED segments of the training and test subjects are projected onto the spatial factor using (5.12). Finally, the time-frequency features of the projected IEDs and non-IEDs are calculated using the spectrogram.

### 5.3.4 *Number of Components*

Identification of the number of components plays an important role in tensor decomposition, and it is also an NP-hard problem. However, we apply nested cross validation to determine the number of suitable components $R$. The training data are split into 5-fold. Each time four folds are used for training the model and the fifth fold for validation. The nested cross validation is performed for $R = \{1,2,3,4\}$. The accuracy is averaged across five folds in each case $R = \{1,2,3,4\}$. Finally, the number of components is determined based on the highest accuracy. In our work, the minimum number of components is one ($R = 1$), and the maximum number is four ($R = 4$), depending on how many components lead to the best performance. It is worth noting that the nested cross validation is performed over the training dataset only. The test dataset does not contribute to estimating the number of components.

### 5.3.5 *Compared Methods*

We compare the performance of our proposed approaches with two other state-of-the-art methods in IED detection, namely TF features [8] and simultaneous multilinear low-rank approximation of tensors (SMLRAT) [10].

#### 5.3.5.1 *Time-frequency features model*

We already discovered that the TF features show an improvement over CWT and chirplet transform for this particular dataset [8]. Therefore, we compare our proposed models with a model based on TF features, which is obtained by employing the spectrogram method – calculated following Section 5.3.3. Each IED or non-IED segment is made of 1782 features (18 scalp channels $\times$11 temporal $\times$9 frequency).

#### 5.3.5.2 *Simultaneous multilinear low-rank approximation of tensors*

We compare our proposed approaches with SMLRAT proposed recently for EEG epileptic spike detection [10]. The model is based on NTD. Having a positive-valued spectrogram allows us to fully benefit from NTD.

In the SMLRAT, the authors apply CWT to the epileptic and non-epileptic spikes and construct a three-way tensor for each segment, $\boldsymbol{\mathcal{X}}_i \in \mathbb{R}^{W \times L \times M}$, where $W, L,$ and $M$ denote wavelet-scale, time, and channel respectively. They concatenate only three-way tensors, $\{\boldsymbol{\mathcal{X}}_i^{ep}\}_{i=1}^{N}$, where $N$ represents the number of epileptic spike segments, in a single four-way tensor $\widetilde{\boldsymbol{\mathcal{X}}}^{ep} \in \mathbb{R}^{W \times L \times M \times N}$. Then, NTD is employed to decompose the tensor $\widetilde{\boldsymbol{\mathcal{X}}}^{ep}$ and obtain the factor matrices:

$$\widetilde{\boldsymbol{\mathcal{X}}}^{ep} = \widetilde{\boldsymbol{\mathcal{X}}}_1^{ep} \boxplus \widetilde{\boldsymbol{\mathcal{X}}}_2^{ep} \cdots \boxplus \widetilde{\boldsymbol{\mathcal{X}}}_N^{ep} = \boldsymbol{\mathcal{G}} \times_1 \mathbf{U} \times_2 \mathbf{A} \times_3 \mathbf{B} \times_4 \mathbf{C}, \quad (5.16)$$

where $\boxplus$ denotes concatenation operation, $\boldsymbol{\mathcal{G}} \in \mathbb{R}^{r_1 \times r_2 \times r_3 \times N}$ is the core tensor, and $\mathbf{U} \in \mathbb{R}^{W \times r_1}, \mathbf{A} \in \mathbb{R}^{L \times r_2}, \mathbf{B} \in \mathbb{R}^{M \times r_3},$ and $\mathbf{C} \in \mathbb{R}^{N \times N}$ denote respectively the wavelet-scale, time, channel, and epileptic spikes. Finally, in order to obtain the feature space of each segment, the spikes (no matter epileptic and non-epileptic) are projected onto the factor matrices as follows:

$$\boldsymbol{\mathcal{F}}_i = \boldsymbol{\mathcal{X}}_i \times_1 \mathbf{U}^{\dagger} \times_2 \mathbf{A}^{\dagger} \times_3 \mathbf{B}^{\dagger}, \quad (5.17)$$

where $(.)^{\dagger}$ represents matrix Moore-Penrose pseudo-inverse.

Here, we transform the IED and non-IED segments through CWT and construct a three-way tensor for each segment – $\boldsymbol{\mathcal{X}}_i \in \mathbb{R}^{38 \times 96 \times 18}$. We concatenate three-way IED tensors into a single four-way tensor, then perform NTD to obtain the factor matrices; $\mathbf{U} \in \mathbb{R}^{38 \times 10}, \mathbf{A} \in \mathbb{R}^{96 \times 15},$ and $\mathbf{B} \in \mathbb{R}^{18 \times 18}$; and consequently the features $\boldsymbol{\mathcal{F}}_i \in \mathbb{R}^{10 \times 15 \times 18}$.

### 5.3.6 *Evaluation and Cross Validation*

We employ a k-fold (k=5) cross validation to validate the methods in the within-subject classification approach. Four folds are used for training the models and the fifth fold as the test data. Increasing the number of folds did not change the outcome. It is worth emphasizing that the classifiers are trained and tested using only the sEEG. In the between-subject approach, the leave-subject-out cross validation is employed. The data of a subject are used as the test data; other subjects' data are used for training the classifiers.

For evaluation of the methods, ACC, SEN, SPC, and F1-S are obtained.

### 5.3.7 *Feature Selection and Classification*

We employ the Fisher score as the feature selection technique to select the most significant features. This method is explained in Section 4.2.5.

In order to classify the IED and non-IED segments, we employ four different classifiers, namely diagonal linear discriminant analysis (DLDA), support vector machines (SVM), NB, and DTE. DLDA is superior to LDA in high-dimensional problems. The covariance matrix needs to be computed in LDA. Calculating all the covariance matrix terms in a high-dimensional case may fail due to the small data size. A strong independence assumption may be applied to the features resulting in a diagonal covariance matrix in LDA. This modification of LDA is called DLDA. We have feature independence assumption in NB as well. SVM is a popular classifier often used for seizure EEG classification [119, 120]. In these cases, linear SVM is more applicable than nonlinear (kernel-based) SVM due to the separability of the data and to avoid overfitting. The DTE classifier has been the last one that we utilized. We have used the bagging technique to perform ensemble decision trees [121]. In the bagging method, the idea is to create several subsets of data from the training samples, which are randomly selected with replacement. Then, each subset is employed to train its decision tree. Hence, we end up with an ensemble of different models, where the average of all the predictions from different trees is used. Therefore, the bagging method is more robust than any single decision tree.

## 5.4 EXPERIMENTAL RESULTS

Here, we first analyze the temporal and spatial components of CP-OPT and CP-WOPT decomposition. We employ four types of classifiers to detect IEDs. Next, we present the obtained results of our proposed SCA and SCA-IEDP models and of the compared methods in the within-subject classification approach. Finally, we report the obtained results of SCA, SCA-IEDP, and SMLRAT in the between-subject classification approach.

### 5.4.1 *Analysis of Temporal and Spatial Factors*

The first temporal vector of CP-OPT and CP-WOPT is illustrated in Figure 5.4(a). The temporal factor of CP-WOPT has a higher amplitude than that of CP-OPT and is sharper and more similar to the source IED. It is because, in CP-WOPT, those IEDs with spike or spike-and-slow-wave complex morphology are given greater weight than those with sharp wave morphology. Whereas in CP-OPT, all IEDs with different morphologies are given the same weight.

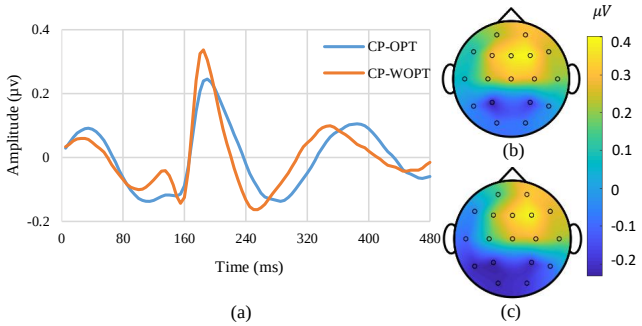



Figure 5.4: The first temporal components and spatial distributions obtained using CP-OPT and CP-WOPT: (a) The temporal components, (b) and (c) the spatial distributions obtained respectively by CP-OPT and CP-WOPT.

In addition, the spatial distributions of the first vector of CP-OPT and CP-WOPT are computed and respectively shown in Figure 5.4(b) and Figure 5.4(c). The spatial distribution of CPWOPT is more focal and right laterally distributed than that of CP-OPT as expected.

### 5.4.2 *IED Detection in the Within-Subject Classification Approach*

Table 5.1 illustrates the classification results for the within-subject approach. We compare our proposed SCA and SCA-IEDP models with TF and SMLRAT methods developed recently. It is worth noting that the first 200 significant features of the TF and the first 100 ones of SMLRAT are classified, giving the highest accuracy in their methods.

Using the NB classifier, SCA-IEDP detects IEDs and non-IEDs with 79.2% accuracy, which is approximately 1%, 8%, and 9% higher than SCA, SMLRAT, and TF accuracy values, respectively. SCA-IEDP provides 73.1% SEN and 87% SPEC which is better than SCA and the compared methods as well. In DLDA, SCA-IEDP provides the best ACC of 76% and SEN of 63.1%. In contrast, its performance is comparable with SCA performance in terms of SPEC criterion.

Using SVM, the best accuracy of 79.2% and sensitivity of 71.4% are obtained using our proposed SCA-IEDP model. However, SCA and SCA-IEDP present the same values of SPEC and F1-S. Except for SEN, where SMLRAT presents the poorest value, the TF model provides the worst in this classifier across all criteria. In the DTE classifier, SCA-IEDP presents the best accuracy of 79.9%, the sensitivity of 77.6%, and the F1-score of 0.79 though the best specificity of 82.8 is achieved using SCA.

Table 5.1 shows that the best accuracy is achieved by SCA-IEDP using the DTE classifier. It provides 79.9% accuracy while the best accuracy values obtained by SCA, SMLRAT, and TF are respectively 78.9%, 72.9%, and 73%. Overall, the SPEC values are higher than SEN values in all detection methods and all classifiers, showing that the

Table 5.1: The performance of classifiers in the within-subject classification approach with results averaged over all subjects and folds. ACC, SEN, and SPC are presented in percent %.

| Classifiers | Models | ACC | SEN | SPC | F1-S |
|---|---|---|---|---|---|
| NB | TF | 70.5 | 68 | 73 | 0.68 |
| | SMLRAT | 71.3 | 61.7 | 80.1 | 0.65 |
| | SCA | 78.1 | 70.3 | 86 | 0.75 |
| | SCA-IEDP | **79.2** | **71.3** | **87** | **0.76** |
| DLDA | TF | 70.7 | 61 | 80.3 | 0.66 |
| | SMLRAT | 64.9 | 46.4 | 83.4 | 0.54 |
| | SCA | 75.3 | 61.7 | **88.9** | 0.7 |
| | SCA-IEDP | **76** | **63.1** | **88.8** | **0.71** |
| SVM | TF | 66.9 | 65.2 | 68.2 | 0.66 |
| | SMLRAT | 72.9 | 63.8 | 82 | 0.69 |
| | SCA | 78.9 | 70.8 | **87** | **0.76** |
| | SCA-IEDP | **79.2** | **71.4** | **87** | **0.76** |
| DTE | TF | 73 | 66.2 | 79.8 | 0.69 |
| | SMLRAT | 72.7 | 70.9 | 74.5 | 0.70 |
| | SCA | 78.3 | 73.9 | **82.8** | 0.76 |
| | SCA-IEDP | **79.9** | **77.6** | 82.1 | **0.79** |

non-IED segments can be detected easier than the IED segments. SMLRAT, SCA, and SCA-IEDP provide their best SEN values using the DTE classifier, while TF obtains its best SEN using the NB classifier. Generally, DTE results in the best trade-off between SEN and SPEC in all methods.

### 5.4.3 *IED Detection in the Between-Subject Classification Approach*

Apart from classifying IEDs and non-IEDs in the within-subject classification approach, the proposed SCA and SCA-IEDP models and the SMLRAT are employed to detect IEDs in the between-subject one. The obtained results are shown in Table 5.2. In SMLRAT [10], epileptic and non-epileptic spikes are classified by concatenating all subjects' three-way tensors (time, wavelet-scale, channel) into a single four-way tensor, and leave-one-subject-out cross validation is used. The TF model proposed in Ref. [8] detects IEDs in the between-subject classification approach by using the ensemble of individual classifiers, rather than by combining the subjects' data. Therefore, the TF approach is not reported here.

SCA-IEDP presents the best accuracy in all the classifiers. It obtains its best accuracy, 63.4%, using both SVM and DTE classifiers. SCA-IEDP and SCA provide significantly better performance than SMLRAT. In all the classifiers, the accuracies of SCA-IEDP

Table 5.2: The performance of SCA, SCA-IEDP and SMLRAT models obtained using the classifiers in the between-subject classification approach. The results have been averaged over 15 subjects. ACC, SEN, and SPC are presented in percent %.

| Classifiers | Models | ACC | SEN | SPC | F1-S |
|---|---|---|---|---|---|
| NB | SMLRAT | 49.2 | **80** | 17.4 | **0.61** |
| | SCA | 58.2 | 37.4 | **79.1** | 0.39 |
| | SCA-IEDP | **62.4** | 56 | 68.7 | 0.52 |
| DLDA | SMLRAT | 50 | **98.5** | 1.6 | **0.66** |
| | SCA | 60.1 | 48.5 | **71.7** | 0.47 |
| | SCA-IEDP | **62.5** | 61.6 | 63.3 | 0.55 |
| SVM | SMLRAT | 50.9 | 47.7 | 54 | 0.36 |
| | SCA | 62.1 | 43.5 | **80.7** | 0.47 |
| | SCA-IEDP | **63.4** | **53.3** | 73.5 | **0.53** |
| DTE | SMLRAT | 49.9 | 51.9 | 48 | 0.50 |
| | SCA | 60.4 | 63.2 | 57.6 | 0.58 |
| | SCA-IEDP | **63.4** | **65.6** | **61.2** | **0.60** |

and SCA are approximately 9-13% higher than that of SMLRAT. In NB and DLDA, SCA-IEDP provides respectively the best accuracy of 62.4% and 62.5%, while the best SEN of 80% and 98.5% are obtained via the SMLRAT model. In these classifiers, most segments are recognized as IEDs through the SMLRAT model, meaning that the model is biased towards the IED class. On the other hand, SCA presents the best SPEC when NB, DLDA, and SVM classifiers are employed, which means that the model is biased towards the non-IED class. Generally, the best trade-off between SEN and SPEC is made using the DTE classifier, in which SCA-IEDP outperformed SCA and SMLRAT in all criteria. In addition, the performance of SMLRAT is around the chance level (which is 50% for binary classification).

## 5.5 DISCUSSION

SCA-IEDP, as the main contribution of this study, outperforms SCA and the compared TF and SMLRAT methods. Meanwhile, in both SCA and SCA-IEDP, we apply CPD and employ spectrogram to extract TF features except that in SCA-IEDP we allocate weight to each IED according to the certainty in its labelling, SCA-IEDP performs significantly better compared to SCA in both within- and between-subject classification approaches. SCA-IEDP is far superior, in terms of sensitivity, to SCA in detecting IEDs in both approaches. This shows that incorporating the IED probabilities in designing an IED detection model can boost performance.

The IEDs with lower uncertainty have more impact on the learning process than those with higher uncertainty. Some brain activities are similar in morphology to IEDs and

sometimes are recognized as IED waveforms by epileptologists. In the proposed SCA-IEDP, the IEDs with higher uncertainty are given lower weights. As a result, the impact of these IEDs in the classification step decreases, leading to better model performance. In contrast, the proposed SCA and the compared methods do not have this advantage. Indeed, there has not been any method that incorporates the uncertainty in labelling the IEDs in their detection, and as far as we know, the proposed SCA-IEDP is the first approach.

The proposed models show superior performance when they are trained and tested over the same subject data (within-subject classification approach) compared to a more generic or between-subject classification approach. However, detecting IEDs in a between-subject-based approach is of paramount importance. Developing a model to automatically detect the IEDs of new subjects without training on them would be worthwhile in clinical practice. Our proposed algorithms enable the detection of IEDs in a between-subject-based approach, and their performances are sufficiently desirable.

Generally, none of the methods performs very well because we use low amplitude (scalp-visible and scalp-invisible) IEDs. However, a small improvement can be significant for many applications, such as seizure prediction or localization. Although some studies have reported higher performance, they only consider scalp-visible IEDs [26, 74]. Here, iEEG has been used for scoring the IEDs. The large proportion of IEDs seen in iEEG is invisible over the scalp. Our dataset includes all scalp-visible and scalp-invisible IEDs thanks to using concurrent iEEG signals for the IED annotation, causing the traditional automatic algorithms not to be effective.

In contrast to the TF model [8] in which the authors used only temporal and spectral IED signatures, we consider spatial IED signatures apart from temporal and spectral ones. The IEDs originate from specific brain regions. Our dataset is from patients with mesial temporal lobe epilepsy whose IEDs originate from the temporal lobe regions. Therefore, intracranial electrodes placed in the temporal lobes and, consequently, the scalp electrodes over those regions can provide more significant features. As a result, considering spatial components in designing an IED detection system can ameliorate its performance. Muti-way analysis provides an opportunity to consider spatial components and consequently boosts the performance of an IED detection system.

In the original TF study, IEDs are detected with 67% accuracy using the logistic regression classifier in the within-subject classification approach using the same dataset. Here, the TF method obtains 73% accuracy using the DTE classifier. In our TF implementation, we use a different frequency band, different IED and non-IED segment lengths, and different classifiers. These differences ameliorate the model performance.

In the referenced SMLRAT method, 95.8% accuracy is obtained using an NB classifier. In our study, SMLRAT respectively provides the maximum accuracy of 72.9% and 50.9% in the within- and between-subject approaches, which are significantly less than the accuracies obtained in the original paper. The main cause of this fall-off in accu-

racy may be due to including scalp-invisible IEDs in the classification. Here, all the scalp-visible and scalp-invisible IEDs are included in our dataset. On the other hand, in the referenced SMLRAT, the authors annotated the spikes using only sEEG, and only scalp-visible spikes were detected. Furthermore, they detected only spikes, while our IED dataset also contains sharp waves – which are wider than spikes and sometimes similar to other brain activities. For SMLRAT, there is a huge difference in the accuracy of within- and between-subject approaches. In the between-subject approach, we do not introduce any new mode to the tensor (motivated by the referenced SMLRAT). In fact, the IED three-way tensors of all the subjects are concatenated into a single four-way tensor (time, wavelet-scale, channel, and segment) in the between-subject approach. Since the morphology and shape of IEDs can vary among subjects because of factors such as age [49], including all IEDs along one tensor mode can deteriorate the decomposition performance. Therefore, SMLART gave a poor performance for our dataset in the between-subject approach. Hence, we added a new mode of subjects leading to a four-way tensor (time, channel, IED segment, and subject) for the between-subject approach, whereas we have a three-way tensor (time, channel, and IED segment) for the within-subject approach.

The most important advantage of the proposed SCA-IEDP method is to incorporate uncertainty levels in the IED scoring in the algorithm. This uncertainty can be mathematically explained by assigning lower probability values to IEDs with higher uncertainty and vice versa, making the impact of certain IEDs become more on the IED detection system.

The limitation of the proposed models is that their performance in the between-subject-based approach is not significantly high. In our dataset, the IEDs of different subjects originate from the left or right temporal region. This means the IED sources' locations are different among the subjects. Since our proposed models are based on spatial components and the IED source locations are different, the performance of the proposed models deteriorates when a between-subject-based approach is applied. Another cause of this fall-off in the performance of the between-subject-based approach is that the consistency of discharges decreases when all patients are pooled together. These limitations exist in the SMLRAT method as well.

## 5.6 CONCLUSION

Improvements in the identification of deep epileptiform discharges from over the scalp would significantly enhance the diagnosis and management of epilepsy. In this chapter, CP-OPT and CP-WOPT techniques are first explained. Then, we propose two models, namely SCA and SCA-IEDP, for IED detection. In SCA-IEDP, as the main contribution of this work, the uncertainty in IED labeling (represented by IED probabilities) is in-

corporated in the IED detection system. We compare our proposed models with two state-of-the-art models, TF [8] and SMLRAT [10]. The IEDs are detected in two different within- and between-subject classification approaches. In both approaches, SCA-IEDP leads to the best performance. It obtains respectively the maximum accuracy of 79.9% and 63.4% in the within- and between-subject classification approaches. Meanwhile, SMLRAT achieves respectively 72.9% and 50.9% accuracy values in the mentioned approaches. In all classifiers, TF and SMLRAT perform worse than our proposed SCA and SCA-IEDP. The results show that incorporating the IED probabilities into the algorithm can improve performance. Here, we use concurrent sEEG and iEEG recordings. The major advantage of our work in using this dataset is to detect the scalp-visible and scalp-invisible IEDs from over the scalp. In a within-subject classification approach, it may not be very useful to record sEEG and iEEG simultaneously and detect IEDs using sEEG, whereas, in a between-subject classification approach, it has a high impact on the diagnosis of epilepsy. That is, a between-subject classifier can be trained using concurrent scalp and intracranial recordings of several subjects (the use of iEEG for scoring IEDs – including the scalp-visible and scalp-invisible IEDs – and the use of sEEG for training the classifier). Afterward, using the trained classifier, the scalp-visible and scalp-invisible IEDs of any new subject can be detected using only sEEG recordings.

## 5.7 AKNOWLEDGE

Electronic version of an article published as "Incorporating Uncertainty in Data Labeling into Automatic Detection of Interictal Epileptiform Discharges from Concurrent Scalp-EEG via Multi-way Analysis," Vol. 31, 2021, pp. 2150019, doi: 10.1142/S0129065721500192

# 6

# SPARSE COMMON FEATURE ANALYSIS FOR DETECTION OF IEDS FROM CONCURRENT SCALP EEG

Common feature analysis (CFA), explained in Section 6.1, has the benefits of exploiting and analyzing common features among different data dimensions. To the best of our knowledge, there exists no IED detection method to extract the common components across all IED segments in the most discriminatory time-space domain only. Therefore, we aim to present two models based on CFA and sparse CFA to detect the IEDs from sEEG using a unique limited set of concurrent sEEG and iEEG recordings.

In this study, we consider that the IED segments for each subject are naturally linked and share spatially and temporally some common features. These common features, which are latent in EEGs, may reflect more accurately the IEDs characteristics. Zhou *et al.* [122] developed an algorithm, namely common orthogonal basis extraction (COBE), for extracting common and individual features to boost image classification performance. We adopt the COBE algorithm to exploit the latent common features among the IED segments to enable the detection of a higher percentage of IEDs from over the scalp using a unique set of simultaneously recorded sEEG and iEEG signals. This method is referred to as the CFA-based method for IED detection. In the second approach, as the main contribution of this chapter, we extend the COBE algorithm to exploit the common features with a sparsity constraint, referred to as sparse common orthogonal basis extraction (SCOBE). Using our developed SCOBE algorithm, we extract common features among the IED segments with sparsity constraints (sparse common features). This method is called the sparse CFA (SCFA)-based method for IED detection. It should be noted that, in our dataset, the IEDs are scored from the iEEG by an expert clinician while detected from the sEEG. This provides an opportunity to automatically detect the scalp-invisible IEDs from sEEG, which is not feasible in the sEEG-based algorithms for IED detection. For classification, three types of classifiers, namely SVM, DLDA, and NB, are employed.

The rest of the chapter is structured as follows. First, CFA and SCFA models are described. The results are then reported. Afterwards, the findings are discussed. Finally, the last section concludes the chapter.

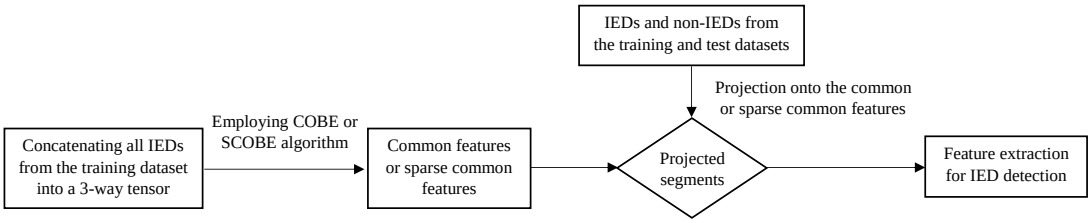



Figure 6.1: The flowchart of the proposed methods representing the overall IED detection system.

## 6.1 COMMON AND SPARSE COMMON FEATURE ANALYSIS

The IEDs are associated with abnormal patterns; thus, it can be assumed that they are independent of other brain activities. Moreover, they have many similarities in shape and morphology; therefore, some features are expected to be shared among them. In contrast, the non-IEDs are random, with no shared feature between them. Therefore, we are interested in a feature space that spans the IEDs only.

In the proposed CFA and SCFA methods, common features and sparse common features among the IED segments are exploited respectively by COBE and SCOBE algorithms. Then, both IED and non-IED segments are projected onto them using the Khatri-Rao product. Finally, the features of projected segments are extracted for classification. Figure 6.1 shows the flowchart of the proposed methods representing the overall IED detection system. The details of the methods are explained in the following subsections.

### 6.1.1 *Common Feature Analysis*

In the proposed CFA model, the common features are exploited using the COBE algorithm [122], explained below.

#### 6.1.1.1 *COBE*

Suppose the training dataset consists of $N$ IED segments, $\boldsymbol{\mathcal{X}} = \{\mathbf{X}_n \in \mathbb{R}^{L\times M} : n \in \mathcal{N}\}$, $\mathcal{N} = \{1, 2, ..., N\}$, where $L$ and $M$ are respectively the number of time samples and channels. Our goal is to extract the common features among all IED segments. According to matrix factorisation solution, for each matrix $\mathbf{X}_n$, we attempt the following minimization:

$$\min_{\mathbf{S}_n, \mathbf{W}_n} \|\mathbf{X}_n - \mathbf{S}_n \mathbf{W}_n^T\|_F^2, \quad n \in \mathcal{N}, \tag{6.1}$$

where the columns of $\mathbf{S}_n \in \mathbb{R}^{L\times P_n}$ denote the sources in $\mathbf{X}_n$, $\mathbf{W}_n \in \mathbb{R}^{M\times P_n}$ indicates the corresponding mixing matrix, and $\|.\|_F$ is the Frobenious norm operator. It is assumed that $P_n < \min(L, M)$, implying that $\mathbf{S}_n\mathbf{W}_n^T$ presents a low-rank representation of $\mathbf{X}_n$.

The sources of the data ensemble $\mathbf{X}_n$ are correlated together, thereby sharing some common features. Hence, we can define $\mathbf{S}_n$ as follows:

$$\mathbf{S}_n = [\bar{\mathbf{S}} \;\; \check{\mathbf{S}}_n], \quad n \in \mathcal{N}, \tag{6.2}$$

where the sub-matrix $\bar{\mathbf{S}} \in \mathbb{R}^{L\times C}$ consists of common features shared by all the matrices in $\boldsymbol{\mathcal{X}}$, and the sub-matrix $\check{\mathbf{S}}_n \in \mathbb{R}^{L\times(P_n-C)}$, $C \leq \min\{P_n : n \in \mathcal{N}\}$, presents the individual sources of each $\mathbf{X}_n$. By doing so, we are able to re-factorize the data matrices $\mathbf{X}_n$ in an augmented way as:

$$\begin{aligned}\mathbf{X}_n = \mathbf{S}_n\mathbf{W}_n^T &= \begin{bmatrix}\bar{\mathbf{S}} & \check{\mathbf{S}}_n\end{bmatrix}\begin{bmatrix}\bar{\mathbf{W}}_n^T \\ \check{\mathbf{W}}_n^T\end{bmatrix} \\ &= \bar{\mathbf{S}}\bar{\mathbf{W}}_n^T + \check{\mathbf{S}}_n\check{\mathbf{W}}_n^T = \bar{\mathbf{X}}_n + \check{\mathbf{X}}_n, \quad n \in \mathcal{N},\end{aligned} \tag{6.3}$$

where $\bar{\mathbf{W}}_n$ and $\check{\mathbf{W}}_n$ consist of the mixing matrices corresponding to $\bar{\mathbf{S}}$ and $\check{\mathbf{S}}_n$, respectively.

There are numerous solutions to minimization of (6.1), which are not unique. To reduce the solution space, the following three constraints are applied:

1. $\bar{\mathbf{S}}^T\bar{\mathbf{S}} = \mathbf{I}_C$

2. $\check{\mathbf{S}}_n^T\check{\mathbf{S}}_n = \mathbf{I}_{P_n-C}$

3. There is no interaction (correlation) between the spaces of common and individual features, i.e., $\bar{\mathbf{S}}^T\check{\mathbf{S}}_n = \mathbf{o}$

By substituting (6.3) in (6.1) and considering the above constraints, (6.1) can be reformulated to:

$$\begin{aligned}\min_{\bar{\mathbf{S}},\bar{\mathbf{W}}_n,\check{\mathbf{S}}_n,\check{\mathbf{W}}_n} \quad & \sum_{n\in\mathcal{N}} \|\mathbf{X}_n - \bar{\mathbf{S}}\bar{\mathbf{W}}_n^T - \check{\mathbf{S}}_n\check{\mathbf{W}}_n^T\|_F^2, \\ s.t. \quad & \bar{\mathbf{S}}^T\bar{\mathbf{S}} = \mathbf{I}_C, \;\; \check{\mathbf{S}}_n^T\check{\mathbf{S}}_n = \mathbf{I}_{P_n-C}, \;\; \bar{\mathbf{S}}^T\check{\mathbf{S}}_n = \mathbf{o}, \;\; n \in \mathcal{N},\end{aligned} \tag{6.4}$$

where the notation $\mathbf{o}$ denotes a $C \times (P_n - C)$ zero matrix.

There is a close relationship between the factorisation problem (6.4) and PCA when $P_n = C, \forall n \in \mathcal{N}$. In this case, $\bar{\mathbf{S}} = \mathbf{S}$ can be found from

$$\min_{\mathbf{S}} \; \textstyle\sum_{n\in\mathcal{N}} \|\mathbf{X}_n - \mathbf{S}\mathbf{W}_n^T\|_F^2, \quad s.t. \;\; \mathbf{S}^T\mathbf{S} = \mathbf{I}_C. \tag{6.5}$$

Problem (6.5) can be considered as a partitioned version of the global PCA of $\tilde{\mathbf{X}}$ when the data matrices $\mathbf{X}_n$ are stacked to construct a global matrix $\tilde{\mathbf{X}} = [\mathbf{X}_1 \quad \mathbf{X}_2 \quad \cdots \quad \mathbf{X}_N]$ and similarly $\tilde{\mathbf{W}} = [\mathbf{W}_1 \quad \mathbf{W}_2 \quad \cdots \quad \mathbf{W}_N]$, that is

$$\min_{\mathbf{S}} \|\tilde{\mathbf{X}} - \mathbf{S}\tilde{\mathbf{W}}^T\|_F^2, \quad s.t. \quad \mathbf{S}^T\mathbf{S} = \mathbf{I}_C. \tag{6.6}$$

However, the factorisation problem (6.4) is not equivalent to PCA when $C < P_n$. The main difference between problems (6.4) and (6.5) is owing to the individual components $\check{\mathbf{S}}_n\check{\mathbf{W}}_n^T$, meaning that the common components found by (6.4) can be interpreted as the principal components of the common subspace $\bar{\mathbf{X}}_n = \mathbf{X}_n - \check{\mathbf{S}}_n\check{\mathbf{W}}_n^T$. For more details, the reader is referred to [122].

To solve (6.4), finding the common features $\bar{\mathbf{S}}$ plays a vital role. From (6.3), we have:

$$[\bar{\mathbf{S}} \quad \check{\mathbf{S}}_n] = \mathbf{X}_n\mathbf{W}_n^{T\dagger}, \quad \mathbf{S}_n^T\mathbf{S}_n = \mathbf{I}_{P_n}, \quad n \in \mathcal{N}, \tag{6.7}$$

where $(.)^\dagger$ denotes Moore-Penrose pseudo-inverse of a matrix. To estimate $\bar{\mathbf{S}}$, we can employ QR decomposition to decompose $\mathbf{X}_n = \mathbf{Q}_n\mathbf{R}_n$, where $\mathbf{Q}_n$ is an orthogonal and $\mathbf{R}_n$ is an upper triangular matrix. By defining $\mathbf{Z}_n = \mathbf{R}_n\mathbf{W}_n^{T\dagger}$, (6.7) can be reformulated to:

$$[\bar{\mathbf{S}} \quad \check{\mathbf{S}}_n] = (\mathbf{Q}_n\mathbf{R}_n)\mathbf{W}_n^{T\dagger} = \mathbf{Q}_n\mathbf{Z}_n, \quad n \in \mathcal{N}. \tag{6.8}$$

Therefore, for any given $n_1, n_2 \in \mathcal{N}$, $n_1 \neq n_2$, we have:

$$\begin{cases} \mathbf{Q}_{n_1}\mathbf{z}_{n_1,k} = \mathbf{Q}_{n_2}\mathbf{z}_{n_2,k} = \bar{\mathbf{s}}_k, & \text{if } k \leq C; \\ \mathbf{Q}_{n_1}\mathbf{z}_{n_1,k} \neq \mathbf{Q}_{n_2}\mathbf{z}_{n_2,k} & \text{if } k > C, \end{cases} \tag{6.9}$$

where $\mathbf{z}_{n,k}$ and $\bar{\mathbf{s}}_k$ are respectively the $k$th columns of $\mathbf{Z}_n$ and $\bar{\mathbf{S}}$. It should be noted that the condition (6.9) is valid when there is a similarity among all the $\mathbf{X}_n$ segments (and consequently $\mathbf{Q}_n$). Due to this fact, our $\mathbf{X}_n$ consists of only IED segments, falling in the same frequency range having similar morphologies (i.e. spkies and sharp waves). From (6.9), we can compute the first column of $\bar{\mathbf{S}}$, signified by $\bar{\mathbf{s}}_1$, by minimizing

$$\min_{\bar{\mathbf{s}}_1, \mathbf{z}_{n,1}} \quad J_1 = \sum_{n \in \mathcal{N}} \|\mathbf{Q}_n\mathbf{z}_{n,1} - \bar{\mathbf{s}}_1\|_2^2, \quad s.t. \quad \bar{\mathbf{s}}_1^T\bar{\mathbf{s}}_1 = 1, \tag{6.10}$$

where $\|.\|_2$ denotes $l_2$-norm. Based on (6.8) and (6.9), the objective function $J_1$ has to be very small (very close to zero) to ensure that $\bar{\mathbf{s}}_1$ is a common basis vector among the trials.

An alternating least-square (ALS) optimization algorithm can be utilized to minimize (6.10). First, by fixing $\mathbf{z}_{n,1}$, the optimal $\bar{\mathbf{s}}_1$ is obtained by

$$\bar{\mathbf{s}}_1 = \sum_n \mathbf{Q}_n \mathbf{z}_{n,1}, \tag{6.11}$$

which is then normalized to have a unit norm. Repeating for a fixed $\bar{\mathbf{s}}$, we calculate $\mathbf{z}_{n,1}$ as

$$\mathbf{z}_{n,1} = \mathbf{Q}_n^T \bar{\mathbf{s}}_1, \quad n \in \mathcal{N}, \tag{6.12}$$

and repeat until convergence. For the proof of convergence of ALS algorithm see [123]. The vector $\bar{\mathbf{s}}_1$ is considered to be a common basis vector as long as $min\ J_1 \leq \epsilon$ for a very small threshold $\epsilon \geq 0$; otherwise, there is no common feature among the trials and iterations (6.11) and (6.12) stop.

Given the estimated set of common basis vectors, $[\bar{\mathbf{s}}_1, \bar{\mathbf{s}}_2, \ldots, \bar{\mathbf{s}}_k]$, it needs to be ensured that the new sought vector $\bar{\mathbf{s}}_{k+1}$ is not repeated. We can achieve this by considering the following property of $\mathbf{Z}_n$. Suppose $\mathbf{Z}_{n,C} = [\mathbf{z}_{n,1} \ \ \mathbf{z}_{n,2} \ \ \cdots \ \ \mathbf{z}_{n,C}]$, then according to (6.8) we have:

$$\mathbf{Z}_{n,C}^T \mathbf{Z}_{n,C} = \mathbf{Z}_{n,C}^T \mathbf{Q}_n^T \mathbf{Q}_n \mathbf{Z}_{n,C} = \bar{\mathbf{S}}^T \bar{\mathbf{S}} = \mathbf{I}_C. \tag{6.13}$$

This means $\mathbf{z}_{n,k+1}^T \mathbf{z}_{n,k} = 0$ and $\mathbf{z}_{n,k+1}$ is the null space of $\mathbf{z}_{n,k}^T$, allowing us to update $\mathbf{Q}_n$ as

$$\mathbf{Q}_n^{(k+1)} = \mathbf{Q}_n^{(k)} (\mathbf{I} - \mathbf{z}_{n,k} \mathbf{z}_{n,k}^T). \tag{6.14}$$

Finally, this leads to finding $\bar{\mathbf{s}}_{k+1}$ through minimizing the following objective function:

$$\min_{\bar{\mathbf{s}}_{(k+1)}, \mathbf{z}_{n,k+1}} J_{k+1} = \sum_{n \in \mathcal{N}} \|\mathbf{Q}_n^{(k+1)} \mathbf{z}_{n,k+1} - \bar{\mathbf{s}}_{k+1}\|_2^2, \qquad s.t. \quad \bar{\mathbf{s}}_{k+1}^T \bar{\mathbf{s}}_{k+1} = 1. \tag{6.15}$$

The ALS algorithm is repeated till $J_{k+1}$ is minimized.

#### 6.1.1.2 *IED detection based on CFA*

In the proposed CFA-based method for IED detection, we employ the COBE algorithm to extract the common basis vectors $\bar{\mathbf{S}} \in \mathbb{R}^{L \times C}$ among the IEDs. Then, both IED and non-IED segments $\mathbf{X}_k \in \mathbb{R}^{L \times M}$ are projected onto the extracted vectors using Khatri-Rao product as follows:

$$\mathbf{P}_k = \bar{\mathbf{S}}^T \odot \mathbf{X}_k^T, \tag{6.16}$$

for $k = 1, \ldots, K$, where $K$ is the total number of both IEDs and non-IEDs in the training and test data, the symbol '$\odot$' denotes Khatri-Rao product, $\mathbf{X}_k$ is an IED or non-IED segment, and $\mathbf{P}_k \in \mathbb{R}^{(MC)\times L}$ represents the same segment after projection. The epileptiform spikes (whether scalp-visible or scalp-invisible IEDs) have similar behavior. Most channels have the same trends, including a sharp excitatory and a damped inhibitory oscillation during the spike onsets. In addition, these trends are similar to the common basis vector trend. By Kharti-Rao product, the time samples of each channel are separately elementwise multiplied by each of the common vectors $[\bar{\mathbf{s}}_1, \bar{\mathbf{s}}_2, \ldots, \bar{\mathbf{s}}_C]$. Therefore, the epileptiform spikes (or the background activities of scalp-invisible IEDs) are magnified by projection. On the other hand, since there is no common feature among the non-IEDs, this projection has no significant impact on them. Furthermore, the kurtoses of projected segments are extracted for being used as the classification features. This strength in the amplitude leads to an increase in the IED kurtosis, while does not significantly affect the non-IED kurtosis.

### 6.1.2 *Sparse Common Feature Analysis*

During the past decade, sparse representation has attracted much attention in various signal processing areas, including epilepsy study [124, 125]. The train of spikes emitted from individual neurons in the brain can be considered sparse in some domains, such as time and space. One of the interesting characteristics of an IED is its sparsity in the time domain. The original COBE algorithm does not exploit this property, making it inefficient for spike detection. Therefore, we develop a new algorithm, namely SCOBE, with a sparsity constraint to exploit the common features. Then, we propose a model based on sparse common features called SCFA to detect the IEDs.

#### 6.1.2.1 *SCOBE*

This approach extracts the common basis vectors with a sparsity constraint. In other words, the number of non-zero elements of each basis vector is sparsified. To this end, the sparsity condition is incorporated into (6.10) to change it into a constrained problem as follows:

$$\begin{aligned} \min_{\mathbf{D},\mathbf{a}_1,\mathbf{z}_{n,1}} \quad & J_1 = \textstyle\sum_{n\in\mathcal{N}} \|\mathbf{Q}_n\mathbf{z}_{n,1} - \mathbf{D}\mathbf{a}_1\|_2^2 \\ & s.t. \quad \|\mathbf{a}_1\|_0 \leq T_0, \quad (\mathbf{D}\mathbf{a}_1)^T(\mathbf{D}\mathbf{a}_1) = 1. \end{aligned} \tag{6.17}$$

for $\bar{\mathbf{s}}_1 = \mathbf{D}\mathbf{a}_1$, where $\mathbf{D} \in \mathbb{R}^{L\times F}$ is the dictionary (whose columns are the atom signals), $\mathbf{a}_1 \in \mathbb{R}^F$ includes the first sparse representation vector of the signals, $\|.\|_0$ denotes $l_0$-norm which accounts for the number of non-zero entries, and $T_0$ is a small threshold set empirically, here it is selected to be 8. However, (6.17) is an NP-hard problem but can be

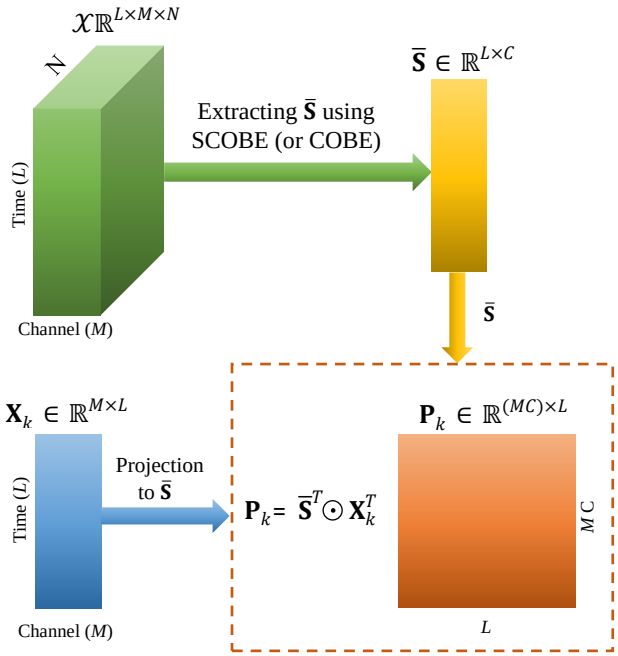



Figure 6.2: The proposed SCFA-based (or CFA-based) model for IED detection. $\mathcal{X}$ includes the IED segments ($N$) from the training set only. $\bar{\mathbf{S}}$ denotes the sparse common basis vectors (or common basis vectors) extracted by applying the SCOBE (or COBE) algorithm to $\mathcal{X}$. $\mathbf{X}_k$ can be an IED or non-IED segment from the training and test datasets and $\mathbf{P}_k$ represents the same segment after projection. The notation '$\odot$' denotes Khatri-Rao product.

efficiently solved using several available approximation techniques such as orthogonal matching pursuit (OMP) [126, 127] and basis pursuit [128, 129].

Apart from $\mathbf{D}$ and $\mathbf{a}_1$, we need to minimize $\mathbf{z}_{n,1}$, thereby employing ALS iteration. Suppose $\mathbf{D}\mathbf{a}_1$ is fixed, $\mathbf{z}_{n,1}$ is computed as:

$$\mathbf{z}_{n,1} = \mathbf{Q}_n^T \mathbf{D}\mathbf{a}_1, \quad n \in \mathcal{N}. \tag{6.18}$$

Then, by keeping $\mathbf{z}_{n,1}$, we have

$$\min_{\mathbf{D},\mathbf{a}_1} \; J_1 = \|\mathbf{y} - \mathbf{D}\mathbf{a}_1\|_2^2, \quad s.t. \quad \|\mathbf{a}_1\|_0 \leq T_0, \tag{6.19}$$

where $\mathbf{y} = \sum_{n\in\mathcal{N}} \mathbf{Q}_n \mathbf{z}_{n,1}$. To optimize the objective function (6.19), the OMP technique [127] is used to approximate the sparsity and the K-SVD algorithm [130] to train the dictionary $\mathbf{D}$. In terms of stability, it should be noted that in our proposed method the objective (6.19) is a matrix-based problem optimized by the OMP technique. The stability of OMP has been proven in [131].

The OMP algorithm is an iterative algorithm that finds the sparse vector $\mathbf{a}_1$ element-by-element in a step-by-step iterative manner. In this algorithm, the atom $\mathbf{d}_f$ – $f$th column of $\mathbf{D}$ – with the highest correlation to the current residue, denoted by $\mathbf{r}$, is selected at each step; $\hat{f} := \underset{f}{\operatorname{argmax}} \; |\mathbf{d}_f^T \mathbf{r}|$. Once the atom is selected, the signal is

Algorithm 6.1: The SCOBE pseudocode.

**Input**: $\mathbf{X}_n, n \in \mathcal{N}, \epsilon \geq 0$
Decompose $\mathbf{X}_n$=$\mathbf{Q}_n\mathbf{R}_n$ *s.t.* $\mathbf{Q}_n^T\mathbf{Q}_n$=$\mathbf{I}_M, n \in \mathcal{N}$.
**Train** a dictionary $\mathbf{D}$ using the K-SVD algorithm.
$\bar{\mathbf{S}} = [\,], \mathbf{Q}_n^{(1)} = \mathbf{Q}_n$ and $k = 1$.
**while** $J_k \leq \epsilon$ **do**
    **Initialize** $\mathbf{a}_k$ randomly and sparsely with a unit norm.
    **while** not converged **do**
        $\mathbf{z}_{n,k} = [\mathbf{Q}_n^{(k)}]^T\mathbf{D}\mathbf{a}_k,\ n \in \mathcal{N}$;
        $\mathbf{y}$=$\sum_{n\in\mathcal{N}} \mathbf{Q}_n\mathbf{z}_{n,1}$
        Use any pursuit algorithm to calculate $\mathbf{a}_k$, through approximating the solution of $\min_{\mathbf{a}_k} \|\mathbf{y} - \mathbf{D}\mathbf{a}_k\|_2^2$, *s.t.* $\|\mathbf{a}_k\|_0 \leq T_0$;
        $\bar{\mathbf{s}}_k = {}^{\mathbf{D}\mathbf{a}_k}/_{\|\mathbf{D}\mathbf{a}_k\|_2}$;
    **end while**
    $J_k = \frac{1}{N}\sum_n \|\mathbf{Q}_n^{(k)}\mathbf{z}_{n,k} - \bar{\mathbf{s}}_k\|_2^2$;
    $\bar{\mathbf{S}} = [\bar{\mathbf{S}} \quad \bar{\mathbf{s}}_k]$;
    $k = k + 1$;
    $\mathbf{Q}_n^{(k)} = \mathbf{Q}_n^{(k-1)}(\mathbf{I} - \mathbf{z}_{n,k-1}\mathbf{z}_{n,k-1}^T), n \in \mathcal{N}$;
**end while**
**return** $\bar{\mathbf{S}} = [\bar{\mathbf{s}}_1 \ \ \bar{\mathbf{s}}_2 \ \ \cdots \ \ \bar{\mathbf{s}}_C]$, where $C = k - 1$.

orthogonally projected to the span of selected atoms $I$; $\mathbf{a}_{1_I}$:=$\mathbf{D}_I^{\dagger}\mathbf{y}$. After recalculating the residue, $\mathbf{r} = \mathbf{y} - \mathbf{D}_I\mathbf{a}_{1_I}$, the procedure repeats until meeting a stopping condition. For the implementation of the OPM algorithm, we utilize the Matlab toolbox provided by Rubinstein *et al.* [127].

The K-SVD algorithm trains a dictionary for sparse approximation through SVD [130]. The goal of the algorithm is to iteratively learn a dictionary to achieve the sparsest representations of the signals in $\Psi \in \mathbb{R}^{L\times T}$ by optimizing the following constrained objective function:

$$\min_{\mathbf{D},\Gamma} \|\Psi - \mathbf{D}\Gamma\|_F^2, \quad s.t. \quad \forall i \|\gamma_i\|_0 \leq K_0, \tag{6.20}$$

where $\Gamma \in \mathbb{R}^{F\times T}$ is the sparse representation matrix of the signals $\Psi$ using the dictionary $\mathbf{D} \in \mathbb{R}^{L\times F}$. $K_0$ is selected to be 4 in our study. It is important to note that $T \gg F \gg L$ and that the columns of $\mathbf{D}$ need to be normalized. At first, $\mathbf{D}$ is selected randomly from $\Psi$. Then, a sparse approximation algorithm, here OPM algorithm, is utilized to compute the sparse representation vectors $\gamma_i$ for each example $\psi_i$. For updating, at first, the group of examples using this atom, $\phi_f = \{i | 1 \leq i \leq F, \gamma_T^f \neq 0\}$, where $\gamma_T^f$ is the $f$th row in $\Gamma$, is defined, and the overall error matrix, $\mathbf{E}_f$, is approximated by

$$\mathbf{E}_f = \Psi - \sum_{b\neq f} \mathbf{d}_b\gamma_T^b. \tag{6.21}$$

The error is then limited to only the columns corresponding to $\phi_f$ as follows:

$$\mathbf{E}_f^{lim} = \mathbf{E}_f \Phi_f, \tag{6.22}$$

where $\Phi_f$ is a matrix of size $T \times |\phi_f|$, with ones on the $(\phi_f(i), i)$th entries and zeros elsewhere. Finally, the SVD decomposition is applied to $\mathbf{E}_f^{lim} = \mathbf{U}\Lambda\mathbf{V}$, and the dictionary column $\mathbf{d}_f$ is updated using the first column of $\mathbf{U}$ as well as the coefficient vector $\gamma_{lim}^f$ using the first column of $\mathbf{V}$ multiplied by $\Lambda(1,1)$, the first singular value of the diagonal matrix. This procedure is repeated until convergence. The reader is referred to [130] to see the proof of convergence. For implementation, we used the Matlab toolbox described in [127].

However, after solving (6.19), from (6.17) the first sparse common basis vector is obtained as:

$$\bar{\mathbf{s}}_1 = \mathbf{D}\mathbf{a}_1, \tag{6.23}$$

and normalized to have a unit norm. $\bar{\mathbf{s}}_1$ and $\mathbf{z}_{n,1}$ are iteratively and in an alternating manner computed. It should be noted that the condition of $min\ J_1 \leq \epsilon$ needs to be met for a very small threshold $\epsilon \geq 0$ for $\bar{\mathbf{s}}_1$ to be a sparse common basis vector among the trials.

In order to avoid repeating the sparse common basis vectors, we need to update $\mathbf{Q}_n$. Here, the property of $\mathbf{Z}_{n,C}^T \mathbf{Z}_{n,C} = \mathbf{I}$ is also verifiable like COBE, (6.13). Therefore, $\mathbf{Q}_n$ is updated through (6.14).

Finally, after computing $\mathbf{Q}_n^{(k+1)}$, the new sparse common basis vector is obtained by solving the following objective function:

$$\begin{aligned} &\min_{\mathbf{D},\mathbf{a}_{k+1},\mathbf{z}_{n,k+1}} J_{k+1} = \textstyle\sum_{n\in\mathcal{N}} \|\mathbf{Q}_n^{k+1}\mathbf{z}_{n,k+1} - \mathbf{D}\mathbf{a}_{k+1}\|_2^2 \\ &\quad s.t. \quad \|\mathbf{a}_{k+1}\|_0 \leq T_0, \quad (\mathbf{D}\mathbf{a}_{k+1})^T(\mathbf{D}\mathbf{a}_{k+1}) = 1, \end{aligned} \tag{6.24}$$

which can be minimized by repeating the procedure in solving (6.17). New sparse common basis vectors are considered the vectors which make $J_k$ smaller than a very small threshold $\epsilon$ ($J_k < \epsilon$). In other words, the number of common or sparse common components are determined by $\epsilon$. Accordingly, $\epsilon$ should be small enough to avoid extracting uncommon factors. The pseudo-code of SCOBE is illustrated in Algorithm 1.

To avoid confusion, it should be noted that the number of non-zero elements of each basis vector – not the number of basis vectors – is sparsified. From (6.17) and (6.19), it can be seen that the number of non-zero elements of vector $\mathbf{a}_c$, where $\bar{\mathbf{s}}_c = \mathbf{D}\mathbf{a}_c$, is sparsified.

#### 6.1.2.2 *IED detection based on SCFA*

In the proposed SCFA-based method for IED detection, the sparse common basis vectors among the IEDs from the training dataset are extracted using the developed SCOBE algorithm. After obtaining the sparse common basis vectors, the IEDs and non-IEDs from the training and test datasets are projected onto them using the Khatri-Rao product according to the procedure given in (6.16). The schematic diagram of CFA- and SCFA-based methods for IED detection is illustrated in Figure 6.2.

## 6.2 EXPERIMENTS

### 6.2.1 *Data Description*

We analyzed 20-minute EEG recordings from 18 subjects suffering from temporal lobe epilepsy. The dataset details are given in Section 3.1. IEDs were scored by an expert epileptologist based on the morphologies and spatial distributions of the observed waveforms from the iEEG.

In order to increase the SNR and avoid the 50 Hz power line (or national grid), a bandpass filter with cutoff frequencies of 4 Hz and 48 Hz is applied to the sEEG signals. In addition, contra-lateral (CL) reference method is employed as re-referencing method to the sEEG signals [132]. In CL, the right and left hemisphere electrodes are re-referenced to the right and left earlobe electrodes, respectively. In our work, "Z" electrodes are re-referenced to the average of the two earlobe electrodes.

For analysis and classification, the length of the segments with IED is selected to be 480 ms (96 samples) – 160 ms before and 320 ms after the peak positions marked as IED. The non-IED segments with 480 ms length are extracted from the time segments in which no scored IED exists, with no overlap with the IED segments. The number of non-IED segments is the same as the number of IED segments for each subject—the number of trials is summarized in Table 4.1. Then, both IED and non-IED segments are linearly detrended to alleviate the undesired drifts.

### 6.2.2 *Feature Extraction*

We construct $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{96\times 18\times N}$ (whose dimensions 96, 18, and $N$ correspond respectively to the time samples, scalp channels, and IED segments from the training dataset). Common (or sparse common) basis vectors $\bar{\mathbf{S}} \in \mathbb{R}^{96\times C}$, where $C$ is the number of vectors, are exploited by employing COBE (or SCOBE) to $\boldsymbol{\mathcal{X}}$. Then, both IED and non-IED segments from the training and test datasets $\mathbf{X}_k \in \mathbb{R}^{18\times 96}$ are projected onto the extracted vectors using the Khatri-Rao product (6.16), $\mathbf{P}_k = \bar{\mathbf{S}}^T \odot \mathbf{X}_k^T$. Finally, the kurtosis values

of components of the projected IEDs and non-IEDs $\mathbf{P}_k \in \mathbb{R}^{(18C)\times 96}$ are computed as the classification features.

Kurtosis is a statistical measure of whether the data are heavy-tailed or light-tailed and describes the shape of a distribution. For each component – here 18$C$ components – the kurtosis can be computed as:

$$\text{Kurtosis} = \frac{\sum_{l=1}^{96}(\mu_l - \bar{\mu})^4/96}{\sigma^4}, \tag{6.25}$$

where $\bar{\mu}$ and $\sigma$ are respectively the component mean and standard deviation. Each scalp IED or non-IED segment consists of $18 \times C$ features.

### 6.2.3 *Competing Models*

We compare the performance of our proposed models with those claimed by very recent publications in this area.

KURTOSIS FEATURES: We compare our proposed methods with a method in which the kurtosis features (KFs) are extracted by (6.25) from the raw data after preprocessing. The corresponding method is referred to as the KFs method. This method is selected for comparison mainly because, in our algorithm, we extract the kurtosis features from the projected segments. Therefore, it is helpful to see the effects of kurtosis on the feature space and IED detection performance.

TIME-FREQUENCY FEATURES: TF features are selected as a compared method. More information about how we obtain TF features is given in Section 5.3.5.1.

SIMULTANEOUS MULTILINEAR LOW-RANK APPROXIMATION OF TENSORS: The proposed models are compared with SMLRAT [10]. The model is summarized in Section 5.3.5.2.

### 6.2.4 *Feature Selection and Classification*

We utilize the Fisher score algorithm to find the significant features. Fisher score is explained in Section 4.2.5.

Classification of the IED and non-IED segments is performed following two approaches, namely within- and between-subject classification approaches. In the within-subject classification approach, an individual classifier is trained for each subject, and a k-fold (k=5) cross-validation is employed to validate the models. Increasing the number of folds does not change the outcome. In this approach, subjects 13, 14, and 16 are

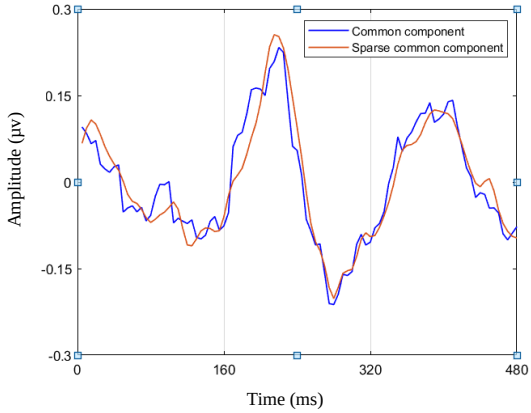



Figure 6.3: The first common and sparse common basis vectors extracted respectively using COBE and SCOBE from the IED segments of a subject. The IEDs start at 160ms.

excluded from classification because of having less number of trials; thus, the results of 15 subjects are reported. In the between-subject classification approach, one-subject-leave-out cross-validation is employed to validate the models. In other words, a subject is used as the test data, and other subjects (17) are employed to train a classifier. This is repeated for all the subjects.

In order to classify the IED and non-IED segments, we employed three different classifiers, namely SVM, DLDA, and NB. ACC, SEN, SPEC, and F1-S are obtained as the evaluation criteria.

## 6.3 EXPERIMENTAL RESULTS

The obtained results are presented in three sections. In Section 6.3.1, the extracted common components and the impact of their projection onto the IED segments are investigated. The within- and between-subject classification approaches results (performance $\pm$ standard error (SE)) are respectively reported in Section 6.3.2 and Section 6.3.3. DLDA, SVM, and NB classifiers are employed for classification. We use the first 36 significant features according to Fisher scores in CFA and SCFA, 18 significant features in KFs, 100 significant features in SMLRAT, and 200 significant features in TF. Those numbers of features give the highest accuracy in their models.

### 6.3.1 *Components and Projection*

The first common and sparse common basis vectors extracted respectively using COBE and SCOBE are illustrated in Figure 6.3. The sparse common basis vector not only is sharper but also has a higher amplitude than the common basis vector. Furthermore, the sparse common basis vector does not fluctuate as much as the common basis vector does.

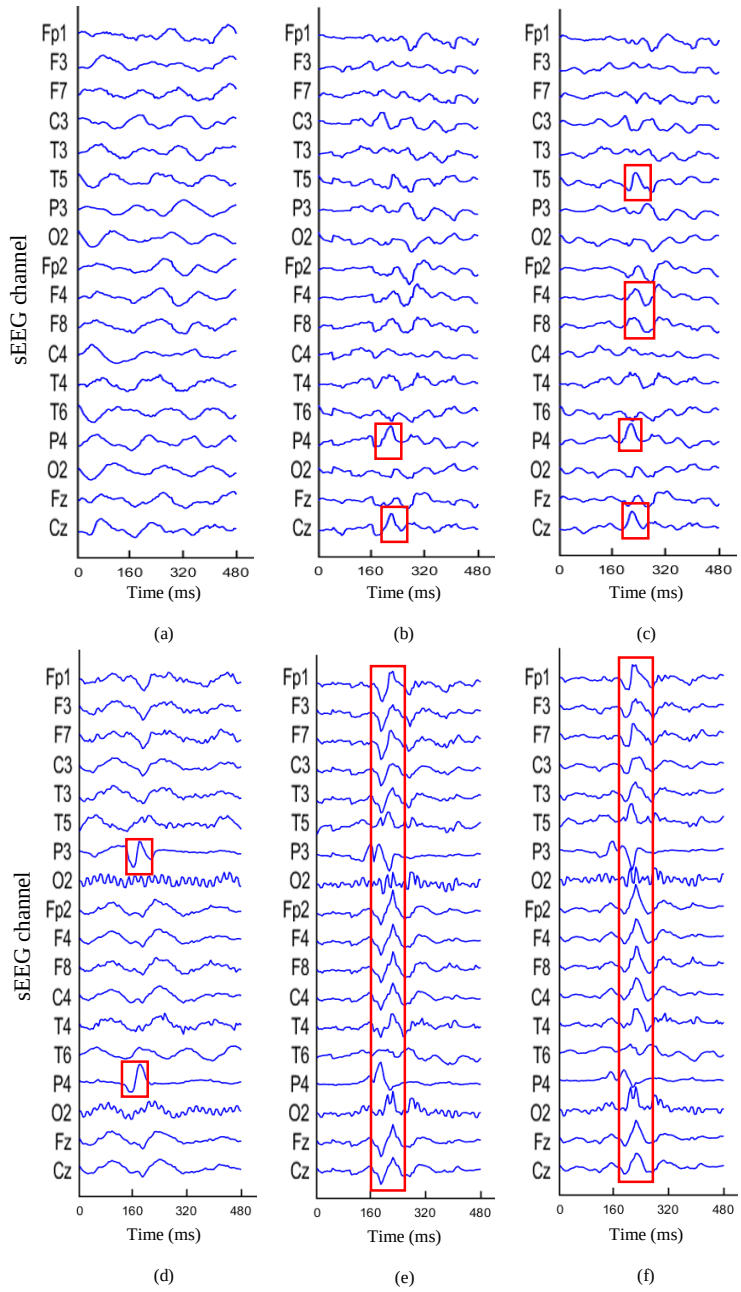



Figure 6.4: (a)-(c) show a non-IED segment before and after applying CFA and SCFA; (a) a non-IED segment before applying the CFA or SCFA, (b) after applying CFA, and (c) after applying SCFA. (d)-(f) present an IED segment before and after applying CFA and SCFA; (d) an IED segment before applying CFA or SCFA, (e) after applying CFA, and (f) after applying SCFA. The segments were projected onto the first common basis vector and sparse common basis vector only. The IED start at 160 ms. The region of interest is shown by red rectangles.

Figure 6.4 shows a non-IED and an IED segment before and after projection onto the common and sparse common basis vectors. Although slow waves appear over a few scalp channels after projecting the non-IED segment onto the first common and sparse common basis vectors (Figure 6.4 (b) and (c), respectively), strong spikes and sharp waves appear over all scalp channels after projecting the IED segment onto the

Table 6.1: The performances of classifiers and the performance SE are shown based on the within-subject approach with results averaged over 15 subjects and 5 folds. ACC, SEN, and SPEC are presented in percent (%).

| Classifier | Model | ACC | SEN | SPEC | F1-S |
|---|---|---|---|---|---|
| DLDA | TF | 70.7±1.7 | 61.0±2.2 | 80.3±3.0 | 0.66±0.02 |
| | SMLRAT | 64.9±2.0 | 46.4±3.7 | 83.4±2.9 | 0.54±0.04 |
| | KFs | 70.7±2.6 | 57.0±2.3 | 84.3±3.2 | 0.65±0.02 |
| | CFA | 73.2±2.3 | **64.1**±2.6 | 82.3±3.1 | **0.70**±0.02 |
| | SCFA | **74.2**±2.5 | 63.7±2.7 | **84.6**±2.9 | **0.70**±0.02 |
| SVM | TF | 66.9±2.3 | 65.2±2.2 | 68.2±2.6 | 0.66±0.02 |
| | SMLRAT | 72.9±2.7 | 63.8±2.5 | 82.0±3.5 | 0.69±0.03 |
| | KFs | 69.8±2.6 | 56.1±2.8 | **83.5**±2.6 | 0.64±0.03 |
| | CFA | 72.3±2.5 | 64.2±2.6 | 80.3±2.7 | 0.69±0.02 |
| | SCFA | **74.3**±2.4 | **68.0**±2.8 | 80.5±2.5 | **0.72**±0.02 |
| NB | TF | 70.5±1.9 | **68.0**±3.5 | 73.0±3.9 | 0.68±0.02 |
| | SMLRAT | 71.3±2.4 | 61.7±4.3 | 80.1±3.3 | 0.65±0.03 |
| | KFs | 71.3±2.7 | 54.4±4.2 | 88.3±1.6 | 0.63±0.04 |
| | CFA | 73.6±2.6 | 64.5±4.4 | 82.6±2.4 | 0.69±0.03 |
| | SCFA | **75.1**±2.6 | 65.5±4.0 | **84.7**±2.2 | **0.71**±0.03 |

first common and sparse common basis vectors (FIGURE 6.4 (e) and (f), respectively). Before projecting the IED, Figure 6.4 (d), the IED waveforms are observable only over channels P3 and P4. After projecting the IED segment onto the common basis vector obtained using COBE, Figure 6.4 (e), IED waveforms as spikes and sharp waves are observable over almost all the channels. After projecting the IED segment onto the sparse common basis vector, Figure 6.4 (f), the IED waveforms become sharper.

### 6.3.2 *IED Detection based on Within-Subject Classification Approach*

Both scalp-visible and scalp-invisible IEDs scored by an expert clinician from the iEEG recordings are detected from the sEEG recordings. The obtained IED detection results based on the within-subject classification approach are illustrated in Table 6.1.

In DLDA, SCFA outperforms other methods, providing the best performance with 74.2% accuracy, 63.7% sensitivity, 84.6% specificity, and 0.70 F1-score values. CFA presents the best sensitivity value of 64.1%, which is approximately 4%, 18%, and 7% more than TF, SMLRAT, and KFs sensitivity values, respectively. In SMLRAT, the DLDA classifier is biased to the non-IED class, meaning that most segments are recognized as non-IED segments.

Table 6.2: The performances of classifiers and the performance SE are shown based on the between-subject approach with results averaged over all 18 subjects. ACC, SEN, and SPEC are presented in percent (%).

| Classifier | Model | ACC | SEN | SPEC | F1-S |
|---|---|---|---|---|---|
| DLDA | SMLRAT | 57.1±1.6 | 48.7±4.9 | 65.5±5.6 | 0.50±0.03 |
| | KFs | 61.7±1.9 | 35.4±5.0 | **87.8**±3.8 | 0.45±0.04 |
| | CFA | 67.2±2.7 | 50.0±5.8 | 84.3±4.4 | 0.55±0.06 |
| | SCFA | **67.6**±2.6 | **54.5**±4.7 | 80.7±4.7 | **0.58**±0.05 |
| SVM | SMLRAT | 55.5±1.9 | 29.0±6.1 | 82.0±3.9 | 0.32±0.06 |
| | KFs | 61.6±2.1 | 32.8±5.1 | **90.4**±3.6 | 0.42±0.04 |
| | CFA | **67.3**±2.6 | **52.7**±6.7 | 82.0±4.3 | **0.57**±0.05 |
| | SCFA | 67.1±2.7 | 52.0±6.9 | 82.3±4.3 | 0.56±0.06 |
| NB | SMLRAT | 51.2±1.0 | 17.5±7.0 | 84.8±6.0 | 0.17±0.06 |
| | KFs | 64.8±3.0 | 35.1±6.6 | **94.5**±3.2 | 0.43±0.07 |
| | CFA | 66.3±2.7 | 40.5±6.4 | 92.1±3.4 | 0.49±0.06 |
| | SCFA | **67.8**±2.9 | **43.1**±6.5 | 92.5±3.5 | **0.52**±0.06 |

The best accuracy of SVM classifier is obtained using SCFA, which is 74.3%. Regarding SEN and F1-S, the SCFA model outperforms other methods as well. In terms of SPEC, KFs provide the best value. TF is the worst method versus all the criteria except in terms of SEN.

SCFA achieves the best accuracy of 75.1%, specificity of 84.7%, and F1-score of 0.71% using the NB classifier. TF obtains the best sensitivity of 68%. CFA classifies the IEDs and non-IEDs with 73.6% accuracy, which is higher than those by TF, SMLRAT, and KFs.

### 6.3.3 *IED Detection based on Between-Subject Classification Approach*

The obtained IED detection results based on the between-subject classification approach are shown in Table 6.2. Here, the performance of SMLRAT, KFs, CFA, and SCFA are reported. The TF method is not employed here in the between-subject classification approach. The authors of [8] use the TF method to detect IEDs in the within- and between-subject classification approach. However, in the between-subject classification approach, the authors train a classifier for each subject using the data of the same subject. Then, all the trained classifiers are combined to detect IEDs of a new subject. Since the data of different subjects are not combined in the TF model, it is not employed in this approach.

Using DLDA, SCFA achieves the best accuracy of 67.6%, which is approximately 10% and 6% more than the accuracy obtained respectively by SMLRAT and KFs. In addition, it presents the best sensitivity and F1-score values. In terms of specificity, KFs achieve the best value of 87.8%.

Using SVM, CFA provides the best accuracy of 67.3%, which is slightly more than the accuracy of SCFA, 67.1%. In terms of specificity, KFs achieve significantly better performance. However, KFs and SMLRAT are biased toward the non-IED class. While they achieve respectively the specificity values of 90.4% and 82%, they detect IEDs respectively with only 32.8% and 29% sensitivity values.

Using NB, SCFA outperforms CFA, KFs, and SMLRAT in all the criteria except specificity. It achieves the accuracy of 67.8%, which is respectively 1.5%, 3.0%, and 16.6% more than those of CFA, KFs, and SMLRAT. The SMLRAT model provides the worst performance with the sensitivity value of 17.5%. Generally, the NB classifier is biased toward the non-IED class in all the methods.

## 6.4 DISCUSSION

In both IED classification approaches, SCFA, a new sparse common feature analysis method, outperforms TF and SMLRAT (which is based on non-negative Tucker decomposition). The major advantage of our proposed models is that they extract the components in a trial-, subject-, and channel-independent-based approach, which enables the algorithms to effectively capture the background EEG activities and the intracranial biomarkers of epilepsy. Furthermore, SCFA outperforms CFA, while common components are extracted and used for classification in both. The only difference between CFA and SCFA is that SCFA exploits common components with sparsity constraints. This shows that our proposed algorithm, SCOBE, is superior to the plain COBE algorithm.

SMLRAT [10] has been reported to have a high performance where scalp-visible epileptic and non-epileptic spikes are detected with an accuracy as high as 95.8% from scalp recordings. In contrast, here, both scalp-visible and scalp-invisible IEDs are detected, causing a fall in the performance of all the methods. SMLRAT performs significantly better in the within-subject approach compared to the between-subject approach. The best accuracy obtained for the within-subject classification approach is 72.9%, while it is 57.1% for the between-subject classification approach. SMLRAT is based on spectral, temporal, and spatial components. In our dataset, the locations of IED sources are different among subjects. For each subject, the IEDs may be originated from the right, left, or both temporal lobes. That is, when the data of different subjects are combined, spatial components not only become meaningless but also deteriorate the performance of a classifier.

In many studies, IEDs are detected with high accuracy [26, 74], while in our study, the performances of both proposed and compared methods are not high. However, only scalp IEDs are included in their dataset, while a large proportion of IEDs is not observable from over the scalp. Therefore, they naturally ignore a large proportion of IEDs by default. Our dataset consists of both scalp-visible and scalp-invisible IEDs. The importance of this dataset is that we access concurrent iEEG and sEEG from the training subjects. After training a model, it can be employed to detect scalp-visible and scalp-invisible IEDs of a new subject from only the scalp recordings. The proposed SCFA and CFA methods detect respectively IEDs of new subjects with the accuracy values of 67.8% and 67.3%. This can bring a huge benefit to clinicians in monitoring epilepsy.

The only limitation of our approach is that the IED morphologies are different and also change with age [49]. Therefore, to generalize the application, we should have access to the data from a wider age range. In our another approach, we include the IED shape diversity into a higher dimensional tensor decomposition approach, presented in Chapter 4.

## 6.5 CONCLUSION

Automated detection of as many as possible IEDs from over the scalp is of paramount importance for epilepsy diagnosis and management. This is due to the fact that the majority of IEDs are invisible on the scalp, and a large proportion of IEDs can be missed out during the clinical examination. To overcome this deficiency, we effectively use a limited set of concurrent iEEG-sEEG recordings to design an algorithm that can be applied to the sEEGs only. In this work, we adopt the COBE algorithm proposed in [122] to extract the common components among the IEDs and then extend it to exploit the common features with sparsity constraints, called SCOBE. We propose two models, namely CFA and SCFA, based on COBE and SCOBE algorithms, respectively. We show that by employing the proposed models for the scalp-invisible IEDs, they become detectable from the sEEG signals. We have employed SVM, DLDA, and NB for classification and compared our proposed methods with two benchmark models, i.e., TF [8] and SMLRAT [10]. IEDs are detected via two different classification approaches, within- and between-subject classification approaches. The SCFA model outperforms other methods in both approaches and achieves respectively the best accuracy values of 75.1% and 67.8% using the NB classifier. These findings show that common component analysis can be very effective in capturing IED signatures, and exploiting the common components among IED segments with sparsity constraints is superior to exploiting the common components without any constraint.

## 6.6 AKNOWLEDGE

Electronic version of an article published as"Sparse Common Feature Analysis for Detection of Interictal Epileptiform Discharges From Concurrent Scalp EEG," in IEEE Access, vol. 10, pp. 49892-49904, 2022, doi: 10.1109/ACCESS.2022.3167433.

# 7

# HIGHER-ORDER TENSOR DECOMPOSITION BASED SCALP-TO-INTRACRANIAL EEG PROJECTION FOR IED DETECTION

The iEEG recordings enjoy high temporal and spatial resolution and hence a high sensitivity in detecting IEDs/spikes. Therefore, by mapping sEEG to iEEG through an effective projection model, the sensitivity of sEEG signals in the identification of IEDs can significantly increase. In a brain-computer interface application, the sEEG was modeled using a linear combination of iEEG via ordinary least-squares regression (LSR) [133]. Spyrou and Sanei [83] trained a dictionary and a mapping function to map the sEEG to iEEG to increase the quality of data. Quite recently, Antoniades et al. [9] developed a deep learning architecture using AE and CNNs to map the sEEG to iEEG to detect IEDs from concurrent sEEG and iEEG recordings. These studies have motivated us to develop a model to map the TF features of sEEG to the TF features of iEEG recordings to detect IEDs from concurrent sEEG and iEEG recordings. Our proposed method is based on tensor factorisation.

Tensor factorisation has been used in mapping models in image and video processing domains [134–136]. Macêdo, Brazil, and Velho [134] mapped a photographed expression performed by a given subject onto the photograph of another person's face by employing TD. In [135], the authors mapped the video-recorded performance of one person to the facial animation of another. Wang and Ahuja [136] used tensor decomposition to model and synthesize the facial expressions of new persons. They extracted facial features by employing principle component analysis and concatenated them onto a three-way tensor with the dimension of subject, expression, and feature, $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{I \times J \times K}$, where $I$ and $J$ are respectively the number of subjects and facial expressions, and $K$ indicates the dimension of the facial features. TD is employed to decompose the tensor into a core tensor and matrix factors: $\boldsymbol{\mathcal{X}} = \boldsymbol{\mathcal{S}} \times_1 \mathbf{U}^{person} \times_2 \mathbf{U}^{expression} \times_3 \mathbf{U}^{feature}$. Then, two tensors are defined for the expression and person modes:

$$\boldsymbol{\mathcal{T}}^{expression} = \boldsymbol{\mathcal{S}} \times_2 \mathbf{U}^{expression} \times_3 \mathbf{U}^{feature}, \tag{7.1}$$

$$\boldsymbol{\mathcal{T}}^{person} = \boldsymbol{\mathcal{S}} \times_1 \mathbf{U}^{person} \times_3 \mathbf{U}^{feature}. \tag{7.2}$$

Now, by having an expression of a new person, his or her other expressions can be synthesized by mapping the features of a given expression to the extracted expression

tensor in (7.1). Similarly, if an unknown expression of a known person is given, it can be mapped onto the extracted person tensor in (7.2) to synthesize the same expression of other persons.

Here, we propose a model to map the TF features of sEEG to those of iEEG signals via tensor factorisation techniques. The concurrent sEEG and iEEG recordings are utilized in this study. At first, both sEEG and iEEG recordings are decomposed by CWT to obtain the TF features. All IED segments of iEEG recordings from the training dataset are concatenated into a four-way tensor of time, frequency, channel, and IED segment. The tensor is then decomposed into temporal, spectral, spatial, and segmental factors using TD and CPD. Finally, both IED and non-IED segments of sEEG from both training and test datasets are projected onto the temporal components to extract the discriminative features for IED detection. Two types of classifiers, namely DTE and KNN, are then applied for classification.

The rest of this chapter is structured as follows. First, the mapping method is described. Then, the experimental matters and results are presented. The last section concludes the chapter.

## 7.1 SCALP-TO-INTRACRANIAL EEG PROJECTION

Here, we present a method based on tensor factorisation to map/project the TF features of sEEG to those of iEEG. The proposed method has three stages: 1) concatenating the TF features of iEEG into a tensor, 2) decomposing the constructed tensor, and 3) mapping the TF features of sEEG to those of iEEG recordings.

### 7.1.1 *Concatenating the TF Features of iEEG (Intracranial IEDs) into a Tensor*

Suppose our training dataset consists of $N$ IED segments from the concurrent iEEG recordings, $\mathbf{X}_n \in \mathbb{R}^{L_t \times M_{f_0}}$, $n = \{1, \cdots, N\}$, where $L_t$ is the number of time samples and $M_{f_0}$ corresponds to the number of FO channels. CWT is applied to $\mathbf{X}_n$, and the magnitude of wavelet coefficients is calculated as TF features. Then, the TF features of IEDs are concatenated into a four-way tensor, $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{L \times F \times M_{f_0} \times N}$, where $L$ and $F$ respectively correspond to the number of temporal and spectral features. In CWT, a complex Morlet wavelet is used as the mother wavelet. The wavelet toolbox from Matlab 2018b is utilized to calculate CWT. In the Matlab wavelet toolbox, the minimum and maximum scales are determined automatically based on the energy spread of the wavelet in frequency and time.

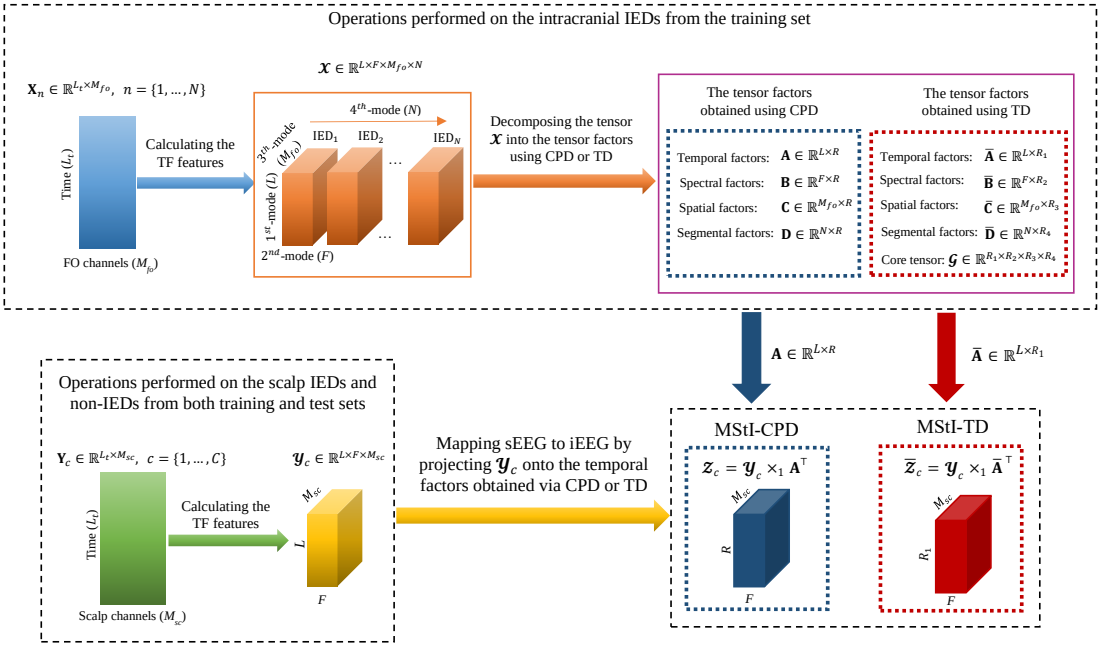



Figure 7.1: The schematic of the proposed MStI-CPD and MStI-TD methods.

### 7.1.2 *Decomposing the Constructed Tensor*

In the second step, the constructed tensor is decomposed into temporal, spectral, spatial, and segmental factors. TD and CPD, which are the most popular tensor decomposition techniques, are employed in this study.

CPD-BASED TENSOR DECOMPOSITION: The CPD algorithm is employed to decompose the tensor into the factor matrices as follows: temporal $\mathbf{A} \in \mathbb{R}^{L \times R}$, spectral $\mathbf{B} \in \mathbb{R}^{F \times R}$, spatial $\mathbf{C} \in \mathbb{R}^{M_{fo} \times R}$, and segmental $\mathbf{D} \in \mathbb{R}^{N \times R}$ factors.

TD-BASED TENSOR DECOMPOSITION: TD enjoys more flexibility than CPD. Unlike CPD, in which the core-tensor is diagonal, the core-tensor does not need to be diagonal in TD, while the factors are orthogonal matrices. We employ the HOOI algorithm [112] described in Section 3.2.2.2 to solve the TD problem. The tensor $\mathcal{X}$ is decomposed to

$$\mathcal{X} = \mathcal{G} \times_1 \bar{\mathbf{A}} \times_2 \bar{\mathbf{B}} \times_3 \bar{\mathbf{C}} \times_4 \bar{\mathbf{D}}, \tag{7.3}$$

where $\bar{\mathbf{A}} \in \mathbb{R}^{L \times R_1}$, $\bar{\mathbf{B}} \in \mathbb{R}^{F \times R_2}$, $\bar{\mathbf{C}} \in \mathbb{R}^{M_{fo} \times R_3}$, and $\bar{\mathbf{D}} \in \mathbb{R}^{N \times R_4}$, all with orthonormal columns, are respectively temporal, spectral, spatial, and segmental factor matrices. $\mathcal{G} \in \mathbb{R}^{R_1 \times R_2 \times R_3 \times R_4}$ is the core tensor, and its entries illustrate the level of interaction between the factor matrices. Recall that $\times_i$ indicates the $i$-th mode product.

### 7.1.3 *Mapping the TF Features of sEEG to those of iEEG Recordings*

Suppose our training and test datasets consist of $K$ IED and non-IED segments from sEEG recordings, $\mathbf{Y}_c \in \mathbb{R}^{L \times M_{sc}}$, $c = \{1, \cdots, C\}$, where $M_{sc}$ indicates the number of scalp channels. CWT is employed to decompose the segments. The magnitude of wavelet coefficients is obtained as TF features. Then, the TF features of each segment are concatenated into a three-way tensor, $\boldsymbol{\mathcal{Y}}_c \in \mathbb{R}^{L \times F \times M_{sc}}, c \in C$. Finally, $\boldsymbol{\mathcal{Y}}_c$ are projected onto the temporal factors obtained using CPD and TD, respectively referred to as mapping scalp to intracranial recordings via CPD (MStI-CDP) and via TD (MStI-TD). Here, the temporal components can provide the most discriminative features since intracranial and scalp electrodes capture the IED signatures at the same time. Meanwhile, the spatial distributions of intracranial and scalp signals are generally different. Thus, projecting the segments onto the spatial components is meaningless.

#### 7.1.3.1 *MStI-CDP*

The TF features of IED and non-IED segments extracted from sEEG recordings, $\boldsymbol{\mathcal{Y}}_c$, are projected onto the temporal factors achieved by CPD, $\mathbf{A}$, as follows:

$$\boldsymbol{\mathcal{Z}}_c = \boldsymbol{\mathcal{Y}}_c \times_1 \mathbf{A}^\top, \tag{7.4}$$

where $\boldsymbol{\mathcal{Z}}_c \in \mathbb{R}^{R \times F \times M_{sc}}$ for $c = \{1, \cdots, C\}$ is the projected segment of $\{\boldsymbol{\mathcal{Y}}_c, c \in C\}$ and $(.)^\top$ indicates transpose of a matrix. Now, the projected segments, $\boldsymbol{\mathcal{Z}}_c$, are used for classification. Figure 7.1 shows the schematic of the proposed MStI-CPD model.

#### 7.1.3.2 *MStI-TD*

After obtaining the temporal factors using TD, $\bar{\mathbf{A}}$, the TF features of IED and non-IED segments extracted from sEEG recordings, $\boldsymbol{\mathcal{Y}}_c$, are projected onto them as follows:

$$\bar{\boldsymbol{\mathcal{Z}}}_c = \boldsymbol{\mathcal{Y}}_c \times_1 \bar{\mathbf{A}}^\top, \tag{7.5}$$

where $\bar{\boldsymbol{\mathcal{Z}}}_c \in \mathbb{R}^{R_1 \times F \times M_{sc}}$ for $c = \{1, \cdots, C\}$ indicates the projected segment of $\{\boldsymbol{\mathcal{Y}}_c, c \in C\}$. Finaly, the projected segments, $\bar{\boldsymbol{\mathcal{Z}}}_c$, are employed to classify IEDs and non-IEDs. The schematic of the proposed MStI-TD model is presented in Figure 7.1.

## 7.2 EXPERIMENTS

### 7.2.1 *Dataset*

The sEEG signals of 18 epileptic subjects are analysed. The dataset details are given in Section 3.1. Here, a Butterworth highpass filter with a cut frequency of 4 Hz and the

order of 6 has been employed to remove the DC shift, baseline fluctuations, and eye movement artefacts. In addition, a notch filter with a notch frequency of 50 Hz was applied to eliminate the power line interference. The filters are applied to both sEEG and iEEG recordings.

For removing artefacts, a spatial filter is applied to only sEEG recordings. The average amplitude of sEEGs from four electrodes – namely F3, F4, F7, and F8 – is calculated and subtracted from all the scalp signals. These electrodes are placed over the mesial temporal lobes, where the IEDs originate. Empirically, we have found that these are the most effective electrodes in this application for re-referencing.

Here, the signal segment within 160 ms before and 320 ms after the peaks marked as the onset of IED waveforms are selected as IED segments. That is, the length of each IED segment is 480 ms (96 samples). Non-IED segments with the same length of 480 ms are selected without overlap with IED segments. The number of IED and non-IED segments is set to be the same for each subject. Both IED and non-IED segments are linearly detrended to remove the undesired drifts. The total number of IED and non-IED segments is shown in Table 4.1.

### 7.2.2 *Applying the Proposed Tensor-based Methods to Map sEEG to iEEG*

#### 7.2.2.1 *Constructing the training tensor*

As described in Section 7.1.1, CWT is applied to the IED segments of training dataset from iEEG recordings – $\mathbf{X}_n \in \mathbb{R}^{96\times 12}$, $n = \{1, \cdots, N\}$, where 96 and 12 are respectively the number of time samples and FO channels, and $N$ is the number of IED segments in the training dataset – to obtain the TF features. Finally, the TF features of IEDs are concatenated into a four-way tensor, $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{96\times 37\times 12\times N}$, where 37 is the number of wavelet scales which is automatically determined by Matlab wavelet toolbox. Next, either CPD or TD is employed to exploit the factor matrices.

#### 7.2.2.2 *Decomposing the tensor using CPD or TD*

In decomposing a tensor, the number of components should be large enough to avoid missing important information and small enough to avoid undesired information. In CPD, the number of components is selected to be between 3 and 8. Then, the optimized number of components is automatically obtained by a nested cross-validation technique. The tensor $\boldsymbol{\mathcal{X}}$ is decomposed into temporal $\mathbf{A} \in \mathbb{R}^{96\times R}$, spectral $\mathbf{B} \in \mathbb{R}^{37\times R}$, spatial $\mathbf{C} \in \mathbb{R}^{12\times R}$, and segmental $\mathbf{D} \in \mathbb{R}^{N\times R}$ factors.

In TD, the numbers of temporal, spectral, and spatial components are selected to be between 3 and 8, and a nested cross-validation method is used to find the optimum number of components. However, the number of segmental components is selected to be the same as the number of segments $R_4 = N$. The TD algorithm is employed to

decompose the tensor $\mathcal{X}$ into the core tensor $\mathcal{G} \in \mathbb{R}^{R_1 \times R_2 \times R_3 \times N}$ and the factor matrices, i.e. temporal $\bar{\mathbf{A}} \in \mathbb{R}^{96 \times R_1}$, spectral $\bar{\mathbf{B}} \in \mathbb{R}^{37 \times R_2}$, spatial $\bar{\mathbf{C}} \in \mathbb{R}^{12 \times R_3}$, and segmental $\bar{\mathbf{D}} \in \mathbb{R}^{N \times N}$ factor matrices.

#### 7.2.2.3 *Mapping sEEG to iEEG through MStI-CPD and MStI-TD*

For mapping the sEEG to iEEG, the TF features of all IED and non-IED segments from both the training and test datasets of sEEG recordings – $\mathbf{Y}_c \in \mathbb{R}^{96 \times 18}, c = \{1, \cdots, C\}$ – are calculated. CWT with complex Morlet wavelet is applied for extracting the TF features. Then, the TF features of each segment are concatenated into a three-way tensor, $\mathcal{Y}_c \in \mathbb{R}^{96 \times 37 \times 18}, c \in C$.

APPLYING MSTI-CPD: In MStI-CPD, $\mathcal{Y}_c$ is projected onto the temporal factors obtained using CPD, $\mathbf{A}$, through equation (7.4), which gives $\mathcal{Z}_c \in \mathbb{R}^{R \times 37 \times 18}$ for $c = \{1, \cdots, C\}$. Finally, $\mathcal{Z}_c$ is vectorized and used for classification.

APPLYING MSTI-TD: In MStI-TD, $\mathcal{Y}_c$ is projected onto the temporal factors obtained using TD, $\bar{\mathbf{A}}$, via (7.5), resulting in $\bar{\mathcal{Z}}_c \in \mathbb{R}^{R_1 \times 37 \times 18}$ for $c = \{1, \cdots, C\}$. Finally, we vectorize $\bar{\mathcal{Z}}_c$ and use it for classification.

### 7.2.3 *Feature Selection*

The number of extracted features depends on the number of temporal factors, selected to be between 3 and 8 by nested cross validation. Thus, the minimum number of features is $3 \times 37 \times 18 = 1998$, and their maximum $8 \times 37 \times 18 = 5328$. However, the number of features is relatively high for applying directly to a classifier. Therefore, we employ a feature selection method to find the most discriminative features.

Minimum redundancy — maximum relevance (MRMR) approach is used as a feature selection method [137]. In MRMR, the mutual information is calculated to measure the level of similarity between features. Those features that are mutually maximally dissimilar (their mutual Euclidean distances are maximized, or their pairwise correlations are minimized) are selected as discriminative features.

### 7.2.4 *Classification and Cross-validation*

DTE (with bagging technique) and KNN are employed as classifiers. ACC, SEN, SPEC, and F1-S are obtained as evaluation criteria. Two approaches of within- and between-subject classifications are used for evaluation. In the within-subject approach, the IEDs and non-IEDs of each subject are classified separately, meaning that the training and test data come from the same subject. In this approach, k-fold ($k = 5$) cross-validation

is employed for classification, that is, four folds are used for training and the rest for testing. On the other hand, in the between-subject approach, IEDs and non-IEDs of all the subjects are classified together. Leave-one-subject-out cross-validation is used in this approach. That is, the IED and non-IED segments of a subject are used as the test data, and the other subjects' data are used for training the model. This process is repeated for all subjects one-by-one. Note that the data of those subjects who provide high performance in the within-subject classification approach are included in the training dataset. For finding the hyperparameters, including the optimized number of neighbors in KNN and the number of components in tensor decomposition techniques, nested cross-validation is used. In the nested cross-validation, a subset of training data is employed for training the model, and the rest is used for validation to optimize the hyperparameters.

### 7.2.5 *Competing Approaches*

Our proposed method is compared with the following four methods.

- **SCA:** SCA is proposed in Chapter 3. In SCA, all IEDs from sEEG are concatenated into a three-way tensor. Temporal, spatial, and segmental factors are obtained by employing CPD. Finally, all IEDs and non-IEDs are projected onto the spatial components. In this method, only sEEG recordings are used in both training and test datasets.
- **TF:** In the TF-based method [8], the TF features are extracted using the spectrogram method and used as the classification features for IED detection. In this work, an individual classifier is trained for each training subject. All classifiers of training subjects are ensemble to detect the IEDs of the test subject. Note that only sEEG is used for training the classifiers. The logistic regression classifier is employed as classification.
- **LSR:** In [133], an LSR model is employed to model the iEEG from sEEG recordings. A stepwise discriminant analysis classifier is employed for classification.
- **AAE & ASAE:** In [9], the authors map the sEEG to iEEG by developing an asymmetric AE (AAE). They call it asymmetric since the number of inputs and outputs of AE are not the same. The output of AAE is referred to as pseudo-iEEG. The pseudo-iEEG is again mapped to the real iEEG by feeding it to a symmetric AE. The overall method is called asymmetric-symmetric AE (ASAE). A CNN was employed for feature exploitation and classification. These methods are referred to respectively as AAE and ASAE.

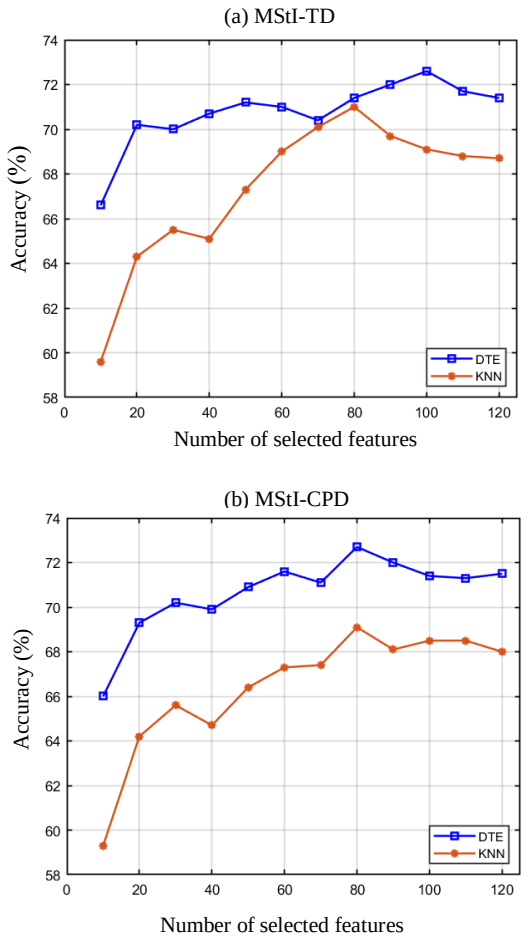



Figure 7.2: The accuracy of between-subject IED and non-IED classification versus the number of selected features based on MRMR feature selection method for combined scalp-visible and scalp-invisible spikes. Scalp-invisible IEDs comprise more than 90% of the total spikes.

## 7.3 EXPERIMENTAL RESULTS

An effective subset of features is selected based on the MRMR feature selection method. Figure 7.2 shows the performances of classifiers in the between-subject approach versus the number of selected features. The DTE classifier has the best accuracy respectively through the first 100 and 80 features using MStI-TD and MStI-CPD methods. On the other hand, the KNN classifier achieves the best accuracy through the first 80 features in both MStI-TD and MStI-CPD methods.

Table 7.1 shows the obtained results from MStI-TD and MStI-CPD using the DTE classifier. The performance of each subject, the mean of performance across subjects, and the standard error (SE) of the performance are illustrated. In the between-subject classification approach, the obtained accuracy values vary among subjects from 47.3% to 92.1% in both MStI-TD and MStI-CPD. However, the average accuracy values across subjects achieved using MStI-TD and MStI-CPD are respectively 72.6% and 72.7% with

the SE of approximately 2.7% and 3.1%. This is promising as the accuracy value includes those of scalp-invisible spikes – which include over 80% of the total spikes – too. The obtained results from the within-subject classification approach are shown in Table 7.1 in parentheses. MStI-TD provides accuracy values of higher than 80% for most subjects. On average, it achieves 84.2% accuracy, 80.5% sensitivity, 87.9% specificity, and 0.82 F1-score. The performance of MStI-CPD is comparable with that of MStI-TD, and it classifies IED and non-IEDs with 81.3% accuracy and 75.5% sensitivity.

Overall, the performance of the KNN classifier shown in Table 7.2 is relatively lower than that of the DTE classifier. It provides an accuracy of around 70% in the between-subject classification approach and around 80% in the within-subject classification approach (shown in parentheses) using both MStI-TD and MStI-CPD methods. However, the KNN classifier detects the IEDs with 62.4% and 62.0% sensitivity values using respectively MStI-TD and MStI-CPD in the between-subject approach. It also achieves the sensitivity value of 73.1% using both MStI-TD and MStI-CPD in the within-subject approach.

In subject 1, none of the IEDs (SEN) is detected, and both classifiers are biased to the non-IED class in both mapping models. In other words, most EEG segments are recognized as non-IED segments. In subjects 3, 4, and 6, the KNN classifier is also biased toward the non-IED class. The sensitivity value is less than 20% for these subjects, while specificity is 100%. The DTE classifier also does not achieve high sensitivity values for these subjects. It can be concluded that these subjects may generate IEDs with different morphologies compared to other subjects. In the projection procedure, the segments are projected onto the temporal factors. When the IED morphologies of a test subject are different from those of the training subjects, the projection cannot give discriminative features, and thus the classifier fails to detect the IEDs and non-IEDs.

The number of segments does not affect the results. In the within-subject approach, high performance has been achieved for subjects with less than 20 IED segments and those with more than 200 IED segments. Conversely, the performance differs for different subjects in the between-subject approach, where the number of segments in the training dataset is not significantly different due to using leave-one-subject-out cross-validation. Therefore, it can be concluded that the number of segments does not affect the model performance.

The proposed mapping models have been compared with SCA, TF, LSR, AAE, and ASAE. The obtained results are illustrated in Table 7.3. The results of our proposed methods are reported based on the DTE classifier. The results of AAE, ASAE, and LSR are based on the results obtained in [9]. The performance of TF is based on [8]. The results of SCA are based on our previous work reported in Chapter 5. TF, LSR, AAE, and ASAE studies reported the sensitivity, specificity, and F1-S only in the between-subject classification approach. All these studies [8, 9, 138] use the same dataset we use to evaluate our methods. Our proposed mapping models outperform the others.

Table 7.1: The performance of our proposed MStI-TD and MStI-CPD models obtained using the DTE classifier. SE shows the standard error of the performance. The numbers in parentheses are the within-subject classification performance. ACC, SEN, and SPC are presented in percent (%).

| Subject | MStI-TD | | | | MStI-CPD | | | |
|---|---|---|---|---|---|---|---|---|
| | ACC | SEN | SPC | F1-S | ACC | SEN | SPC | F1-S |
| 1 | **47.3** (80.0) | **0.0** (93.3) | 94.7 (66.6) | 0.0 (0.83) | **47.3** (80.0) | **0.0** (86.6) | 94.7 (73.3) | 0.0 (0.78) |
| 2 | 65.6 (90.0) | 97.7 (85.7) | 33.6 (94.2) | 0.74 (0.89) | 60.9 (89.0) | 98.4 (85.0) | 23.2 (93.1) | 0.71 (0.88) |
| 3 | 63.9 (80.6) | 34.4 (78.6) | 93.4 (82.6) | 0.48 (0.80) | 61.6 (75.0) | 27.1 (72.0) | 96.0 (78.0) | 0.41 (0.74) |
| 4 | 64.8 (83.0) | 29.6 (80.0) | 100 (86.0) | 0.45 (0.76) | 59.2 (85.0) | 20.3 (78.0) | 98.1 (92.0) | 0.33 (0.81) |
| 5 | 90.5 (90.6) | 86.1 (82.6) | 94.9 (98.6) | 0.90 (0.89) | 90.5 (88.6) | 82.3 (78.6) | 98.7 (98.6) | 0.89 (0.87) |
| 6 | 73.0 (82.6) | 48.7 (78.4) | 97.2 (86.88) | 0.64 (0.82) | 71.4 (78.3) | 45.3 (70.0) | 97.5 (86.5) | 0.61 (0.76) |
| 7 | 62.8 (77.2) | 31.2 (73.6) | 94.4 (80.8) | 0.45 (0.74) | 64.8 (77.2) | 32.8 (72.0) | 96.8 (82.4) | 0.48 (0.73) |
| 8 | 59.9 (79.3) | 95.3 (74.2) | 23.9 (84.3) | 0.70 (0.78) | 59.4 (72.9) | 94.5 (69.8) | 24.3 (76.0) | 0.70 (0.72) |
| 9 | **92.1** (93.3) | 89.5 (86.6) | 94.7 (100) | 0.91 (0.92) | **92.1** (86.6) | 89.4 (73.3) | 94.7 (100) | 0.92 (0.84) |
| 10 | 59.3 (79.5) | 36.6 (77.1) | 82.0 (82.0) | 0.47 (0.79) | 59.3 (71.9) | 34.1 (69.7) | 84.7 (74.2) | 0.45 (0.71) |
| 11 | 76.4 (77.0) | 77.4 (69.5) | 75.4 (84.3) | 0.76 (0.75) | 74.7 (75.2) | 80.0 (72.2) | 69.3 (78.2) | 0.76 (0.75) |
| 12 | 75.8 (83.5) | 86.1 (74.7) | 65.7 (92.3) | 0.78 (0.82) | 75.8 (84.1) | 86.0 (72.9) | 65.7 (95.3) | 0.78 (0.82) |
| 13 | 80.7 (90.0) | 61.5 (90.0) | 100 (90.0) | 0.76 (0.89) | 80.7 (75.0) | 61.5 (70.0) | 100 (80.0) | 0.76 (0.65) |
| 14 | 85.0 (80.0) | 90.0 (76.0) | 80.0 (90.0) | 0.85 (0.76) | 90.0 (85.0) | 90.0 (80.0) | 90.0 (90.0) | 0.90 (0.82) |
| 15 | 71.2 (84.8) | 92.5 (81.4) | 50.0 (88.1) | 0.76 (0.84) | 72.4 (81.9) | 91.3 (73.3) | 93.5 (90.4) | 0.76 (0.80) |
| 16 | 81.8 (85.0) | 72.7 (80.0) | 90.9 (90.0) | 0.80 (0.76) | 90.9 (85.0) | 90.9 (70.0) | 90.9 (100) | 0.91 (0.73) |
| 17 | 75.8 (87.6) | 73.0 (85.8) | 78.6 (89.4) | 0.75 (0.87) | 75.8 (82.3) | 73.0 (80.0) | 78.6 (84.7) | 0.75 (0.82) |
| 18 | 81.6 (90.9) | 92.3 (86.6) | 71.0 (95.1) | 0.83 (0.90) | 82.0 (89.4) | 92.3 (85.4) | 71.6 (93.3) | 0.83 (0.89) |
| Mean | **72.6** **(84.2)** | 66.4 (80.5) | 78.9 (87.9) | 0.67 (0.82) | **72.7** **(81.3)** | 66.1 (75.5) | 79.3 (87.0) | 0.66 (0.78) |
| SE | ±2.7 (±1.2) | ±6.8 (±1.5) | ±5.3 (±1.8) | ±.05 (±.01) | ±3.1 (±1.3) | ±7.3 (±1.3) | ±5.7 (±2.1) | ±.05 (±.01) |

Table 7.2: The performance of our proposed MStI-TD and MStI-CPD models obtained using the KNN classifier. SE shows the standard error of the performance. The numbers in parentheses are the within-subject classification performance. ACC, SEN, and SPC are presented in percent (%).

| Subject | MStI-TD | | | | MStI-CPD | | | |
|---|---|---|---|---|---|---|---|---|
| | ACC | SEN | SPC | F1-S | ACC | SEN | SPC | F1-S |
| 1 | **50.0** (80.0) | **0.0** (80.0) | 100 (80.0) | 0.0 (0.78) | **50.0** (90.0) | **0.0** (86.6) | 100 (93.3) | 0.0 (0.87) |
| 2 | 67.0 (87.7) | 96.5 (77.3) | 37.4 (98.0) | 0.74 (0.86) | 62.6 (86.0) | 99.2 (76.1) | 25.9 (95.7) | 0.72 (0.84) |
| 3 | **33.0** (78.3) | **6.0** (76.0) | 100 (80.0) | 0.11 (0.78) | **51.3** (70.0) | **2.6** (68.0) | 100 (72.0) | 0.05 (0.69) |
| 4 | **55.5** (79.0) | **11.1** (70.0) | 100 (88.0) | 0.20 (0.70) | **55.1** (81.0) | **10.2** (74.0) | 100 (88.0) | 0.16 (0.75) |
| 5 | 91.7 (88.0) | 86.1 (76.0) | 97.4 (100) | 0.91 (0.86) | 84.1 (86.6) | 84.8 (73.3) | 83.5 (100) | 0.84 (0.84) |
| 6 | **59.2** (80.3) | **18.5** (72.8) | 100 (87.8) | 0.31 (0.86) | **56.5** (73.9) | **13.2** (62.5) | 99.7 (85.3) | 0.23 (0.70) |
| 7 | 63.2 (73.2) | 28.0 (59.2) | 98.4 (87.2) | 0.43 (0.59) | 60.0 (72.8) | 20.1 (62.4) | 99.5 (83.2) | 0.32 (0.64) |
| 8 | 63.4 (72.9) | 87.0 (62.2) | 39.8 (83.6) | 0.70 (0.69) | 60.5 (69.8) | 89.1 (61.8) | 31.9 (77.8) | 0.69 (0.67) |
| 9 | **94.7** (96.6) | 94.7 (93.3) | 94.7 (100) | 0.94 (0.96) | 86.8 (83.3) | 84.2 (66.6) | 89.4 (100) | 0.86 (0.78) |
| 10 | 58.5 (76.7) | 31.8 (74.8) | 85.2 (78.7) | 0.43 (0.76) | 60.7 (65.1) | 43.1 (62.9) | 78.4 (67.4) | 0.52 (0.64) |
| 11 | 72.7 (73.9) | 75.4 (66.6) | 69.9 (81.1) | 0.73 (0.72) | 74.1 (73.0) | 80.3 (64.9) | 67.9 (81.1) | 0.52 (0.64) |
| 12 | 73.0 (80.8) | 88.9 (62.3) | 57.0 (99.4) | 0.77 (0.76) | 68.9 (80.6) | 88.3 (61.7) | 49.4 (99.4) | 0.74 (0.76) |
| 13 | 33.1 (80.0) | 46.1 (70.0) | 100 (90.0) | 0.63 (0.69) | 73.1 (80.0) | 46.1 (70.0) | 100 (90.0) | 0.63 (0.69) |
| 14 | 85.0 (80.0) | 90.0 (70.0) | 80.0 (90.0) | 0.85 (0.76) | 80.0 (85.0) | 80.0 (80.0) | 80.0 (90.0) | 0.80 (0.82) |
| 15 | 67.3 (81.6) | 95.3 (71.6) | 39.3 (91.6) | 0.75 (0.79) | 61.1 (79.1) | 94.2 (69.8) | 28.0 (88.4) | 0.71 (0.77) |
| 16 | 86.3 (75.0) | 90.0 (70.0) | 81.8 (80.0) | 0.87 (0.66) | **90.9** (90.0) | 100 (90.0) | 81.8 (90.0) | 0.91 (0.89) |
| 17 | 81.4 (87.6) | 86.5 (83.5) | 76.4 (91.7) | 0.82 (0.87) | 82.5 (82.9) | 88.7 (77.0) | 76.4 (88.2) | 0.83 (0.82) |
| 18 | 83.1 (90.0) | 91.1 (80.6) | 75.1 (99.4) | 0.84 (0.88) | 85.5 (87.5) | 91.7 (75.7) | 79.3 (99.4) | 0.86 (0.85) |
| Mean | **71.0** **(81.2)** | 62.4 (73.1) | 79.6 (89.3) | 0.61 (0.77) | 69.1 (79.8) | 62.0 (71.3) | 76.1 (88.3) | 0.59 (0.76) |
| SE | ±3.2 (±1.5) | ±8.5 (±1.9) | ±5.3 (±1.8) | ±.07 (±.02) | ±3.1 (±1.7) | ±8.7 (±2) | ±6.0 (±2.2) | ±.07 (±.02) |

Table 7.3: The performance of our proposed MStI-TD and MStI-CPD models (using the DTE classifier) and the competing methods with results averaged over all subjects. The numbers in parentheses are the within-subject average of IED and non-IED classification rate. ACC, SEN, and SPC are presented in percent (%).

| Models | ACC | SEN | SPC | F1-S |
|---|---|---|---|---|
| SCA | 62 (78) | 43 (70) | **80** (87) | 0.47 (0.76) |
| TF | 65 (67) | 65 | 64 | 0.63 |
| LSR | 62 (65) | 61 | 59 | 59 |
| AAE | 66 (68) | 66 | 65 | 0.65 |
| ASAE | 68 (73) | **67** | 68 | **0.68** |
| MStI-TD | **72 (84)** | **66 (80)** | 78(87) | 0.67 **(0.82)** |
| MStI-CPD | **72 (81)** | **66 (75)** | 79(87) | 0.66 **(0.78)** |

Among the compared methods, ASAE achieves the best accuracy of 68%, which is 4% less than that of our proposed MStI-TD and MStI-CPD models in the between-subject approach. AAE, ASAE, and LSR map the sEEG to iEEG. However, LSR achieved 62% accuracy, which is 3% less than the accuracy of TF, while the TF model is based on only sEEG recordings. SCA is biased to the non-IED class and provides a low sensitivity value of 43% in the between-subject classification approach.

Though our proposed method is biased toward the non-IED class in very few subjects, there is a good balance between sensitivity and specificity. We project the IED and non-IED segments onto the temporal factors. Therefore, the bias occurs when the IED morphology of the test subject differs from those of the training subjects, which is rare. On the other hand, in SCA, the IED and non-IED segments are projected onto the spatial components. Because of spatial projection, SCA depends on the IED source location. Therefore, when the location of IED sources for the test subject differs from that of the training subjects, the system tends to be biased toward the non-IED class. As it can be seen from Table 7.3, SCA performs significantly better in the within-subject approach (the accuracy of 78% and sensitivity of 70%) as compared to the between-subject approach (the accuracy and sensitivity values of 62% and 43%, respectively). This difference between the performances of within- and between-subject approaches indicates that SCA is sensitive to the epileptiform source location and is not necessarily suitable for being employed in a between-subject classification approach.

## 7.4 DISCUSSION

IED identification can establish a guideline for pre-ictal state monitoring, seizure prediction, treatment, and surgical planning. However, the sensitivity of scalp recordings is low in capturing epileptiform discharges, and around 30% to 40% of patients con-

sidered for epilepsy surgery require intracranial recordings [5]. On the other hand, intracranial recordings are invasive and have side effects. Therefore, ameliorating the sensitivity of sEEG to detect IEDs motivates developing a new direction for more concise seizure prediction.

In some studies, the IEDs are detected with high sensitivity from sEEG signals [10, 85]. Thanh *et al.* proposed a tensor-based method to detect epileptic and non-epileptic spikes [10]. Their model achieved a sensitivity value of 83%. In [85], the authors developed a method based on deep neural networks. The model detects the epileptiform discharges with 81% sensitivity, while it suffers from low specificity of 46%. However, in both studies, scalp recordings are utilized for scoring IEDs. The main disadvantage of studies in which sEEG signals are used as ground truth for labeling IEDs is that they miss the scalp-invisible IEDs or spikes, which often contribute to more than 80% of the IEDs by default.

Among the compared methods, TF and SCA detect IEDs from the concurrent sEEG recordings. In these methods, the iEEG recordings are used as ground truth for labeling both scalp-visible and scalp-invisible IEDs, while only sEEG recordings are employed to detect IEDs. There is an appropriate balance between the sensitivity and specificity values in TF, although they are not high enough. On the other hand, SCA suffers from low sensitivity, though it achieves high specificity. SCA is based on spatial components. Different subjects may have different epileptiform source locations. When the data from different subjects are combined, the spatial factors are corrupted, and consequently, the model performance deteriorates. Therefore, SCA fails to detect the IEDs in the between-subject classification approach. However, it is a powerful method for the within-subject classification approach and enables to classify the IEDs and non-IEDs with high accuracy.

LSR, AAE, and ASAE map the sEEG to iEEG recordings. The performance of LSR is worse than all other methods. ASAE achieves the highest accuracy and sensitivity values among the competing methods in the between-subject approach, 68% and 67% respectively. Meanwhile, it provides a higher accuracy of 73% in the within-subject approach. However, when the training and test datasets come from different subjects (between-subject classification) and have different distributions, AAE and ASAE cannot be successful since the neural networks are highly sensitive to the data distribution.

One of the advantages of our proposed tensor-based method is that it enjoys flexibility in terms of TF feature extraction and tensor decomposition structure. Here, the magnitude of continuous wavelet coefficients is extracted as TF features. Other TF methods, such as discrete wavelet transform, spectrogram, or synchrosqueezed transform highly used in biomedical signal processing [8, 139, 140], can be employed to extract the TF features. In addition, CPD and TD are applied here for decomposing the tensor into its factor matrices. Other algorithms, for example, nonnegative- or sparse-based tensor decomposition methods, can be employed to decompose the tensor into the factor ma-

trices. Here, the performances of both mapping methods are comparable. Therefore, it cannot be concluded which one is superior to another.

The main limitation of our proposed method is that it depends on the morphologies and shapes of IEDs. In our mapping model, the projection is performed onto the temporal factors that can vary if the IED morphologies significantly differ. Therefore, the model's performance may be adversely affected for a new subject with an IED of very different morphology and shape.

Intracranial recordings are necessary for training the mapping-based methods. This can be a limitation for such methods. However, these methods can detect IEDs from new subjects without any need for intracranial signals. Furthermore, we need to have some recorded iEEG signals from the training subjects to score the scalp-invisible IEDs.

## 7.5 CONCLUSION

This chapter proposes a method based on TF features and tensor decomposition to design a mapping algorithm for projecting sEEG to iEEG. At first, the TF features of intracranial IEDs from the training set are extracted and concatenated into a four-way tensor with time, frequency, channel, and IED segment slabs. Then, two tensor factorisation algorithms, CPD and TD, are employed to decompose the tensor into factor matrices. In the mapping procedure, the TF features of scalp IEDs and non-IEDs from both training and test sets are computed and projected onto the temporal factors obtained from CPD and TD, respectively called MStI-CPD and MStI-TD. The results of the other four methods, namely SCA [138], TF [8], LSR [133], and AAE as well as ASAE [9], are reported here for comparison. The methods are validated in within- and between-subject classification approaches. MStI-CPD and MStI-TD outperform the competing methods in both approaches and achieve respectively the accuracy values of around 84.2% and 72.6% in within- and between-subject classifications. Meanwhile, among the competing methods, ASAE obtains the highest accuracy value of 68% in the between-subject classification approach. SCA and LSR provide the worse accuracy value of 62%. These findings show that when the training and test data come from different subjects, ASAE and SCA cannot perform favorably because of being sensitive respectively to the data distribution and epileptiform source location. However, the proposed method does not suffer from these limitations. Firstly, unlike AAE and ASAE, tensor factorisation is not highly sensitive to data distribution. Secondly, unlike SCA, in which the data is projected onto the spatial factors, in the proposed MStI-CPD and MStI-TD methods, the data is projected onto the temporal factors, which is independent from the IED source location.

## 7.6 AKNOWLEDGE

Electronic version of an article published as "Higher-order tensor decomposition based scalp-to-intracranial EEG projection for detection of interictal epileptiform discharges," Journal of Neural Engineering, 18(6), p.066039., doi: 10.1088/1741-2552/ac3cc4.

# 8

# EEG-TO-EEG: SCALP-TO-INTRACRANIAL EEG TRANSLATION USING GENERATIVE ADVERSARIAL NETWORKS

In the previous chapter, we have proposed two methods based on tensor factorisation to map the TF features of sEEG to those of iEEG to improve an IED detection system performance. In this chapter, we develop and examine two novel mapping methods based on GANs to map raw sEEG to iEEG.

Mapping sEEG to iEEG using deep neural networks can be considered as generating iEEG by feeding sEEG to a network. GANs [141] and variational AEs (VAEs) [142] are powerful tools for generating data (such as image, text, and speech). Generally, a GAN consists of a generator and a discriminator network. The generator network generates the data, and the discriminator network discriminates the real and generated data. GANs map either a noise vector [141, 143] or an observed data sample [144–148] to a data space. When the generator generates a data space from the observed data, it is called the conditioning generator. Conditions may also be applied to the discriminator. In a conditioning discriminator, the generator input data is concatenated with either the real or generated data and fed to the discriminator. Generally, if a condition (or conditions) is applied to a generator, a discriminator, or both generator and discriminator networks of GAN, it is called conditional GAN (cGAN). cGANs have been employed in different applications, such as text-to-image [144, 149], speech-to-image [145], speech-to-speech [146], and particularly image-to-image [147, 148, 150, 151]. Here, our main contribution is to design two types of cGANs for EEG-to-EEG mapping. We feed sEEG to the generator to generate an estimation of iEEG, called scalp-to-intracranial EEG translation. In this study, we propose two EEG-to-EEG mapping techniques. The first one is a cGAN based on U network (U-net) [152], called Unet-cGAN. In the second method, we propose a novel cGAN technique based on VAE, called VAE-cGAN. In other words, in this method, VAE is incorporated in the generator of cGAN.

The rest of the chapter is structured as follows. First, an introduction to deep learning methods, including CNN, LSTM, VAE, and GAN, is presented. Then, the proposed Unet-cGAN and VAE-cGAN for EEG-to-EEG translation are explained. Next, the experimental setup is provided. Afterward, the results are presented. Finally, the last section concludes the chapter.

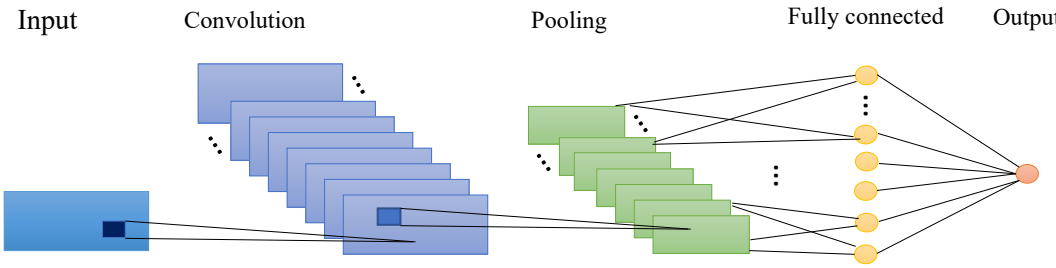



Figure 8.1: Schematic diagram of a basic convolutional neural network architecture.

## 8.1 DEEP LEARNING TECHNIQUES

Deep neural networks (DNN) have been provided in different structures for different applications. The most popular architectures are CNN, LSTM, GAN, and VAE.

### 8.1.1 *CNN*

CNNs mainly comprise three types of layers: convolutional, Pooling, and fully connected (FC). A CNN architecture is formed by stacking these layers. A simplified CNN architecture is shown in Figure 8.1. The convolution layer applies a convolution to the previous layer data and forwards the result to the next layer. A non-linear activation function follows it. The pooling layer gradually reduces the dimensionality of the representation to decrease the computational complexity. Finally, a FC layer performs like standard neural networks.

CONVOLUTIONAL LAYER: A convolution is performed by sliding a filter (or mask) over the input, multiplying it by the data, and aggregating the result. The layers' parameters focus on the use of learnable filters. In a convolution layer, the data is convolved with a sliding mask (a.k.a. filter) which can be two dimensions for images or one dimension for signals. After each convolution, the results are accumulated using the following pooling layer.

POOLING LAYER: A pooling layer is usually applied after a convolutional layer. It downsamples the data along the spatial (for images) and temporal (for signals) dimensions of the given input. Consequently, it reduces the number of parameters and the computational complexity of the model. The pooling layer operates over each activation map in the input. It scales its dimensionality using either a "Max" or an "Average" function—the Max and Average pooling functions return respectively the maximum and the average values of each filter mask.

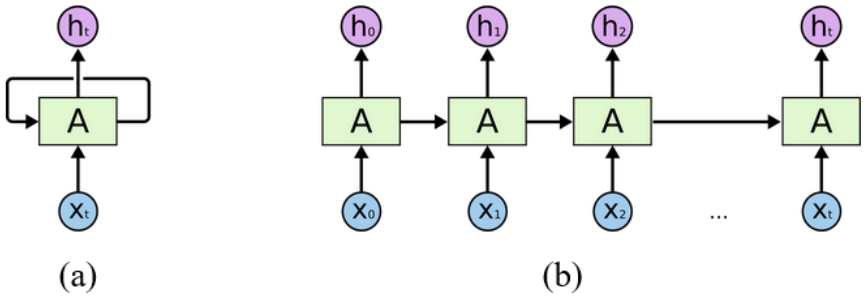



Figure 8.2: An RNN (a) and its unfolded representation.

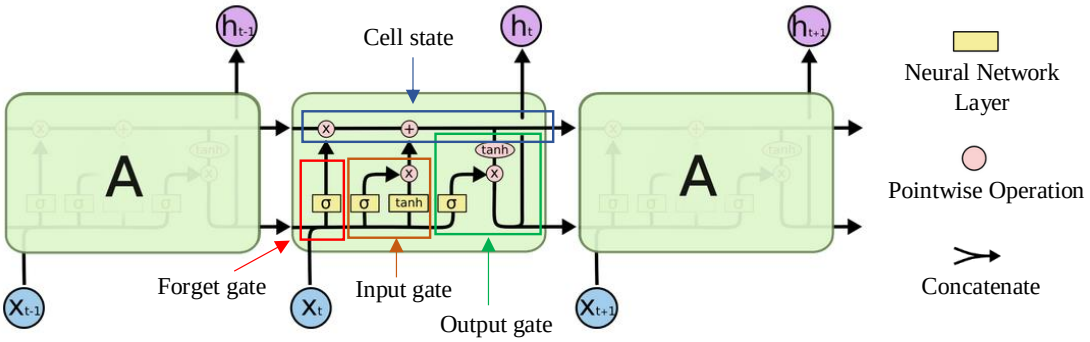



Figure 8.3: A long short-term memory network architecture.

FC: In an FC network, each neuron in the current layer is connected to all the neurons of the next layer.

### 8.1.2 *LSTM*

LSTM is a modified model of RNN. In an RNN based model, shown in Figure 8.2, the output from previous step is fed as input to the current step. In traditional neural networks, all the inputs and outputs are independent. However, in data prediction or in places (such as speech and video) where the data history is helpful in data classification, RNNs become advantageous.

However, standard RNNs fail in learning to connect the information when the gap length increases. LSTM proposed by Hochreiter and Schmidhuber [153] can tackle the problem of long-term dependencies of RNN in which the RNN cannot predict the word stored in the long-term memory but can give more accurate predictions from the recent information. LSTMs are explicitly designed to avoid the long-term dependency problem. Remembering information for long periods is practically their default behavior.

Standard RNN and LSTM can be presented as a chain of repeating neural network modules. The standard RNN has a single layer performing hyperbolic tangent (Tanh)

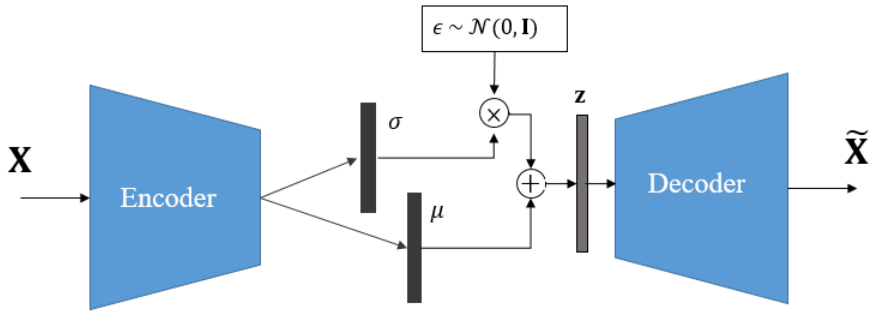



Figure 8.4: The variational autoencoder architecture.

activation. Meanwhile, LSTM contains four neural networks and different memory blocks called cells, shown in Figure 8.3.

The key factor of LSTMs is the cell state, the horizontal line running through the top of the diagram. The LSTM has the ability to remove or add information to the cell state, carefully regulated by structures called gates. As a result, the cells retain the information, and the gates carry out the memory manipulations. There are three gates, namely, Forget gate, Input gate, and Output gate, to protect and control the cell state.

FORGET GATE: The Forget gate decides which information should be removed from the cell state. It looks at $x_t$ (input at the particular time) and $h_{t-1}$ (previous cell output), and outputs a number between 0 and 1 for each number in the cell state $c_{t-1}$. If the output is 0 for a particular cell state, the piece of information is forgotten, and for output 1, the information is retained for future use.

INPUT GATE: The Input gate decides which values are to be updated. First, the information is regulated using the sigmoid function similar to the Forget gate using inputs $x_t$ and $h_{t-1}$. Next, a layer with the Tanh function creates a vector of new candidate values, giving an output from -1 to +1. At last, the vector and regulated values are multiplied to create an update to the state.

OUTPUT GATE: The Output gate filters the cell state and presents an output. First, a vector is generated by applying the Tanh function to the cell to push its values between -1 and 1. Then, the information is regulated using the sigmoid function. At last, the vector and regulated values are multiplied to be sent as an output and input to the next cell.

### 8.1.3 *VAE*

A VAE is made up of two networks, namely, encoder and decoder. The encoder network $\mathcal{E}$ encodes a data sample $\mathbf{X}$ into a latent representation $\mathbf{z}$ based on the distribution, and

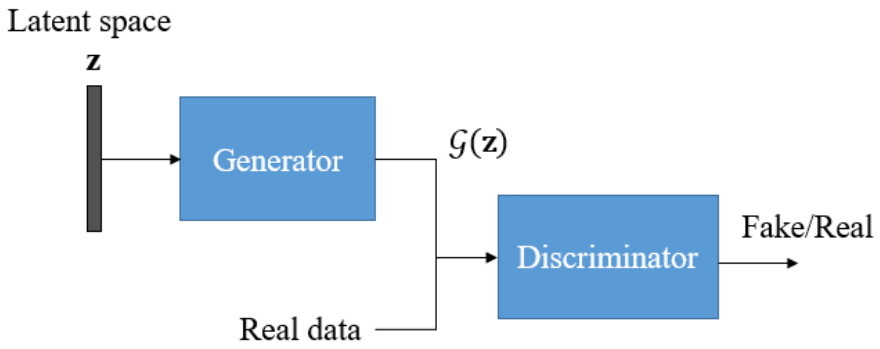



Figure 8.5: The generative adversarial network architecture.

the decoder network decodes the latent representation to reconstruct the data (called the reconstruction or generated data $\widetilde{\mathbf{X}}$) with minimum error. An architecture of VAE is presented in Figure 8.4

A VAE imposes a prior over the latent space $p(\mathbf{z})$ to ensure that it follows a Gaussian distribution, $\boldsymbol{\epsilon} \sim \mathcal{N}(0, \mathbf{I})$. In other words, rather than mapping the data sample $\mathbf{X}$ directly to $\mathbf{z}$, the encoder network $\mathcal{E}$ maps $\mathbf{X}$ into two different vectors that are the mean $\boldsymbol{\mu}$ and the standard deviation (STD) $\sigma$ of multivariate Gaussian distribution. Then, the Gaussian sample $\boldsymbol{\epsilon}$ is shifted and scaled by the predicted $\boldsymbol{\mu}$ and $\sigma$ as follows

$$\mathbf{z} = \boldsymbol{\mu} + \boldsymbol{\epsilon} \odot \sigma, \tag{8.1}$$

where $\boldsymbol{\epsilon} \sim \mathcal{N}(0, \mathbf{I})$ is an auxiliary noise variable, and $\odot$ indicates the element-wise product.

To enforce the encoder to map the data sample $\mathbf{X}$ into a Gaussian distribution, the Kullback-Leibler (KL) divergence $D_{KL}$ is calculated as follows:

$$\mathcal{L}_{D_{KL}} = D_{KL}(\mathcal{E}(\mathbf{X}) \| p(\mathbf{z})). \tag{8.2}$$

The reparameterization trick in [142] is employed for estimating the objective loss (8.2) as follows:

$$\mathcal{L}_{D_{KL}} \simeq -\frac{1}{2} \sum_{z=1}^{Z} (1 + log(\sigma_z^2) - \mu_z^2 - \sigma_z^2), \tag{8.3}$$

where $Z$ is the dimension of latent space.

### 8.1.4 *GAN*

A GAN consists of a generator $\mathcal{G}$ and a discriminator $\mathcal{D}$, shown in Figure 8.5. The generator $\mathcal{G}$ maps the latent space $\mathbf{z}$ with a prior probability distribution $p(\mathbf{z})$ to a data

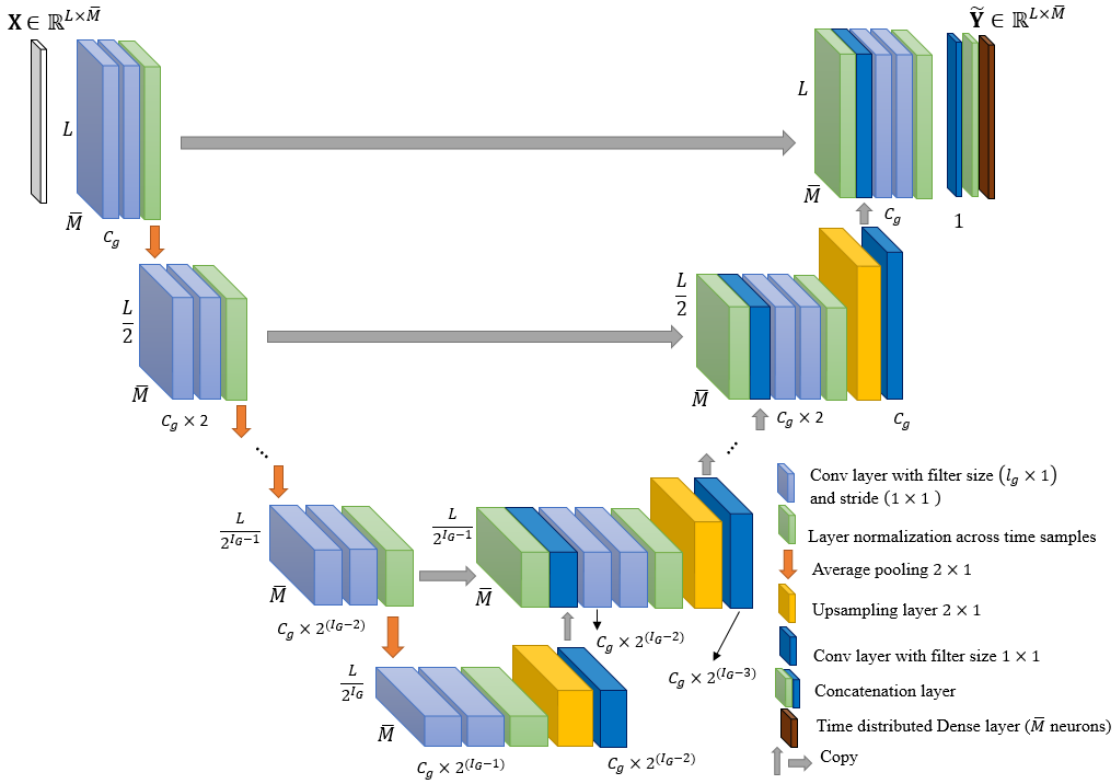



Figure 8.6: The generator network architecture of Unet-cGAN.

space to generate a data sample. The discriminator $\mathcal{D}$ takes either real or generated data samples as input and predicts a binary class of real or fake (generated). An adversarial loss is employed to train the generator and discriminator. The original GAN study [141] introduced the binary cross entropy in a min-max game approach according to the following loss function:

$$\min_{\mathcal{G}} \max_{\mathcal{D}} \mathcal{L}_{GAN}(\mathcal{G}, \mathcal{D}) = \mathbb{E}_{\mathbf{Y}}[log(\mathcal{D}(\mathbf{Y}))] + \mathbb{E}_{\mathbf{z}}[log(1 - \mathcal{D}(\mathcal{G}(\mathbf{z})))], \tag{8.4}$$

where $\mathcal{G}$ minimizes the objective loss function against an adversarial $\mathcal{D}$ maximizing it, $\mathbf{Y}$ is the real data sample, $\mathcal{G}(\mathbf{z})$ is the generated sample, and $\mathbf{z}$ is the latent space (random noise).

In cGANs, both the generator and discriminator can be conditioned. In conditioning the generator, instead of latent space, an observed data sample is fed to the generator to generate a data space and, consequently, the adversarial loss function is defined as follows:

$$\min_{\mathcal{G}} \max_{\mathcal{D}} \mathcal{L}_{cGAN}(\mathcal{G}, \mathcal{D}) = \mathbb{E}_{(\mathbf{X},\mathbf{Y})}[log(\mathcal{D}(\mathbf{X}, \mathbf{Y}))] + \mathbb{E}_{\mathbf{X}}[log(1 - \mathcal{D}(\mathbf{X}, \mathcal{G}(\mathbf{X})))], \tag{8.5}$$

where $\mathbf{X}$ is the observed data (given data to the generator).

## 8.2 SCALP-TO-INTRACRANIAL EEG TRANSLATION

Let $\mathbf{X} \in \mathbb{R}^{L \times M}$ and $\mathbf{Y} \in \mathbb{R}^{L \times \bar{M}}$ be respectively sEEG and iEEG, where $L$ is the number of time samples, and $M$ and $\bar{M}$ are respectively the number of channels in sEEG and iEEG. Each observed source of sEEG is considered to comprise of a number of iEEG sources plus noise. In other words, the sEEG is treated as a noisy version of the iEEG. Their relationship can be modeled as

$$\mathbf{X} = \mathbf{H}\mathbf{Y} + \textit{noise}. \tag{8.6}$$

where $\mathbf{H}$ is the mapping model. Here, we propose two methods based on GANs, namely Unet-cGAN and VAE-cGAN, as the mapping methods.

In contrast to the original GAN that maps a noise vector to a data space [141], the cGANs map an observed data sample $\mathbf{X}$ into data space $\bar{\mathbf{X}}$ [150]. Inspired by image-to-image translation methods [147, 148, 150, 151, 154] and Speech Enhancement GAN [146], which are based on cGANs, the proposed scalp-to-intracranial EEG translation methods map sEEG to iEEG using the proposed cGANs.

### 8.2.1 *Unet-cGAN*

In this mapping model, we employ a conditioning generator and discriminator networks. Our generator network is designed based on a U-net, referred to as Unet-cGAN. Our conditioning generator is fed with sEEG to generate an estimation of iEEG. Either the generated or real iEEG is concatenated with sEEG and applied to our conditioning discriminator to be classified as real or fake. Here, we set the number of scalp channels the same as that of intracranial channels since the sEEG is concatenated with the iEEG to be fed to the discriminator. Since the IEDs originate from these brain regions, we select scalp channels from temporal and frontal areas.

#### 8.2.1.1 *Generator of Unet-cGAN*

The architecture of the generator is shown in Figure 8.6. It is made up of a contracting path (left side) and an expanding path (right side). The contracting path consists of repeated convolutional layers with the filter size of $l_g \times 1$, each followed by a normalization layer operation across the time domain and an average pooling operation with stride $2 \times 1$ for downsampling. At each downsampling step, we double the number of feature channels.

Every step in the expanding path consists of an upsampling of the feature map with the size of $2 \times 1$ performed bilinearly, a $1 \times 1$ convolution, a concatenation with the corresponding feature map from the contracting path, two $l_g \times 1$ convolutions, and a

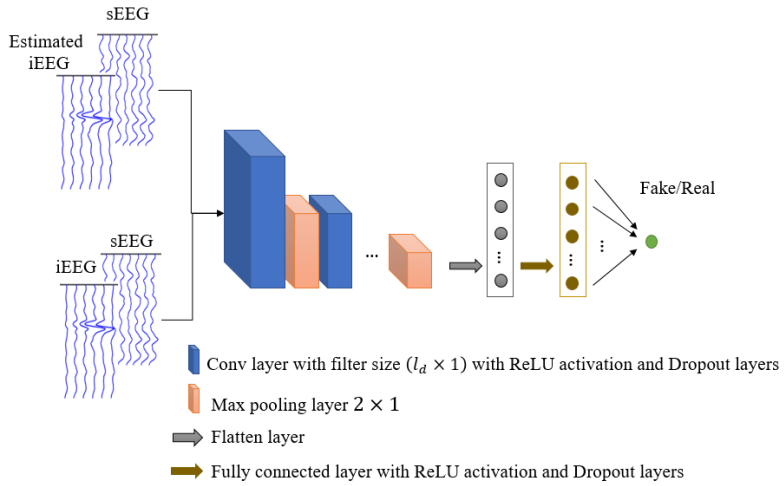



Figure 8.7: The discriminator network architecture of Unet-cGAN.

normalization layer operation across time domain. The output of the final expanding layer is fed to a $1 \times 1$ convolutional layer. Finally, a time-distributed dense layer maps each time component to $\bar{M}$ neurons (the same number as iEEG channels).

#### 8.2.1.2 *Discriminator of Unet-cGAN*

Our discriminator follows the typical architecture of a convolutional network, Figure 8.7. It consists of an $l_d \times 1$ convolutional layer followed by a rectified linear unit (ReLU) activation layer and a dropout layer, max pooling with the size of $2 \times 1$, and an FC layer. The sEEG is concatenated with either the estimated or real iEEG to be fed to the discriminator. The effectiveness of this concatenation technique has been proven in some mapping studies [146, 150].

#### 8.2.1.3 *Optimization and loss function of Unet-cGAN*

To have a more accurate estimation, we regularize the cGAN objective function, shown in (8.5), with $L_2$ loss $\mathcal{L}_{L_2}$. $L_1$ and $L_2$ loss functions are defined as the distance between the real data sample $\mathbf{Y}$ and the generated data $\widetilde{\mathbf{X}}$. In Unet-cGAN, $L_2$ loss is estimated as follows:

$$\mathcal{L}_{L_2} = \mathbb{E}_{(\widetilde{\mathbf{X}},\mathbf{Y})}[\|\mathbf{Y} - \widetilde{\mathbf{X}}\|_2]. \tag{8.7}$$

Here, the discriminator network remains unchanged, but the generator loss is coupled with $L_2$ loss and applied to train the generator:

$$\mathcal{L}_{\mathcal{G}} = \min_{\mathcal{G}} \max_{\mathcal{D}} \mathcal{L}_{cGAN}(\mathcal{G}, \mathcal{D}) + \lambda \mathcal{L}_{L_2} \tag{8.8}$$

where $\lambda$ is the coefficient of the $L_2$ loss function.

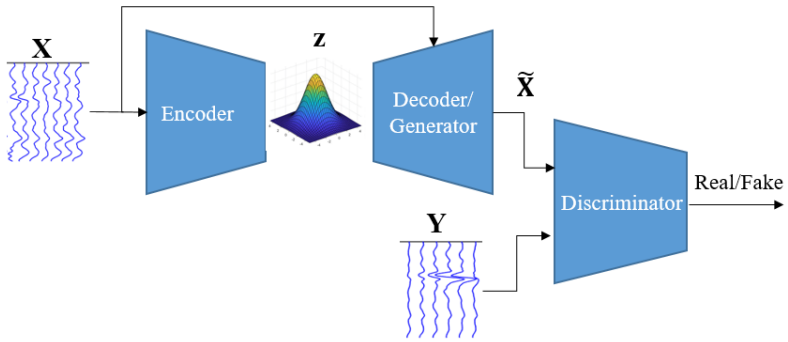



Figure 8.8: The overview of our proposed VAE-cGAN. **X** and **Y** are respectively sEEG and iEEG.

### 8.2.2 *VAE-cGAN*

In this mapping model, we combine the generator $\mathcal{G}$ of our cGAN with VAE, called VAE-cGAN, by letting them share their parameters and train jointly. The overview of our networks is shown in Figure 8.8.

Apart from differences in structures, loss functions, and their applications [155–157], the main difference between the previously developed GAN-VAE with our proposed VAE-cGAN is that both the input data sample $\mathbf{X}$ and the encoded latent space $\mathbf{z}$ are given to the generator $\mathcal{G}$ in our proposed method, while in others the input of generator is only the latent space $\mathbf{z}$.

In our VAE-cGAN model, the encoder $\mathcal{E}$ encodes the sEEG $\mathbf{X}$ to the latent space $\mathbf{z}$. The objective of the generator $\mathcal{G}$ is to translate/map the sEEG $\mathbf{X}$ as well as the latent space $\mathbf{z}$ to the iEEG $\mathbf{Y}$, where the translated signal is named estimated iEEG $\widetilde{\mathbf{X}}$. On the other hand, the goal of the discriminator $\mathcal{D}$ is to distinguish the iEEG $\mathbf{Y}$ from the estimated iEEG $\widetilde{\mathbf{X}}$. As a result, VAE-cGAN is

$$\begin{aligned} \mathbf{z} &= \mathcal{E}(\mathbf{X}), \\ \widetilde{\mathbf{X}} &= \mathcal{G}(\mathbf{X}, \mathbf{z}), \end{aligned} \tag{8.9}$$

where $\widetilde{\mathbf{X}} \in \mathbb{R}^{L \times \tilde{M}}$ is the estimated iEEG.

#### 8.2.2.1 *Encoder*

The sEEG $\mathbf{X}$ is given as the input to the encoder network. As shown in Figure 8.9, our encoder $\mathcal{E}$ consists of a sequence of convolutional layers with the filter size of $(l_e \times 1)$ and stride of $(2 \times 1)$, followed by the instance normalization (IN) [158] and leaky ReLU (LReLU) layers. The output of the last convolutional layer is projected onto a couple of dense layers which are the mean $\mu$ and STD $\sigma$ of a Gaussian distribution. Finally, the latent space with the dimension of $Z$ is obtained by scaling the Gaussian distribution $\epsilon \sim \mathcal{N}(0, \mathbf{I})$ using $\mu$ and $\sigma$, shown in (8.1).

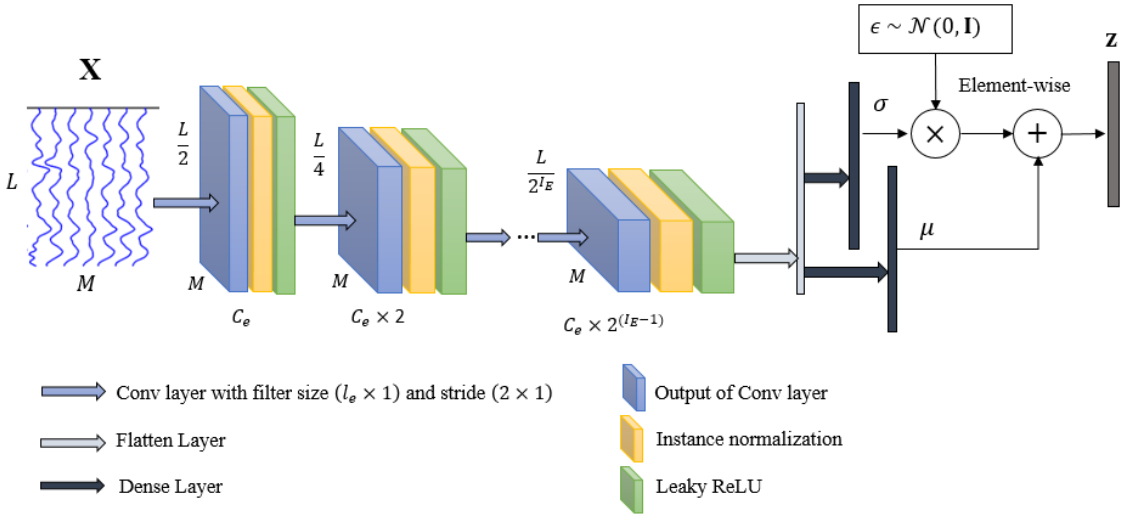



Figure 8.9: The encoder network $\mathcal{E}$ of VAE-cGAN. $C_e$ is the coefficient of the number of filters in the Conv layer, and $I_E$ is the total number of layers in the encoder.

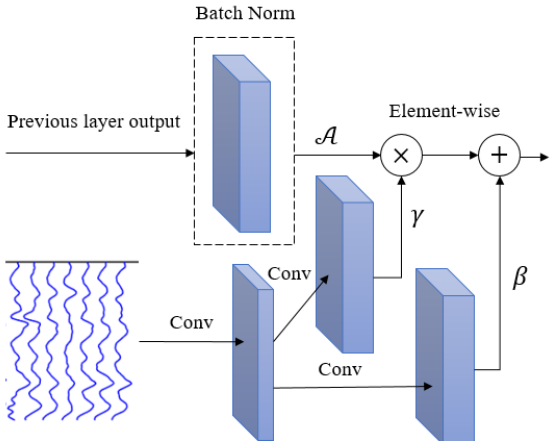



Figure 8.10: In the SPADE block, the sEEG is first projected onto an embedding space and then convolved with two different filter masks to produce the modulation parameters $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$. After normalizing the activation layer, it is multiplied by $\gamma$ and added to $\boldsymbol{\beta}$ element-wise. The SPADE is firstly proposed in [147].

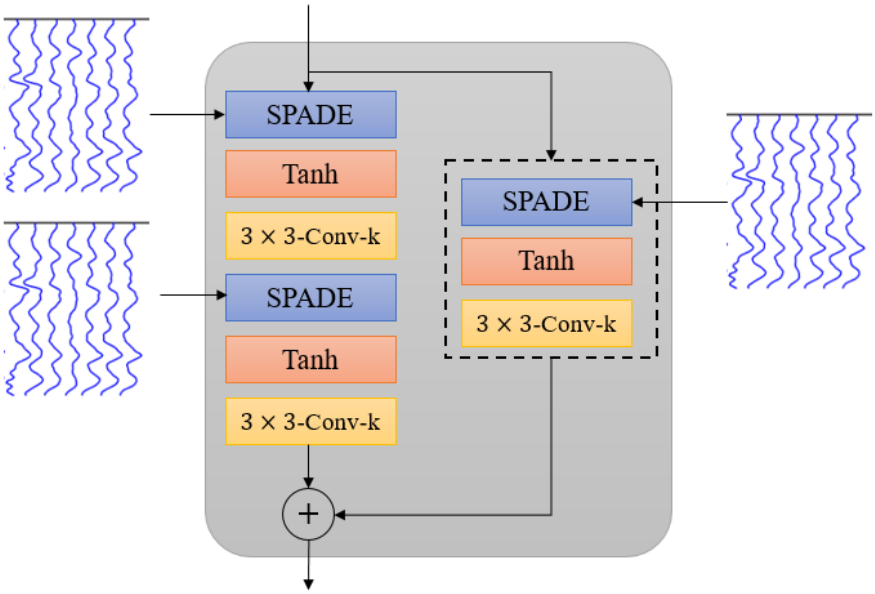



Figure 8.11: The SPADE ResNet consists of two SAPDE blocks followed by Tanh activation and convolutional layers.

#### 8.2.2.2 *Generator of VAE-cGAN*

We develop the generator $\mathcal{G}$ based on SPatially-Adaptive (DE)normalization (SPADE) [147]. SPADE is a type of conditional normalization. In contrast to standard normalization methods, conditional normalization techniques require external data. In conditional normalization, the layer activations are firstly normalized to have zero mean and unit STD. Then, the external data, here sEEG $\mathbf{X}$, is employed to denormalize the normalized activations [147, 159, 160].

The SPADE block is shown in Figure 8.10. In SPADE, the activation is normalized in a channel-wise manner, like batch normalization [161]. Then, it is multiplied by $\gamma$ and added to $\beta$ element-wise:

$$SPADE_{out} = \gamma \times \mathcal{A} + \beta, \tag{8.10}$$

where $\mathcal{A}$ is the normalized activation layer and $\gamma$ and $\beta$ are the learned modulation parameters of the normalization layer whose parameters are inferred from $\mathbf{X}$. The SPADE networks are used in the ResNet approach, called SPADE ResNet, shown in Figure 8.11.

The SPADE ResNet is made up of two SAPDE blocks which are followed by Tanh activation and convolutional layers. When the input and output dimensions of SPADE ResNet are different, a SPADE block is used instead of the skip connection for matching the dimensions as well. For more details, the reader is referred to [147].

The generator, whose architecture is shown in Figure 8.12, is fed by both the latent space $\mathbf{z}$ and the sEEG $\mathbf{X}$. Our generator is made up of a series of SPADE ResNets, which are followed by nearest neighbor upsampling with the size of $2 \times 1$ and a couple

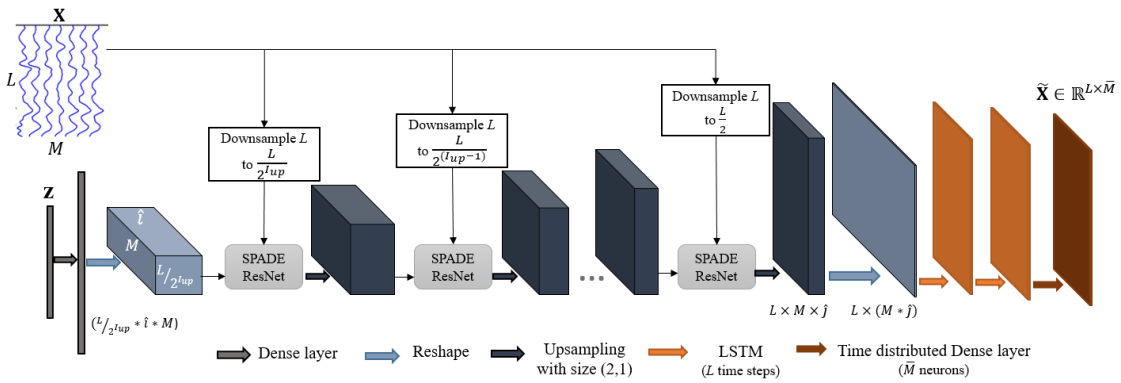



Figure 8.12: The generator network $\mathcal{G}$ of VAE-cGAN. $I_{up}$ is the number of upsampling or SPADE ResNet layers.

of LSTM layers. The sEEG given to a SPADE ResNet is downsampled according to the dimension of activation layer $\mathcal{A}$ to match the resolution. LSTM is powerful in capturing time-varying features. Therefore, the output of the last upsampling layer is applied to a couple of LSTM layers. The LSTM layers are employed with $L$ number of time steps. The output of each time step of the second LSTM is given to a dense layer with $\bar{M}$ dimensions, called time distributed dense layer, followed by the Tanh activation function. The output dimension $\widetilde{\mathbf{X}}$ is $L \times \bar{M}$, the same dimension as $\mathbf{Y}$.

#### 8.2.2.3 *Discriminator of VAE-cGAN*

In VAE-cGAN, a Markovian discriminator is employed [162]. For these discriminators, the output dimension is not 1, it is rather $p_1 \times p_2$, called patch, and the discriminator distinguishes each patch as fake or real [162, 163].

The input of the discriminator is either the iEEG $\mathbf{Y}$ or the estimated iEEG $\widetilde{\mathbf{X}}$, having $L$ time samples and $\bar{M}$ source observations. Our discriminator network is made up of a sequence of convolutional layers. Except for the last convolutional layer, the filter size of all the convolutional layers is selected to be $l_d \times 1$ with stride $2 \times 1$, followed by IN and LReLU layers. The last convolutional layer with the filter size of $l_d \times 1$ and stride of 1 does not have a normalization or activation layer. The size of the discriminator patch is set to $1 \times \bar{M}$. This means that the discriminator tries to classify each observation as either real iEEG or estimated iEEG.

#### 8.2.2.4 *Optimization and loss function of VAE-cGAN*

Depending on the dataset and application, various types of loss functions are employed to train the discriminator and generator in GANs. Here, the hinge loss, which has

already been used in an adversarial loss approach [147, 164, 165], is employed to train the discriminator $\mathcal{D}$:

$$\begin{aligned}\mathcal{L}_{\mathcal{D}} &= -\mathbb{E}_{\mathbf{Y}}[min(0, -1 + \mathcal{D}(\mathbf{Y}))] \\ &\quad -\mathbb{E}_{\mathbf{X},\mathbf{z}}[min(0, -1 - \mathcal{D}(\mathcal{G}(\mathbf{X}, \mathbf{z})))] \\ \mathcal{L}_{\mathcal{G}_h} &= -\mathbb{E}_{\mathbf{X},\mathbf{z}}[\mathcal{D}(\mathcal{G}(\mathbf{X}, \mathbf{z}))].\end{aligned} \tag{8.11}$$

Apart from the generator hinge loss $\mathcal{L}_{\mathcal{G}_h}$, another three loss functions – namely KL divergence loss $\mathcal{L}_{D_{KL}}$, $L_1$ loss $\mathcal{L}_{L_1}$, and feature matching loss $\mathcal{L}_{FM}$ – are coupled together and applied to train the generator:

$$\mathcal{L}_{\mathcal{G}} = \mathcal{L}_{\mathcal{G}_h} + \lambda_1 \mathcal{L}_{D_{KL}} + \lambda_2 \mathcal{L}_{L_1} + \lambda_3 \mathcal{L}_{FM}, \tag{8.12}$$

where $\lambda_1$, $\lambda_2$, and $\lambda_3$ are the coefficients of the corresponding loss function.

$\mathcal{L}_{D_{KL}}$ is estimated using (8.3). In Unet-cGAN, $L_2$ loss function is used as the reconstruction error. However, here $L_1$ loss provides better performance and is estimated as follows:

$$\mathcal{L}_{L_1} = \mathbb{E}_{(\widetilde{\mathbf{X}},\mathbf{Y})}[\|\mathbf{Y} - \widetilde{\mathbf{X}}\|_1]. \tag{8.13}$$

Feature matching loss $\mathcal{L}_{FM}$ has been used in image-to-image studies to improve the adversarial loss [148, 166]. It is estimated from the discriminator and makes the GAN module learn the global information using multi-scale features. The feature matching loss $\mathcal{L}_{FM}$ is obtained from multiple layers of the discriminator. It tries to match the intermediate maps of the iEEG $\mathbf{Y}$ with those of the estimated iEEG $\widetilde{\mathbf{X}}$. Let $\mathcal{D}^{(i)}$ be the $i^{th}$ layer in the discriminator $\mathcal{D}$. Then, the feature matching loss $\mathcal{L}_{FM}$ is estimated as follows:

$$\mathcal{L}_{FM} = \mathbb{E}\sum_{i=1}^{I_D} \frac{1}{F_i} |\mathcal{D}^{(i)}(\mathbf{Y}) - \mathcal{D}^{(i)}(\widetilde{\mathbf{X}})|_1, \tag{8.14}$$

where $I_D$ is the total number of layers in the discriminator, and $F_i$ is the number of features in layer $i$.

## 8.3 EXPERIMENT

The proposed mapping methods are employed to translate the sEEG to iEEG recordings. After mapping, the IEDs are detected from the estimated iEEG.

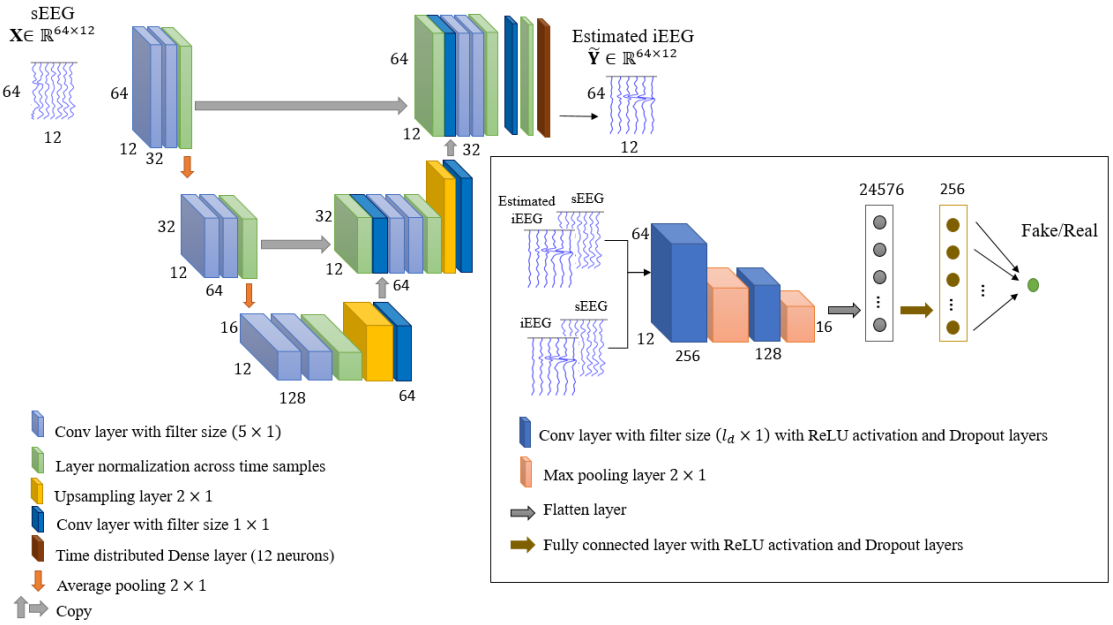



Figure 8.13: The proposed Unet-cGAN for mapping sEEG to iEEG.

### 8.3.1 *Dataset*

We use a concurrent EEG dataset in which sEEG and iEEG were recorded simultaneously from 18 people with epilepsy. The details of the dataset are given in Section 3.1.

An expert epileptologist annotated the IEDs based on the morphology and spatial distribution of the observed waveforms from the iEEG, meaning that the iEEG recordings were used as ground truth in the IED annotation. For mapping and classification, 32 samples before and after the peaks marked as the onset of IED waveforms (totally 64 samples) were selected as IED segments. The non-IED segments were randomly selected from time segments with no annotated IEDs. The same number of IEDs and

Table 8.1: The total number of IED and non-IED segments for each subject. The same number of IED and non-IED segments have been chosen for each subject.

| Subject | No. of segments | Subject | No. of segments |
|---|---|---|---|
| S1 | 684 | S10 | 488 |
| S2 | 100 | S11 | 1696 |
| S3 | 144 | S12 | 1906 |
| S4 | 330 | S13 | 1658 |
| S5 | 316 | S14 | 1082 |
| S6 | 944 | S15 | 520 |
| S7 | 398 | S16 | 1212 |
| S8 | 634 | S17 | 228 |
| S9 | 682 | S18 | 236 |

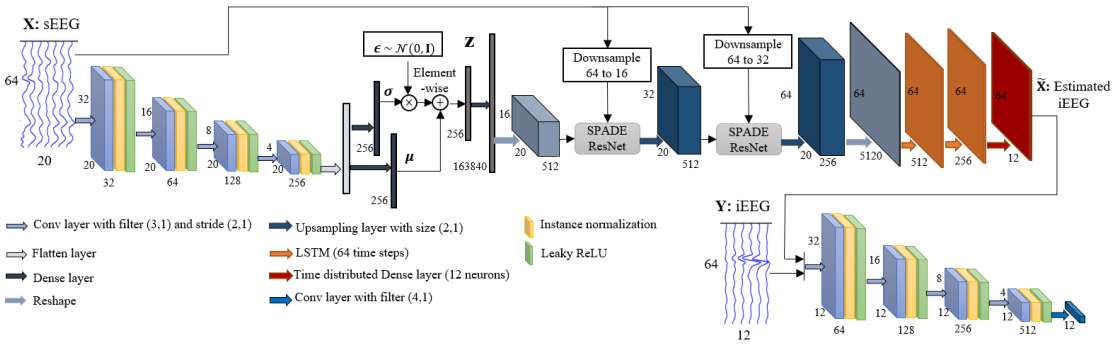



Figure 8.14: The proposed VAE-cGAN for mapping the sEEG to iEEG.

non-IEDs were selected. Table 8.1 shows the total number of IED and non-IED segments for each subject.

A highpass filter with a cutoff frequency of 1 Hz was applied to both sEEG and iEEG recordings. In addition, a notch filter with a notch frequency of 50 Hz was employed to eliminate the power line interference. Furthermore, CAR is applied to the sEEG to eliminate the artefacts.

### 8.3.2 *Translating/Mapping sEEG to iEEG*

The proposed sEEG-to-iEEG translation methods are employed to map sEEG to iEEG.

The architecture of Unet-cGAN is shown in Figure 8.13. In Unet-cGAN, the sEEG is concatenated with either the real or estimated iEEG, then fed to the discriminator. Therefore, the number of scalp channels and FO channels must be the same. Twelve out of twenty scalp channels are selected and fed to the generator network. These twelve channels, namely Fp1, F3, F7, C3, T3, Fp2, F4, F8, C4, T4, Fz, and Cz, are selected from temporal and frontal regions, where the IEDs originate from.

The whole architecture of VAE-cGAN is shown in Figure 8.14. In the training stage, the sEEG with 64 time samples and 20 observations is projected onto a latent space $\mathbf{z}$ with the dimension of 256. Then, the latent space, as well as the sEEG, is fed to the generator $\mathcal{G}$ to generate an estimation of the corresponding iEEG. Finally, the estimated iEEG and the real iEEG are given to a patch discriminator.

### 8.3.3 *Classification*

We employ EEGNet [167] to detect IEDs from the estimated iEEG. The effectiveness of EEGNet in EEG classification has been proven [167, 168]. For classifying IEDs and non-IEDs, an EEGNet with minor changes is employed. The EEGNet consists of three blocks.

Table 8.2: EEGNet architecture. The rate of Dropout layer is set to 0.2, r=0.2.

| Block | Layer | No. of filters | Size of filters | Output size | Options |
|---|---|---|---|---|---|
| 1 | Input | - | - | (64, 12, 1) | - |
| | Conv2D | 128 | (5, 1) | (64, 12, 128) | Mode=same |
| | DepthwiseConv | $2 \times 128$ | (1, 12) | (64, 1, 256) | Mode=valid, depth=2 |
| | Activation | - | - | (64, 1, 256) | Type=ReLU |
| | MaxPooling | - | (2,1) | (32, 1, 256) | - |
| | Dropout | - | - | (32, 1, 256) | r=0.2 |
| 2 | SeparableConv | 256 | (5, 1) | (32, 1, 256) | Mode=same, depth=1 |
| | Activation | - | - | (32, 1, 256) | Type=ReLU |
| | MaxPooling | - | (2,1) | (16, 1, 256) | - |
| | Dropout | - | - | (16, 1, 256) | r=0.2 |
| 3 | Flatten | - | - | 4096 | - |
| | Dense | - | - | 512 | No. of neurons=512 |
| | Activation | - | - | 512 | Type=ReLU |
| | Dropout | - | - | 512 | r=0.2 |
| | Sigmoid | - | - | 1 | - |

- In block 1, the estimated iEEG is fed to a convolutional layer with 128 filters of size $5 \times 1$. We then use a "depthwise" convolution layer of size $12 \times 1$ to learn a spatial filter. The size is set to 12 to contain all channels. The "depthwise" convolution is a type of convolution in which each input channel is convolved with a different kernel (called a depthwise kernel). It is the first step in a depthwise separable convolution [169]. The depthwise kernel is set to 2. We add ReLU as the nonlinear activation function and the Dropout layer for regularization. Then, Max pooling of size $2 \times 1$ is added.

- In block 2, the output of block 1 is fed to a Separable convolution layer with 256 filters of size $5 \times 1$. In the Separable convolution layer, a depthwise spatial convolution (which acts on each input channel separately) is first performed, then it is followed by a pointwise convolution which mixes the resulting output channels. The Separable convolution is followed by the ReLU, Dropout, and Max pooling layers like block 1.

- In block 3, the output of block 2 is flattened, then fed to a dense layer of size 512, followed by the ReLU and Dropout layers. Finally, we have a sigmoid layer to identify samples as IED or non-IED.

### 8.3.4 *Cross Validation*

The IEDs are detected in the within- and between-subject classification approaches. In the within-subject classification approach, the data of a subject is divided into training

(70%), validation (10%), and test datasets (20%). On the other hand, in the between-subject classification, the data from $N$ subjects are used for training and the data of a subject for testing. Here, $N$ is the number of subjects whose IEDs are detected with high accuracy in a within-subject classification approach. This approach is repeated for all 18 subjects. In this classification approach, the sEEG of all $N$ training subjects and the test subject are mapped to the iEEG using each of the trained Unet-cGAN or VAE-cGAN ($\mathcal{G}_n,\ n = \{1, 2, \ldots, N\}$) to obtain the estimated iEEG, $\mathbf{X} \xrightarrow{\mathcal{G}_n} \widetilde{\mathbf{X}}_n$. Then, each of the estimated iEEG $\widetilde{\mathbf{X}}_n$ is given to the EEGNet to classify IEDs and non-IEDs. Finally, to find the segment labels in the test data, the output probabilities of $N$ EEGNets are averaged (average voting classification). Figure 8.15 shows the diagram of the between-subject classification approach.

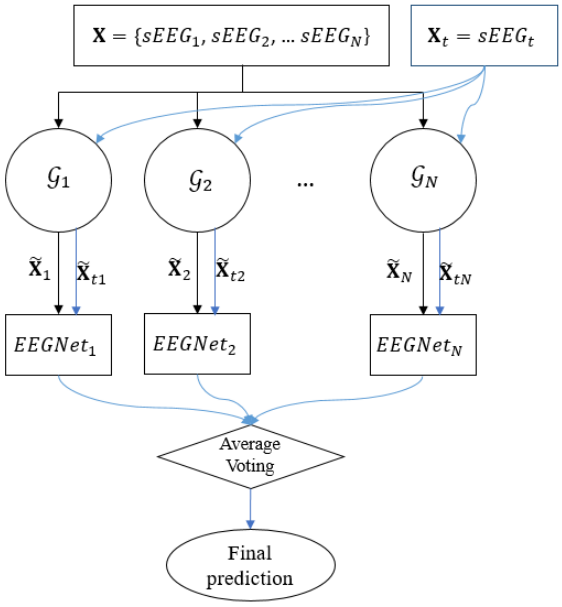



Figure 8.15: The diagram of the between-subject classification approach.

## 8.4 EXPERIMENTAL RESULTS

ACC, SEN, and SPC are obtained as the evaluation criteria. Our proposed Unet-cGAN and VAE-cGAN techniques are compared with three previously developed methods [9, 133], namely LSR, AAE, and ASAE, explained in Section 7.2.5.

The sEEG, the iEEG, and the estimated iEEG obtained from VAE-cGAN are shown in Figure 8.16. The estimated iEEG (middle row) exactly follows the iEEG in the IED samples. While the epileptic spikes are not observable in the sEEG, the estimated iEEG nicely follows the real iEEG trend, and the IEDs are well pronounced. In the non-IEDs, the estimated iEEG follows the trend of the iEEG.

Table 8.3: The classification accuracy of our proposed Unet-cGAN and VAE-cGAN methods and the compared LSR, AAE, and ASAE methods in the between-subject classification approach. The amounts in parenthesis show the within-subject classification performance. The values are in percent (%).

| Subject | LSR [133] | AAE[9] | ASAE[9] | Unet-GAN | VAE-cGAN |
|---|---|---|---|---|---|
| S1 | 65 (72) | 85 (80) | 87 (78) | 67 (78) | 75 (88) |
| S2 | 86 (81) | 92 (82) | 94 (88) | 83 (95) | 80 (95) |
| S3 | 65 (69) | 72 (72) | 69 (82) | 74 (90) | 75 (72) |
| S4 | 58 (62) | 58 (71) | 59 (77) | 66 (81) | 64 (76) |
| S5 | 55 (55) | 64 (64) | 65 (75) | 67 (73) | 67 (78) |
| S6 | 61 (59) | 70 (60) | 71 (63) | 68 (68) | 72 (73) |
| S7 | 59 (64) | 54 (62) | 67 (72) | 64 (67) | 63 (61) |
| S8 | 55 (66) | 55 (62) | 57 (68) | 63 (72) | 60 (68) |
| S9 | 63 (65) | 61 (74) | 62 (68) | 61 (71) | 72 (76) |
| S10 | 66 (70) | 71 (65) | 74 (77) | 75 (91) | 87 (87) |
| S11 | 63 (64) | 65 (67) | 65 (68) | 61 (62) | 63 (62) |
| S12 | 73 (79) | 75 (84) | 77 (84) | 79 (84) | 73 (87) |
| S13 | 62 (71) | 62 (72) | 64 (71) | 63 (74) | 68 (77) |
| S14 | 59 (62) | 66 (71) | 67 (65) | 63 (69) | 62 (74) |
| S15 | 50 (46) | 50 (53) | 50 (52) | 55 (59) | 53 (62) |
| S16 | 51 (55) | 67 (77) | 68 (72) | 75 (77) | 74 (82) |
| S17 | 54 (62) | 59 (54) | 62 (71) | 66 (78) | 64 (80) |
| S18 | 66 (64) | 61 (53) | 67 (75) | 65 (72) | 64 (66) |
| Mean | 62 (65) | 66 (68) | 68 (73) | 68 **(76)** | **69 (76)** |

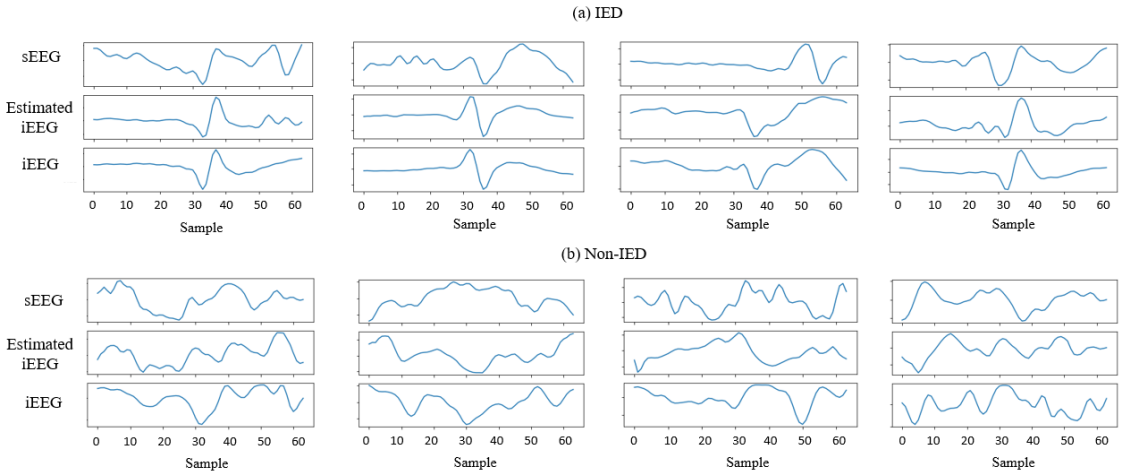



Figure 8.16: Samples of (a) IEDs and (b) non-IEDs averaged over all channels. The sEEG, iEEG, and estimated iEEG obtianed using VAE-cGAN are shown. The IEDs start at time sample 32, and the sampling rate is 200 samples/sec.

Table 8.4: The SEN and SPC obtained using our proposed Unet-GAN and VAE-cGAN methods and the compared LSR, AAE, and ASAE methods in the between-subject classification approach. The values are in percent (%).

| Criteria | LSR[133] | AAE[9] | ASAE[9] | Unet-GAN | VAE-cGAN |
|---|---|---|---|---|---|
| SEN | 61 | 66 | 67 | 65 | **68** |
| SPC | 59 | 65 | 68 | **70** | **70** |

The obtained accuracy levels of our proposed Unet-cGAN and VAE-cGAN methods and the compared methods (LSR, AAE, and ASAE) are provided in Table 8.3. The Table shows the obtained ACC from both within- and between-subject classification approaches. Our proposed methods provide the best ACC in the within-subject classification approach. Both methods obtain 76% ACC, which is respectively 11%, 8%, and 3% higher than the obtained ACC values using LSR, AAE, and ASAE. In the between-subject classification approach, the VAE-cGAN has an increase ACC of respectively 7%, 3%, and 1% over LSR, AAE, and ASAE. It also outperforms Unet-GAN by 1%.

SEN and SPC values are shown in Table 8.4. In both criteria, the VAE-cGAN provides the best values of 68% and 70%, respectively. Unet-cGAN provides 65% SEN and 70% SPC. The SEN and SPC values achieved by ASAE are close to VAE-cGAN. Its reason is that, in ASAE, the data is mapped onto the iEEG two times. In contrast, in the VAE-cGAN, we map the sEEG to iEEG one time.

## 8.5 CONCLUSION

This study proposes two EEG-to-EEG translation methods based on cGAN to transfer a low-resolution EEG to a high-resolution counterpart to best highlight the IEDs. A U-net is employed as the generator in the first one, named Unet-GAN. On the other hand, in VAE-cGAN, the generator is designed based on the VAE architecture. The VAE and the cGAN are trained together by sharing parameters. This improves the performance of cGAN in generating high-quality data samples. A Markovian discriminator (patch discriminator) [162] is used to distinguish the real from the estimated iEEG. In the patch discriminator, each channel of real and estimated iEEG is projected onto a patch using a CNN, meaning that the discriminator classifies each channel as either real or estimated iEEG. The proposed systems learn from jointly recorded iEEG and sEEG and are then applied to the classification of sEEG. This enables the detection of IEDs that are not visible from over the scalp.

Our proposed methods are employed to map the sEEG to iEEG recordings collected from epileptic patients containing epileptic spikes/IEDs. The VAE-cGAN provides higher ACC, SEN, and SPC compared to Unet-cGAN and the previously developed methods [9, 133]. It achieves an average ACC of 69% for the between-subject classification approach, which is 1% higher than the Unet-cGAN ACC and respectively 7%,

3%, and 1% higher than ACC values obtained using LSR, AAE, and ASAE. In addition, the VAE-cGAN and Unet-cGAN outperformed the compared methods in the within-subject classification approach by achieving the average ACC of 76%, which is respectively 11%, 8%, and 3% higher than ACC values achieved using LSR, AAE, and ASAE. However, here the trade-off is between the higher accuracy of the system and its high computational complexity as for other DNN algorithms.

Due to conditioning the discriminator in Unet-cGAN, we used the same number of sEEG and iEEG channels. This means we have not exploited the full sEEG multichannel information in Unet-cGAN. We overcome this limitation using VAE-cGAN, which includes all sEEG channels in our analysis. Also, in this study, we assumed the background brain activity is additive uncorrelated noise. This is not an accurate assumption. The noise can change with the brain's state. Therefore, a more flexible algorithm together with more accurate assumptions may improve the results. This can be a subject of our future work.

# 9 CONCLUSION AND FUTURE RESEARCH

IEDs are transient activities occurring between two seizure onsets. Automatic detection of IEDs helps manage and monitor epilepsy patients. Therefore, in clinical practice, it is crucial to automatically detect IEDs using widely available EEG systems with high sensitivity. However, the problem is that IEDs are generated from deep sources and mainly captured by iEEG recordings. They are attenuated by the skull and either appear as very weak spikes or sharp waves on sEEG recordings or disappear completely. Consequently, a large proportion of IEDs are invisible over the scalp, and 30% to 40% of patients considered for epilepsy surgery require invasive iEEG recordings. To overcome this limitation and detect all visible and invisible IEDs from the sEEG, we started with detecting IEDs from the concurrent sEEG and iEEG recordings and then developed models to extract the iEEG information from the sEEG.

In this thesis, we first developed methods based on tensor factorisation to automatically detect IEDs from the concurrent sEEG. Then, we proposed an algorithm based on sparse regularization. Finally, we developed two different methods based on multiway analysis and GAN to map the sEEG to iEEG to boost the sensitivity of sEEG. This work opens up several new interesting and important fronts for further research and investigation. In this chapter, we conclude the thesis with an overview of the main contributions and propose new possible directions for further investigations.

## 9.1 REAL TIME IED DETECTION

Detecting IEDs from the ongoing EEG is more critical than the segmented EEG. In the real world, the EEG is not segmented, and it is needed to be detected from the ongoing signals. Meanwhile, most studies detected IEDs from only the segmented signals. In Chapter 3, we have proposed a method based on tensor factorisation to detect IEDs from ongoing sEEG and iEEG. In this work, we detected the IEDs from the ongoing sEEG and iEEG recorded simultaneously. Since the data has not been segmented, we slide a window with a high percentage of overlap over the signal.

IEDs originate from the temporal brain regions, and their signatures can be captured using several electrodes simultaneously. Therefore, it is expected to have temporal and spatial correlation among IED segments. Thus, multi-channel and multi-segment EEG processing using higher dimensional decomposition methods can effectively detect IEDs. On the other hand, non-IEDs are random, and there is no shared feature among them. Therefore, spatial and temporal components of IED can provide the most

distinguishable features. To exploit temporal and spatial components, we concatenate all IEDs of training data into a three-way tensor (time, channel, segment) and decompose the tensor into temporal, spatial, and segmental factors using CPD. Finally, we employ temporal and spatial components for IED detection. The proposed approach provides promising SEN values in detecting IEDs from the ongoing sEEG and iEEG.

## 9.2 EXPLOITING THE MORPHOLOGY OF IEDS IN THE TENSOR STRUCTURE

A tensor is a multi-way data representation. Tensor factorisation applies to both non-stationary and under-determined cases. The signal, which generally changes in time, space, and morphology, can be identified using tensor factorisation. In terms of morphology, IEDs can be various. In Chapter 4, we have exploited the morphology of IEDs within the tensor structure. We construct a four-way tensor with the dimension of time, channel, segment, and morphology, then decompose the tensor into temporal, spatial, segmental, and morphological factors using CPD. Finally, we use spatial and morphological factors for IED detection. The method has been applied to detect IEDs from the iEEG. In this work, we show that exploiting the IED morphology can boost the performance of an IED detection system. However, the IED morphologies change with age [49]. Therefore, to generalize the application, we should have access to the data from a wider age range. As part of future work, individuals with a wider age range, from children to older adults, can be participated in the study.

## 9.3 INCORPORATING LABELING UNCERTAINTY IN AN IED DETECTION SYSTEM

As mentioned in Section 9.2, IEDs appear in various morphologies: sharp waves, spikes, or poly-spikes, often followed by an inhibitory damped oscillation. Source locations of IEDs are different among subjects as well. Furthermore, there is a similarity between IED waveforms and physiological artefacts or normal brain activity waveforms. The mentioned characteristics of IEDs cause uncertainty in IED labeling. In Chapter 5, we have incorporated this uncertainty into an IED detection system.

We construct a three-way data tensor (time, channel, and segment) and a probability (uncertainty) tensor with the same size as the data tensor. The probability tensor is constructed based on the uncertainty level assigned to each IED segment by an expert. We employ CP-WOPT for decomposing the data tensor based on the probability tensor into temporal, spatial, and segmental factors. Finally, the spatial components are used for IED detection from the sEEG.

This work has been applied to both between- and within-subject classification approaches. However, this work is based on spatial component analysis. Consequently, its performance is not very good in the between-subject classification approach owing to different IED source locations in different subjects. As part of future studies, a source localization algorithm can be used to group the subjects according to their IED locations before applying a between-subject-based approach.

## 9.4 SPARSE COMMON FEATURE ANALYSIS FOR IED DETECTION

IEDs are naturally sparse in temporal and spatial domains. In addition, as mentioned in Section 9.1, there are common features among the IED segments. In Chapter 6, we have proposed a new algorithm to exploit sparse common features among the IED segments. This algorithm can be used for other types of signals. However, we have evaluated the performance of our proposed method on an IED detection system. By exploiting sparse common features, we have improved the system's performance compared to the systems without the sparsity constraint.

Our proposed method can extract sparse common features from a three-dimensional dataset. As part of future studies, the algorithm can be extended to extract the sparse common features of high-dimensional datasets. By doing so, we can exploit other diversities of data, such as morphology.

## 9.5 MULTI-WAY ANALYSIS FOR MAPPING SEEG TO IEEG

The sEEG is easy to use and is non-invasive. However, the main problem is that the sEEG electrodes are not close to the IED sources. On the other hand, the iEEG channels are placed directly on the exposed surface of the brain to record electrical activity from the cerebral cortex. However, the problem with iEEG recordings is that they are recorded using invasive techniques and have side effects for patients. In Chapter 7, we proposed a method for mapping sEEG to iEEG based on multi-way analysis.

Multi-way analysis provides an opportunity to consider different domains of data together. Our proposed method had three stages. In the first stage, we extract TF features of the iEEG segments, then we construct a four-way tensor with the dimensions of time, frequency, channel, and segment. In the second stage, the tensor is decomposed using CPD and TD into factor matrices (temporal, spectral, spatial, and segmental factors). To map the TF features of the sEEG to iEEG, in the third stage, the TF features of sEEG are projected onto the temporal factors, and the projected segments are used for IED detection. The results show that the mapping model can significantly boost the performance of the model, particularly in the between-subject approach.

For obtaining TF features, we apply CWT in this work. Since the magnitudes of continuous wavelet coefficients are non-negative, non-negative-based tensor decomposition is likely to perform better. In future work, non-negative tensor factorisation methods can be considered.

## 9.6 MAPPING THE SEEG TO IEEG USING GANS

GANs can be used for generating data samples from random noise (latent space) or given samples. In Chapter 8, we have proposed two deep networks based on GANs for mapping sEEG to iEEG. In our proposed methods, the developed GANs generate an estimation of iEEG from the given sEEG. The generated signals are used for IED detection.

The application of DNNs and GANs is fast growing. As part of future work, more complex GANs can be developed for mapping. Furthermore, Transformer neural networks introduced recently [170] provide promising performance for machine translation [171]. They may provide high performance for EEG-to-EEG translation as well. This type of network can be considered in future work.

# BIBLIOGRAPHY


[1] Radek Janca, Petr Jezdik, Roman Cmejla, Martin Tomasek, Gregory A Worrell, Matt Stead, Joost Wagenaar, John GR Jefferys, Pavel Krsek, Vladimir Komarek, et al. "Detection of interictal epileptiform discharges using signal envelope distribution modelling: application to epileptic and non-epileptic intracranial recordings." In: *Brain Topography* 28.1 (2015), pp. 172–183.

[2] Saeid Sanei and Jonathon A Chambers. *EEG Signal Processing and Machine Learning*. John Wiley & Sons, 2021.

[3] World Health Organization. Epilepsy. "https://www.who.int/en/news-room/fact-sheets/detail/epilepsy." In: *Accessed date:* (20 June 2019).

[4] Nick Kane, Jayant Acharya, Sandor Beniczky, Luis Caboclo, Simon Finnigan, Peter W Kaplan, Hiroshi Shibasaki, Ronit Pressler, and Michel JAM van Putten. "A revised glossary of terms most commonly used by clinical electroencephalographers and updated proposal for the report format of the EEG findings. Revision 2017." In: *Clinical Neurophysiology Practice* 2 (2017), p. 170.

[5] Stjepana Kovac, Vejay N Vakharia, Catherine Scott, and Beate Diehl. "Invasive epilepsy surgery evaluation." In: *Seizure* 44 (2017), pp. 125–136.

[6] Dinesh Nayak, Antonio Valentin, Gonzalo Alarcon, Jorge J Garcia Seoane, Franz Brunnhuber, Jane Juler, Charles E Polkey, and Colin D Binnie. "Characteristics of scalp electrical fields associated with deep medial temporal epileptiform discharges." In: *Clinical Neurophysiology* 115.6 (2004), pp. 1423–1435.

[7] Madoka Yamazaki, Don M Tucker, Ayataka Fujimoto, Tomohiro Yamazoe, Tohru Okanishi, Takuya Yokota, Hideo Enoki, and Takamichi Yamamoto. "Comparison of dense array EEG with simultaneous intracranial EEG for interictal spike detection and localization." In: *Epilepsy Research* 98.2-3 (2012), pp. 166–173.

[8] Loukianos Spyrou, David Martín-Lopez, Antonio Valentín, Gonzalo Alarcón, and Saeid Sanei. "Detection of intracranial signatures of interictal epileptiform discharges from concurrent scalp EEG." In: *International Journal of Neural Systems* 26.04 (2016), p. 1650016.

[9] Andreas Antoniades, Loukianos Spyrou, David Martin-Lopez, Antonio Valentin, Gonzalo Alarcon, Saeid Sanei, and Clive Cheong Took. "Deep neural architectures for mapping scalp to intracranial EEG." In: *International Journal of Neural Systems* 28.08 (2018), p. 1850009.

[10] Le Trung Thanh, Nguyen Thi Anh Dao, Nguyen Viet Dung, Nguyen Linh Trung, and Karim Abed-Meraim. "Multi-channel EEG epileptic spike detection by a new method of tensor decomposition." In: *Journal of Neural Engineering* 17.1 (2020), p. 016023.

[11] W Robert S Webber, B Litt, RP Lesser, RS Fisher, and I Bankman. "Automatic EEG spike detection: What should the computer imitate?" In: *Electroencephalography and Clinical Neurophysiology* 87.6 (1993), pp. 364–373.

[12] Jonathan J Halford, M Brandon Westover, Suzette M LaRoche, Micheal P Macken, Ekrem Kutluay, Jonathan C Edwards, Leonardo Bonilha, Giridhar P Kalamangalam, Kan Ding, Jennifer L Hopp, et al. "Interictal epileptiform discharge detection in EEG in different practice settings." In: *Journal of Clinical Neurophysiology* 35.5 (2018), p. 375.

[13] Kosuke Fukumori, Noboru Yoshida, Hidenori Sugano, Madoka Nakajima, and Toshihisa Tanaka. "Epileptic spike detection by recurrent neural networks with self-attention mechanism." In: *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE. 2022, pp. 1406–1410.

[14] Boxuan Wei, Xiaohui Zhao, Lijuan Shi, Lu Xu, Tao Liu, and Jicong Zhang. "A deep learning framework with multi-perspective fusion for interictal epileptiform discharges detection in scalp electroencephalogram." In: *Journal of Neural Engineering* 18.4 (2021), 0460b3.

[15] Loukianos Spyrou, Samaneh Kouchaki, and Saeid Sanei. "Multiview classification and dimensionality reduction of scalp and intracranial EEG data through tensor factorisation." In: *Journal of Signal Processing Systems* 90.2 (2018), pp. 273–284.

[16] CD Binnie, E Dekker, A Smit, and G Van der Linden. "Practical considerations in the positioning of EEG electrodes." In: *Electroencephalography and Clinical Neurophysiology* 53.4 (1982), pp. 453–458.

[17] Robert J Quon, Stephen Meisenhelter, Edward J Camp, Markus E Testorf, Yinchen Song, Qingyuan Song, George W Culler, Payam Moein, and Barbara C Jobst. "AiED: Artificial intelligence for the detection of intracranial interictal epileptiform discharges." In: *Clinical Neurophysiology* 133 (2022), pp. 1–8.

[18] HG Wieser, CE Elger, and SRG Stodieck. "The 'foramen ovale electrode': a new recording method for the preoperative evaluation of patients suffering from mesio-basal temporal lobe epilepsy." In: *Electroencephalography and Clinical Neurophysiology* 61.4 (1985), pp. 314–322.

[19] Maurice Abou Jaoude, Jin Jing, Haoqi Sun, Claire S Jacobs, Kyle R Pellerin, M Brandon Westover, Sydney S Cash, and Alice D Lam. "Detection of mesial temporal lobe epileptiform discharges on intracranial electrodes using deep learning." In: *Clinical Neurophysiology* 131.1 (2020), pp. 133–141.

[20] Andreas Antoniades, Loukianos Spyrou, David Martin-Lopez, Antonio Valentin, Gonzalo Alarcon, Saeid Sanei, and Clive Cheong Took. "Detection of interictal discharges with convolutional neural networks using discrete ordered multichannel intracranial EEG." In: *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 25.12 (2017), pp. 2285–2294.

[21] Sameer A Sheth, Joshua P Aronson, Mouhsin M Shafi, H Wesley Phillips, Naymee Velez-Ruiz, Brian P Walcott, Churl-Su Kwon, Matthew K Mian, Andrew R Dykstra, Andrew Cole, et al. "Utility of foramen ovale electrodes in mesial temporal lobe epilepsy." In: *Epilepsia* 55.5 (2014), pp. 713–724.

[22] Matthew Sparkes, Antonio Valentin, and Gonzalo Alarcon. "Mechanisms involved in the conduction of anterior temporal epileptiform discharges to the scalp." In: *Clinical Neurophysiology* 120.12 (2009), pp. 2063–2070.

[23] Meysam Golmohammadi, Amir Hossein Harati Nejad Torbati, Silvia Lopez de Diego, Iyad Obeid, and Joseph Picone. "Automatic analysis of EEGs using big data and hybrid deep learning architectures." In: *Frontiers in Human Neuroscience* 13 (2019), p. 76.

[24] Loukianos Spyrou, Samaneh Kouchaki, and Saeid Sanei. "Multiview classification of brain data through tensor factorisation." In: *2015 IEEE 25th International Workshop on Machine Learning for Signal Processing (MLSP)*. IEEE. 2015, pp. 1–6.

[25] Andreas Antoniades, Loukianos Spyrou, Clive Cheong Took, and Saeid Sanei. "Deep learning for epileptic intracranial EEG data." In: *2016 IEEE 26th International Workshop on Machine Learning for Signal Processing (MLSP)*. IEEE. 2016, pp. 1–6.

[26] John Thomas, Jing Jin, Prasanth Thangavel, Elham Bagheri, Rajamanickam Yuvaraj, Justin Dauwels, Rahul Rathakrishnan, Jonathan J Halford, Sydney S Cash, and Brandon Westover. "Automated Detection of Interictal Epileptiform Discharges from Scalp Electroencephalograms by Convolutional Neural Networks." In: *International Journal of Neural Systems* (2020), pp. 2050030–2050030.

[27] Thangavel Prasanth, John Thomas, R Yuvaraj, Jing Jing, Sydney S Cash, Rima Chaudhari, Tan Yee Leng, Rahul Rathakrishnan, Srivastava Rohit, Vinay Saini, et al. "Deep Learning for Interictal Epileptiform Spike Detection from scalp EEG frequency sub bands." In: *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. IEEE. 2020, pp. 3703–3706.

[28] Kosuke Fukumori, Noboru Yoshida, Hidenori Sugano, Madoka Nakajima, and Toshihisa Tanaka. "Epileptic Spike Detection by Using a Linear-Phase Convolutional Neural Network." In: (2020). DOI: 10.1101/2020.10.08.330936.

[29] K.P. Indiradevi, Elizabeth Elias, P.S. Sathidevi, S. Dinesh Nayak, and K. Radhakrishnan. "A multi-level wavelet approach for automatic detection of epileptic spikes in the electroencephalogram." In: *Computers in Biology and Medicine* 38.7 (2008), pp. 805–816. ISSN: 0010-4825.

[30] Thanh Xuyen Le, Trung Thanh Le, Van Viet Dinh, Quoc Long Tran, Linh Trung Nguyen, and Duc Thuan Nguyen. "Deep learning for epileptic spike detection." In: *VNU Journal of Science: Computer Science and Communication Engineering* 33.2 (2018), pp. 1–13.

[31] Shaun S Lodder, Jessica Askamp, and Michel JAM van Putten. "Inter-ictal spike detection using a database of smart templates." In: *Clinical Neurophysiology* 124.12 (2013), pp. 2328–2335.

[32] Shaun S Lodder and Michel JAM van Putten. "A self-adapting system for the automated detection of inter-ictal epileptiform discharges." In: *PloS one* 9.1 (2014), e85180.

[33] Z. Wang, D. Wu, F. Dong, J. Cao, T. Jiang, and J. Liu. "A Novel Spike Detection Algorithm Based on Multi-Channel of BECT EEG Signals." In: *IEEE Transactions on Circuits and Systems II: Express Briefs* 67.12 (2020), pp. 3592–3596.

[34] Alexander Rosenberg Johansen, Jing Jin, Tomasz Maszczyk, Justin Dauwels, Sydney S Cash, and M Brandon Westover. "Epileptiform spike detection via convolutional neural networks." In: *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE. 2016, pp. 754–758.

[35] John Thomas, Jing Jin, Justin Dauwels, Sydney S Cash, and M Brandon Westover. "Automated epileptiform spike detection via affinity propagation-based template matching." In: *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE. 2017, pp. 3057–3060.

[36] Janice R Stevens, Brenda L Lonsbury, and Sunder L Goel. "Seizure occurrence and interspike interval: Telemetered electroencephalogram studies." In: *Archives of Neurology* 26.5 (1972), pp. 409–419.

[37] K Vijayalakshmi and Appaji M Abhishek. "Spike detection in epileptic patients EEG data using template matching technique." In: *International Journal of Computer Applications* 2.6 (2010), pp. 5–8.

[38] Zhanfeng Ji, Takenao Sugi, Satoru Goto, Xingyu Wang, Akio Ikeda, Takashi Nagamine, Hiroshi Shibasaki, and Masatoshi Nakamura. "An automatic spike detection system based on elimination of false positives using the large-area context in the scalp EEG." In: *IEEE Transactions on Biomedical Engineering* 58.9 (2011), pp. 2478–2488.

[39] J Jing, J Dauwels, T Rakthanmanon, E Keogh, SS Cash, and MB Westover. "Rapid annotation of interictal epileptiform discharges via template matching under dynamic time warping." In: *Journal of Neuroscience Methods* 274 (2016), pp. 179–190.

[40] R Sankar and J Natour. "Automatic computer analysis of transients in EEG." In: *Computers in Biology and Medicine* 22.6 (1992), pp. 407–422.

[41] Mahmoud El-Gohary, James McNames, and Siegward Elsas. "User-guided interictal spike detection." In: *2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE. 2008, pp. 821–824.

[42] J Gotman and P Gloor. "Automatic recognition and quantification of interictal epileptic activity in the human scalp EEG." In: *Electroencephalography and Clinical Neurophysiology* 41.5 (1976), pp. 513–529.

[43] Claudie Faure. "Attributed strings for recognition of epileptic transients in EEG." In: *International Journal of Bio-Medical Computing* 16.3-4 (1985), pp. 217–229.

[44] WRS Webber, Brian Litt, K Wilson, and Ronald P Lesser. "Practical detection of epileptiform discharges (EDs) in the EEG using an artificial neural network: A comparison of raw and parameterized EEG data." In: *Electroencephalography and Clinical Neurophysiology* 91.3 (1994), pp. 194–204.

[45] Alison A Dingle, Richard D Jones, Grant J Carroll, and W Richard Fright. "A multistage system to detect epileptiform activity in the EEG." In: *IEEE Transactions on Biomedical Engineering* 40.12 (1993), pp. 1260–1268.

[46] R Benlamri, M Batouche, S Rami, and C Bouanaka. "An automated system for analysis and interpretation of epileptiform activity in the EEG." In: *Computers in Biology and Medicine* 27.2 (1997), pp. 129–139.

[47] Yung-Chun Liu, Chou-Ching K Lin, Jing-Jane Tsai, and Yung-Nien Sun. "Model-based spike detection of epileptic EEG data." In: *Sensors* 13.9 (2013), pp. 12536–12547.

[48] Zimeng Wang, Duanpo Wu, Fang Dong, Jiuwen Cao, Tiejia Jiang, and Junbiao Liu. "A novel spike detection algorithm based on multi-channel of BECT EEG signals." In: *IEEE Transactions on Circuits and Systems II: Express Briefs* 67.12 (2020), pp. 3592–3596.

[49] Eivind Aanestad, Nils Erik Gilhus, and Jan Brogger. "Interictal epileptiform discharges vary across age groups." In: *Clinical Neurophysiology* 131.1 (2020), pp. 25–33.

[50] He Sheng Liu, Tong Zhang, and Fu Sheng Yang. "A multistage, multimethod approach for automatic detection and classification of epileptiform EEG." In: *IEEE Transactions on Biomedical Engineering* 49.12 (2002), pp. 1557–1566.

[51] Flavio Sartoretto and Mario Ermani. "Automatic detection of epileptiform activity by single-level wavelet analysis." In: *Clinical Neurophysiology* 110.2 (1999), pp. 239–249.

[52] Hansjerg Goelz, Richard D Jones, and Philip J Bones. "Wavelet analysis of transient biomedical signals and its application to detection of epileptiform activity in the EEG." In: *Clinical Electroencephalography* 31.4 (2000), pp. 181–191.

[53] KP Indiradevi, Elizabeth Elias, PS Sathidevi, S Dinesh Nayak, and K Radhakrishnan. "A multi-level wavelet approach for automatic detection of epileptic spikes in the electroencephalogram." In: *Computers in Biology and Medicine* 38.7 (2008), pp. 805–816.

[54] Elham Bagheri, Jing Jin, Justin Dauwels, Sydney Cash, and M Brandon Westover. "A fast machine learning approach to facilitate the detection of interictal epileptiform discharges in the scalp electroencephalogram." In: *Journal of Neuroscience Methods* 326 (2019), p. 108362.

[55] Malek Adjouadi, Danmary Sanchez, Mercedes Cabrerizo, Melvin Ayala, Prasanna Jayakar, Ilker Yaylali, and Armando Barreto. "Interictal spike detection using the Walsh transform." In: *IEEE Transactions on Biomedical Engineering* 51.5 (2004), pp. 868–872.

[56] Kemal Polat and Salih Güneş. "Classification of epileptiform EEG using a hybrid system based on decision tree classifier and fast Fourier transform." In: *Applied Mathematics and Computation* 187.2 (2007), pp. 1017–1026.

[57] Jin-De Zhu, Chin-Feng Lin, Shun-Hsyung Chang, Jung-Hua Wang, Tsung-Ii Peng, and Yu-Yi Chien. "Analysis of spike waves in epilepsy using Hilbert-Huang transform." In: *Journal of Medical Systems* 39.1 (2015), pp. 1–13.

[58] Sudipta Mukhopadhyay and GC Ray. "A new interpretation of nonlinear energy operator and its efficacy in spike detection." In: *IEEE Transactions on Bomedical Engineering* 45.2 (1998), pp. 180–187.

[59] Qiong Li, Jianbo Gao, Qi Huang, Yuan Wu, and Bo Xu. "Distinguishing Epileptiform Discharges From Normal Electroencephalograms Using Scale-Dependent Lyapunov Exponent." In: *Frontiers in Bioengineering and Biotechnology* 8 (2020), p. 1006.

[60] Bart Vanrumste, Richard D Jones, and Philip J Bones. "Detection of focal epileptiform activity in the EEG: an SVD and dipole model approach." In: *Proceedings of the Second Joint 24th Annual Conference and the Annual Fall Meeting of the Biomedical Engineering Society][Engineering in Medicine and Biology*. Vol. 3. IEEE. 2002, pp. 2031–2032.

[61] Yun Jiang, Wanzhong Chen, Tao Zhang, Mingyang Li, Yang You, and Xiao Zheng. "Developing multi-component dictionary-based sparse representation for automatic detection of epileptic EEG spikes." In: *Biomedical Signal Processing and Control* 60 (2020), p. 101966. ISSN: 1746-8094.

[62] Chunmei Wang, Junzhong Zou, Jian Zhang, Min Wang, and Rubin Wang. "Feature extraction and recognition of epileptiform activity in EEG by combining PCA with ApEn." In: *Cognitive neurodynamics* 4.3 (2010), pp. 233–240.

[63] K Kobayashi, I Merlet, and J Gotman. "Separation of spikes from background by independent component analysis with dipole modeling and comparison to intracranial recording." In: *Clinical Neurophysiology* 112.3 (2001), pp. 405 –413. ISSN: 1388-2457.

[64] K. Kobayashi, C.J. James, T. Nakahori, T. Akiyama, and J. Gotman. "Isolation of epileptiform discharges from unaveraged EEG by independent component analysis." In: *Clinical Neurophysiology* 110.10 (1999), pp. 1755 –1763. ISSN: 1388-2457.

[65] Christian W Hesse and Christopher J James. "Tracking epileptiform activity in the multichannel ictal EEG using spatially constrained independent component analysis." In: *2005 IEEE Engineering in Medicine and Biology 27th Annual Conference*. IEEE. 2006, pp. 2067–2070.

[66] Marzia De Lucia, Juan Fritschy, Peter Dayan, and David S Holder. "A novel method for automated classification of epileptiform activity in the human electroencephalogram-based on independent component analysis." In: *Medical & Biological Engineering & Computing* 46.3 (2008), pp. 263–272.

[67] Russell C Eberhart, Roy W Dobbins, and W Robert S Webber. "CaseNet: a neural network tool for EEG waveform classification." In: *[1989] Proceedings. Second Annual IEEE Symposium on Computer-based Medical Systems*. IEEE. 1989, pp. 60–68.

[68] Tulga Kalayci and Ozcan Ozdamar. "Wavelet preprocessing for automated neural network detection of EEG spikes." In: *IEEE Engineering in Medicine and Biology Magazine* 14.2 (1995), pp. 160–166.

[69] Hyun S Park, Yong H Lee, Doo S Lee, and Sum I Kim. "Detection of epileptiform activity using wavelet and neural network." In: *Proceedings of the 19th Annual International Conference of the IEEE Engineering in Medicine and Biology Society.'Magnificent Milestones and Emerging Opportunities in Medical Engineering'(Cat. No. 97CH36136)*. Vol. 3. IEEE. 1997, pp. 1194–1197.

[70] Hyun S Park, Yong H Lee, Nam G Kim, Doo S Lee, and Sun I Kim. "Detection of epileptiform activities in the EEG using neural network and expert system." In: *MEDINFO'98*. IOS Press, 1998, pp. 1255–1259.

[71] Christopher J James, Richard D Jones, Philip J Bones, and Grant J Carroll. "Detection of epileptiform discharges in the EEG by a hybrid system comprising mimetic, self-organized artificial neural network, and fuzzy logic stages." In: *Clinical Neurophysiology* 110.12 (1999), pp. 2049–2063.

[72] Nurettin Acır. "Automated system for detection of epileptiform patterns in EEG by using a modified RBFN classifier." In: *Expert Systems with Applications* 29.2 (2005), pp. 455–462.

[73] AT Tzallas, PS Karvelis, CD Katsis, DI Fotiadis, S Giannopoulos, and S Konitsiotis. "A method for classification of transient events in EEG recordings: application to epilepsy diagnosis." In: *Methods of Information in Medicine* 45.06 (2006), pp. 610–621.

[74] Fernanda IM Argoud, Fernando M De Azevedo, José Marino Neto, and Eugênio Grillo. "SADE: An effective system for automated detection of epileptiform events in long-term EEG based on context information." In: *Medical and Biological Engineering and Computing* 44.6 (2006), pp. 459–470.

[75] Nurettin Acir, Ibrahim Oztura, Mehmet Kuntalp, Baris Baklan, and Cuneyt Guzelis. "Automatic detection of epileptiform events in EEG by a three-stage procedure based on artificial neural networks." In: *IEEE Transactions on Biomedical Engineering* 52.1 (2004), pp. 30–40.

[76] Ö Özdamar and T Kalayci. "Detection of spikes with artificial neural networks using raw EEG." In: *Computers and Biomedical Research* 31.2 (1998), pp. 122–142.

[77] Howard J Carey, Milos Manic, and Paul Arsenovic. "Epileptic spike detection with EEG using artificial neural networks." In: *2016 9th International Conference on Human System Interactions (HSI)*. IEEE. 2016, pp. 89–95.

[78] Kosuke Fukumori, Noboru Yoshida, Hidenori Sugano, Madoka Nakajima, and Toshihisa Tanaka. "Epileptic spike detection using neural networks with linear-phase convolutions." In: *IEEE Journal of Biomedical and Health Informatics* 26.3 (2021), pp. 1045–1056.

[79] Mehmet Akif Ozdemir, Ozlem Karabiber Cura, and Aydin Akan. "Epileptic eeg classification by using time-frequency images for deep learning." In: *International Journal of Neural Systems* 31.08 (2021), p. 2150026.

[80] AV Medvedev, GI Agoureeva, and AM Murro. "A long short-term memory neural network for the detection of epileptiform spikes and high frequency oscillations." In: *Scientific Reports* 9.1 (2019), pp. 1–10.

[81] Kosuke Fukumori, Hoang Thien Thu Nguyen, Noboru Yoshida, and Toshihisa Tanaka. "Fully data-driven convolutional filters with deep learning models for epileptic spike detection." In: *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE. 2019, pp. 2772–2776.

[82] Marleen C Tjepkema-Cloostermans, Rafael CV de Carvalho, and Michel JAM van Putten. "Deep learning for detection of focal epileptiform discharges from scalp EEG recordings." In: *Clinical Neurophysiology* 129.10 (2018), pp. 2191–2196.

[83] Loukianos Spyrou and Saeid Sanei. "Coupled dictionary learning for multimodal data: An application to concurrent intracranial and scalp EEG." In: *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE. 2016, pp. 2349–2353.

[84] Jonathan J Halford, Robert J Schalkoff, Jing Zhou, Selim R Benbadis, William O Tatum, Robert P Turner, Saurabh R Sinha, Nathan B Fountain, Amir Arain, Paul B Pritchard, et al. "Standardized database development for EEG epileptiform transient detection: EEGnet scoring system and machine learning analysis." In: *Journal of Neuroscience Methods* 212.2 (2013), pp. 308–316.

[85] Franz Fürbass, Mustafa Aykut Kural, Gerhard Gritsch, Manfred Hartmann, Tilmann Kluge, and Sándor Beniczky. "An artificial intelligence-based EEG algorithm for detection of epileptiform EEG discharges: Validation against the diagnostic gold standard." In: *Clinical Neurophysiology* 131.6 (2020), pp. 1174–1179.

[86] David Geng, Ayham Alkhachroum, Manuel A Melo Bicchi, Jonathan R Jagid, Iahn Cajigas, and Zhe Sage Chen. "Deep learning for robust detection of interictal epileptiform discharges." In: *Journal of Neural Engineering* 18.5 (2021), p. 056015.

[87] C Cheong Took, S Alty, D Martin-Lopez, A Valentin, G Alarcon, and S Sanei. "Transfer Learning of EEG for Analysis of Interictal Epileptiform Discharges." In: *2021 International Conference on e-Health and Bioengineering (EHB)*. IEEE. 2021, pp. 1–4.

[88] Anh Huy Phan and Andrzej Cichocki. "Tensor decompositions for feature extraction and classification of high dimensional datasets." In: *Nonlinear Theory and its Applications, IEICE* 1.1 (2010), pp. 37–68.

[89] Samaneh Kouchaki, Saeid Sanei, Emma L Arbon, and Derk-Jan Dijk. "Tensor based singular spectrum analysis for automatic scoring of sleep EEG." In: *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 23.1 (2014), pp. 1–9.

[90] Frank L Hitchcock. "The expression of a tensor or a polyadic as a sum of products." In: *Journal of Mathematics and Physics* 6.1-4 (1927), pp. 164–189.

[91] Frank L Hitchcock. "Multiple invariants and generalized rank of a p-way matrix or tensor." In: *Journal of Mathematics and Physics* 7.1-4 (1928), pp. 39–79.

[92] Raymond B Cattell. ""Parallel proportional profiles" and other principles for determining the choice of factors by rotation." In: *Psychometrika* 9.4 (1944), pp. 267–283.

[93] Raymond B Cattell. "The three basic factor-analytic research designs—their interrelations and derivatives." In: *Psychological Bulletin* 49.5 (1952), p. 499.

[94] Ledyard R Tucker. "Implications of factor analysis of three-way matrices for measurement of change." In: *Problems in Measuring Change* 15.122-137 (1963), p. 3.

[95] Ledyard R Tucker et al. "The extension of factor analysis to three-dimensional matrices." In: *Contributions to Mathematical Psychology* 110119 (1964).

[96] Ledyard R Tucker. "Some mathematical notes on three-mode factor analysis." In: *Psychometrika* 31.3 (1966), pp. 279–311.

[97] Richard A Harshman. "Foundations of the PARAFAC procedure: Models and conditions for an" explanatory" multimodal factor analysis." In: (1970).

[98] J Douglas Carroll and Jih-Jie Chang. "Analysis of individual differences in multidimensional scaling via an N-way generalization of "Eckart-Young" decomposition." In: *Psychometrika* 35.3 (1970), pp. 283–319.

[99] René Henrion. "Body diagonalization of core matrices in three-way principal components analysis: Theoretical bounds and simulation." In: *Journal of Chemometrics* 7.6 (1993), pp. 477–494.

[100] Rasmus Bro. "PARAFAC. Tutorial and applications." In: *Chemometrics and Intelligent Laboratory Systems* 38.2 (1997), pp. 149–171.

[101] Age K Smilde, Yongdong Wang, and Bruce R Kowalski. "Theory of medium-rank second-order calibration with restricted-Tucker models." In: *Journal of Chemometrics* 8.1 (1994), pp. 21–36.

[102] Victoria Hore, Ana Vinuela, Alfonso Buil, Julian Knight, Mark I McCarthy, Kerrin Small, and Jonathan Marchini. "Tensor decomposition for multiple-tissue gene expression experiments." In: *Nature Genetics* 48.9 (2016), pp. 1094–1100.

[103] Yh Taguchi and Turki Turki. "Tensor decomposition-based unsupervised feature extraction applied to single-cell gene expression analysis." In: *Frontiers in Genetics* 10 (2019), p. 864.

[104] Zhenjun Tang, Lv Chen, Xianquan Zhang, and Shichao Zhang. "Robust image hashing with tensor decomposition." In: *IEEE Transactions on Knowledge and Data Engineering* 31.3 (2018), pp. 549–560.

[105] Ning Hao, Misha E Kilmer, Karen Braman, and Randy C Hoover. "Facial recognition using tensor-tensor decompositions." In: *SIAM Journal on Imaging Sciences* 6.1 (2013), pp. 437–463.

[106] Yissel Rodriguez Aldana, Borbala Hunyadi, Enrique J Marañón Reyes, Valia Rodriguez Rodriguez, and Sabine Van Huffel. "Nonconvulsive epileptic seizure detection in scalp EEG using multiway data analysis." In: *IEEE journal of biomedical and health informatics* 23.2 (2018), pp. 660–671.

[107] Shasha Yuan, Jinxing Liu, Junliang Shang, Fangzhou Xu, Lingyun Dai, and Xiangzhen Kong. "Automatic Seizure Prediction based on Modified Stockwell Transform and Tensor Decomposition." In: *2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. IEEE. 2020, pp. 1503–1509.

[108] Tamara G Kolda and Brett W Bader. "Tensor decompositions and applications." In: *SIAM Review* 51.3 (2009), pp. 455–500.

[109] Tamara Gibson Kolda. *Multilinear operators for higher-order decompositions.* Tech. rep. Sandia National Laboratories, 2006.

[110] Johan Håstad. "Tensor rank is NP-complete." In: *Journal of Algorithms* 11.4 (1990), pp. 644–654.

[111] Age K Smilde, Paul Geladi, and Rasmus Bro. *Multi-way analysis: applications in the chemical sciences*. John Wiley & Sons, 2005.

[112] Lieven De Lathauwer, Bart De Moor, and Joos Vandewalle. "On the best rank-1 and rank-(r 1, r 2,..., rn) approximation of higher-order tensors." In: *SIAM Journal on Matrix Analysis and Applications* 21.4 (2000), pp. 1324–1342.

[113] Stephan Rabanser, Oleksandr Shchur, and Stephan Günnemann. "Introduction to tensor decompositions and their applications in machine learning." In: *arXiv preprint arXiv:1711.10781* (2017).

[114] Evrim Acar, Daniel M Dunlavy, and Tamara G Kolda. "A scalable optimization approach for fitting canonical tensor decompositions." In: *Journal of Chemometrics* 25.2 (2011), pp. 67–86.

[115] Evrim Acar, Daniel M Dunlavy, Tamara G Kolda, and Morten Mørup. "Scalable tensor factorizations for incomplete data." In: *Chemometrics and Intelligent Laboratory Systems* 106.1 (2011), pp. 41–56.

[116] Simon Haykin. *Neural networks and learning machines.* New York: Prentice Hall, 2009.

[117] Oliver Faust, U Rajendra Acharya, Hojjat Adeli, and Amir Adeli. "Wavelet-based EEG processing for computer-aided seizure detection and epilepsy diagnosis." In: *Seizure* 26 (2015), pp. 56–64.

[118] Hojjat Adeli, Ziqin Zhou, and Nahid Dadmehr. "Analysis of EEG records in an epileptic patient using wavelet transform." In: *Journal of Neuroscience Methods* 123.1 (2003), pp. 69–87.

[119] U Rajendra Acharya, Yuki Hagiwara, and Hojjat Adeli. "Automated seizure prediction." In: *Epilepsy & Behavior* 88 (2018), pp. 251–261.

[120] Piotr W Mirowski, Yann LeCun, Deepak Madhavan, and Ruben Kuzniecky. "Comparing SVM and convolutional networks for epileptic seizure prediction from intracranial EEG." In: *2008 IEEE Workshop on Machine Learning for Signal Processing*. IEEE. 2008, pp. 244–249.

[121] J Ross Quinlan. "Bagging, boosting, and C4. 5." In: *AAAI/IAAI, Vol. 1*. 1996, pp. 725–730.

[122] Guoxu Zhou, Andrzej Cichocki, Yu Zhang, and Danilo P Mandic. "Group component analysis for multiblock data: Common and individual feature extraction." In: *IEEE Transactions on Neural Networks and Learning Systems* 27.11 (2015), pp. 2426–2439.

[123] Andrzej Cichocki and Rafal Zdunek. "Regularized alternating least squares algorithms for non-negative matrix/tensor factorization." In: *International Symposium on Neural Networks*. Springer. 2007, pp. 793–802.

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

[125] Shasha Yuan, Weidong Zhou, Qi Wu, and Yanli Zhang. "Epileptic seizure detection with log-euclidean gaussian kernel-based sparse representation." In: *International journal of Neural Systems* 26.03 (2016), p. 1650011.

[126] Geoff Davis, Stephane Mallat, and Marco Avellaneda. "Adaptive greedy approximations." In: *Constructive Approximation* 13.1 (1997), pp. 57–98.

[127] Ron Rubinstein, Michael Zibulevsky, and Michael Elad. "Efficient implementation of the K-SVD algorithm and the Batch-OMP method." In: *Department of Computer Science, Technion, Israel, Tech. Rep* (2008).

[128] Scott Shaobing Chen, David L Donoho, and Michael A Saunders. "Atomic decomposition by basis pursuit." In: *SIAM Review* 43.1 (2001), pp. 129–159.

[129] David L Donoho and Michael Elad. "Optimally sparse representation in general (nonorthogonal) dictionaries via l1 minimization." In: *Proceedings of the National Academy of Sciences* 100.5 (2003), pp. 2197–2202.

[130] Michal Aharon, Michael Elad, and Alfred Bruckstein. "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation." In: *IEEE Transactions on Signal Processing* 54.11 (2006), pp. 4311–4322.

[131] Michael Elad. *Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing*. (pp. 79-109), Springer, 2010.

[132] Wady A Ríos-Herrera, Paola V Olguín-Rodríguez, J Daniel Arzate-Mena, Maria Corsi-Cabrera, Joaquín Escalona, Arlex Marín-García, Julieta Ramos-Loyo, Ana Leonor Rivera, Daniel Rivera-López, José F Zapata-Berruecos, et al. "The influence of EEG references on the analysis of spatio-temporal interrelation patterns." In: *Frontiers in Neuroscience* 13 (2019), p. 941.

[133] Komalpreet Kaur, Jerry J Shih, and Dean J Krusienski. "Empirical models of scalp-EEG responses using non-concurrent intracranial responses." In: *Journal of Neural Engineering* 11.3 (2014), p. 035012.

[134] Ives Macêdo, Emilio Vital Brazil, and Luiz Velho. "Expression transfer between photographs through multilinear AAM's." In: *19th Brazilian Symposium on Computer Graphics and Image Processing*. IEEE. 2006, pp. 239–246.

[135] Daniel Vlasic, Matthew Brand, Hanspeter Pfister, and Jovan Popovic. "Face transfer with multilinear models." In: *ACM SIGGRAPH 2006 Courses*. 2006, 24–es.

[136] Hongcheng Wang and Ahuja. "Facial expression decomposition." In: *Proceedings Ninth IEEE International Conference on Computer Vision*. 2003, 958–965 vol.2. DOI: 10.1109/ICCV.2003.1238452.

[137] Hanchuan Peng, Fuhui Long, and Chris Ding. "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy." In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 27.8 (2005), pp. 1226–1238.

[138] Bahman Abdi-Sargezeh, Antonio Valentin, Gonzalo Alarcon, and Saeid Sanei. "Incorporating Uncertainty in Data Labeling into Automatic Detection of Interictal Epileptiform Discharges from Concurrent Scalp-EEG via Multi-way Analysis." In: *International Journal of Neural Systems* (2021), p. 2150019. ISSN: 0129-0657. URL: https://doi.org/10.1142/S0129065721500192.

[139] Delaram Jarchi, Saeid Sanei, and Ales Prochazka. "Detection of sleep apnea/hypopnea events using synchrosqueezed wavelet transform." In: *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE. 2019, pp. 1199–1203.

[140] Bahman Abdi-Sargezeh, Reza Foodeh, Vahid Shalchyan, and Mohammad Reza Daliri. "EEG Artifact Rejection by Extracting Spatial and Spatio-Spectral Common Components." In: *Journal of Neuroscience Methods* (2021), p. 109182.

[141] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. "Generative adversarial nets." In: *Advances in Neural Information Processing Systems* 27 (2014).

[142] Diederik P Kingma and Max Welling. "Auto-encoding variational bayes." In: *arXiv preprint arXiv:1312.6114* (2013).

[143] Xudong Mao, Qing Li, Haoran Xie, Raymond YK Lau, Zhen Wang, and Stephen Paul Smolley. "Least squares generative adversarial networks." In: *Proceedings of the IEEE International Conference on Computer Vision*. 2017, pp. 2794–2802.

[144] Bowen Li, Xiaojuan Qi, Thomas Lukasiewicz, and Philip Torr. "Controllable text-to-image generation." In: *Advances in Neural Information Processing Systems* 32 (2019).

[145] Jiguo Li, Xinfeng Zhang, Chuanmin Jia, Jizheng Xu, Li Zhang, Yue Wang, Siwei Ma, and Wen Gao. "Direct speech-to-image translation." In: *IEEE Journal of Selected Topics in Signal Processing* 14.3 (2020), pp. 517–529.

[146] Santiago Pascual, Antonio Bonafonte, and Joan Serra. "SEGAN: Speech enhancement generative adversarial network." In: *arXiv preprint arXiv:1703.09452* (2017).

[147] Taesung Park, Ming-Yu Liu, Ting-Chun Wang, and Jun-Yan Zhu. "Semantic image synthesis with spatially-adaptive normalization." In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2019, pp. 2337–2346.

[148] Ting-Chun Wang, Ming-Yu Liu, Jun-Yan Zhu, Andrew Tao, Jan Kautz, and Bryan Catanzaro. "High-resolution image synthesis and semantic manipulation with conditional gans." In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2018, pp. 8798–8807.

[149] Tingting Qiao, Jing Zhang, Duanqing Xu, and Dacheng Tao. "Mirrorgan: Learning text-to-image generation by redescription." In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2019, pp. 1505–1514.

[150] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. "Image-to-image translation with conditional adversarial networks." In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2017, pp. 1125–1134.

[151] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. "Unpaired image-to-image translation using cycle-consistent adversarial networks." In: *Proceedings of the IEEE International Conference on Computer Vision*. 2017, pp. 2223–2232.

[152] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. "U-net: Convolutional networks for biomedical image segmentation." In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer. 2015, pp. 234–241.

[153] Sepp Hochreiter and Jürgen Schmidhuber. "Long short-term memory." In: *Neural Computation* 9.8 (1997), pp. 1735–1780.

[154] Bo Zhao, Lili Meng, Weidong Yin, and Leonid Sigal. "Image generation from layout." In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2019, pp. 8584–8593.

[155] Chin-Cheng Hsu, Hsin-Te Hwang, Yi-Chiao Wu, Yu Tsao, and Hsin-Min Wang. "Voice conversion from unaligned corpora using variational autoencoding wasserstein generative adversarial networks." In: *arXiv preprint arXiv:1704.00849* (2017).

[156] Mina Razghandi, Hao Zhou, Melike Erol-Kantarci, and Damla Turgut. "Variational Autoencoder Generative Adversarial Network for Synthetic Data Generation in Smart Home." In: *arXiv preprint arXiv:2201.07387* (2022).

[157] Anders Boesen Lindbo Larsen, Søren Kaae Sønderby, Hugo Larochelle, and Ole Winther. "Autoencoding beyond pixels using a learned similarity metric." In: *International Conference on Machine Learning*. PMLR. 2016, pp. 1558–1566.

[158] Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky. "Instance normalization: The missing ingredient for fast stylization." In: *arXiv preprint arXiv:1607.08022* (2016).

[159] Vincent Dumoulin, Jonathon Shlens, and Manjunath Kudlur. "A learned representation for artistic style." In: *arXiv preprint arXiv:1610.07629* (2016).

[160] Xun Huang and Serge Belongie. "Arbitrary style transfer in real-time with adaptive instance normalization." In: *Proceedings of the IEEE International Conference on Computer Vision*. 2017, pp. 1501–1510.

[161] Sergey Ioffe and Christian Szegedy. "Batch normalization: Accelerating deep network training by reducing internal covariate shift." In: *International Conference on Machine Learning*. PMLR. 2015, pp. 448–456.

[162] Chuan Li and Michael Wand. "Precomputed real-time texture synthesis with markovian generative adversarial networks." In: *European Conference on Computer Vision*. Springer. 2016, pp. 702–716.

[163] Hao Zhang, Jiteng Yuan, Xin Tian, and Jiayi Ma. "GAN-FM: Infrared and visible image fusion using GAN with full-scale skip connection and dual Markovian discriminators." In: *IEEE Transactions on Computational Imaging* 7 (2021), pp. 1134–1147.

[164] Takeru Miyato, Toshiki Kataoka, Masanori Koyama, and Yuichi Yoshida. "Spectral normalization for generative adversarial networks." In: *arXiv preprint arXiv:1802.05957* (2018).

[165] Han Zhang, Ian Goodfellow, Dimitris Metaxas, and Augustus Odena. "Self-attention generative adversarial networks." In: *International Conference on Machine Learning*. PMLR. 2019, pp. 7354–7363.

[166] Yanyun Qu, Yizi Chen, Jingying Huang, and Yuan Xie. "Enhanced pix2pix dehazing network." In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2019, pp. 8160–8168.

[167] Vernon J Lawhern, Amelia J Solon, Nicholas R Waytowich, Stephen M Gordon, Chou P Hung, and Brent J Lance. "EEGNet: a compact convolutional neural network for EEG-based brain–computer interfaces." In: *Journal of Neural Engineering* 15.5 (2018), p. 056013.

[168] Nicholas Waytowich, Vernon J Lawhern, Javier O Garcia, Jennifer Cummings, Josef Faller, Paul Sajda, and Jean M Vettel. "Compact convolutional neural networks for classification of asynchronous steady-state visual evoked potentials." In: *Journal of Neural Engineering* 15.6 (2018), p. 066031.

[169] François Chollet. "Xception: Deep learning with depthwise separable convolutions." In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2017, pp. 1251–1258.

[170] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. "Attention is all you need." In: *Advances in Neural Information Processing Systems* 30 (2017).

[171] Qiang Wang, Bei Li, Tong Xiao, Jingbo Zhu, Changliang Li, Derek F Wong, and Lidia S Chao. "Learning deep transformer models for machine translation." In: *arXiv preprint arXiv:1906.01787* (2019).