Epilepsy diagnosis from scalp EEG : a machine learning approach

Thomas, John

2019

Thomas, J. (2019). Epilepsy diagnosis from scalp EEG : a machine learning approach.
Doctoral thesis, Nanyang Technological University, Singapore.

<https://hdl.handle.net/10356/93579>

<https://doi.org/10.32657/10356/93579>

*Epilepsy Diagnosis from Scalp EEG:
A Machine Learning Approach*

JOHN THOMAS
SCHOOL OF ELECTRICAL AND ELECTRONIC ENGINEERING
2019

*Epilepsy Diagnosis from Scalp EEG:
A Machine Learning Approach*

JOHN THOMAS

SCHOOL OF ELECTRICAL AND ELECTRONIC ENGINEERING

A THESIS SUBMITTED TO THE NANYANG TECHNOLOGICAL UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENT FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

2019

Statement of Originality

I hereby certify that the work embodied in this thesis is the result of original research, is free of plagiarised materials, and has not been submitted for a higher degree to any other University or Institution.

13 - 09 - 2019

.....

Date

A handwritten signature in black ink, appearing to read 'John Thomas', written over a horizontal dotted line.

John Thomas

Supervisor Declaration Statement

I have reviewed the content and presentation style of this thesis and declare it is free of plagiarism and of sufficient grammatical clarity to be examined. To the best of my knowledge, the research and writing are those of the candidate except as acknowledged in the Author Attribution Statement. I confirm that the investigations were conducted in accord with the ethics policies and integrity standards of Nanyang Technological University and that the research data are presented honestly and without prejudice.

13 - 09 - 2019

.....

Date

.....

Assoc. Prof. Justin Dauwels

Authorship Attribution Statement

This thesis contains material from five papers published in the following conference proceedings where I was an author.

Chapter 4, Section 7.1 is accepted for publication as Yuvaraj R., **Thomas J.**, Kluge T., and Dauwels J. “A Deep Learning Scheme for Automatic Seizure Detection from Long-Term Scalp EEG” in Asilomar Conference on Signals, Systems, and Computers 2018. The contributions of the authors are as follows:

- Assoc. Prof. Justin Dauwels co-designed the project, guided throughout the project, and helped in revising the manuscript.
- Dr. Yuvaraj Rajamanickam co-designed the project, prepared the data, evaluated the system, and prepared the manuscript.
- I co-designed the project, prepared the code for developing the Convolutional Neural Network (CNN)-based system, and helped in revising the manuscript.
- Dr. Tilmann Kluge helped in revising the manuscript.

Chapter 4, Section 7.2 is published as **Thomas J.**, Maszczyk T., Sinha N., Kluge T., and Dauwels J. “Deep learning-based classification for brain-computer interfaces” in Systems, Man, and Cybernetics (SMC), 2017 IEEE International Conference on, pp. 234-239. IEEE, 2017. The contributions of the authors are as follows:

- Assoc. Prof. Justin Dauwels co-designed the project, guided throughout the project, and helped in revising the manuscript drafts.

- I co-designed the project, preprocessed the data, performed the evaluation, and prepared the manuscript.
- Mr. Nishant Sinha helped in initializing the project and preparing the dataset.
- Dr. Tomasz Maszczyk helped in evaluating the traditional classifiers results and revising the manuscript.
- Dr. Tilmann Kluge helped in revising the manuscript.

Chapter 5 and 6 contains content from the following published conference papers:

1. **Thomas J.**, Comoretto L., Jin J., Dauwels J., Cash S.S., and Westover M.B. “EEG Classification via Convolutional Neural Network-Based Interictal Epileptiform Event Detection” in Engineering in Medicine and Biology Society (EMBC), 2018 40th Annual International Conference of the IEEE, pp. 3148-3151, 2018.
2. **Thomas J.**, Jin J., Dauwels J., Cash S.S., and Westover M.B. “Automated epileptiform spike detection via affinity propagation-based template matching” in Engineering in Medicine and Biology Society (EMBC), 2017 39th Annual International Conference of the IEEE, pp. 3057-3060. IEEE, 2017.
3. **Thomas J.**, Jin J., Dauwels J., Cash S.S., and Westover M.B. “Clustering of interictal spikes by dynamic time warping and affinity propagation” in Acoustics, Speech, and Signal Processing (ICASSP), 2016 IEEE International Conference on, pp. 749-753. IEEE, 2016.

The contributions of the authors are as follows:

- Assoc. Prof. Justin Dauwels co-designed the project, guided throughout the project, and helped in revising the manuscript drafts.

- I co-designed the project, preprocessed the data, performed the evaluation, and prepared the manuscripts.
- Dr. Jing Jin helped in preparing the dataset.
- Mr. Luca Comeretto helped in conducting some preliminary evaluation of the system for conference paper 1.
- Dr. Brandon Westover and Dr. Sydney S. Cash helped by providing feedback on the project by revising the manuscripts.

13 - 09 - 2019

.....

Date

A handwritten signature in black ink, appearing to read "John Thomas", written over a horizontal dotted line.

John Thomas

Acknowledgments

First and foremost, I would like to express my sincere gratitude to my supervisor Assoc. Prof. Justin Dauwels, not just for his continuous support of my Ph.D. research, but also for his unending patience, motivation, enthusiasm, and immense knowledge. Without his guidance and persistent help, this research would not have been possible.

I would also like to thank the experts who were involved in data recording, guidance, and validation for this research project: Dr. M. Brandon Westover and Dr. Sydney S. Cash from Massachusetts General Hospital and Harvard Medical School. Their passionate participation and input have been imperative to my research. I also thank my fellow labmates in Dauwels Lab: Dr. Yuvaraj Rajamanickam, Prasanth Thangavel, Elham Bagheri, Banishree Ghosh, Nishant Sinha, Dr. Jing Jin, and Dr. Tomasz Maszczyk for both the stimulating discussions as well as all the fun we have had in the last four years.

My sincere thanks are also due to the clinicians Dr. Rahul Rathakrishnan from National University Hospital (NUH) and Dr. YeeLeng, National Neuroscience Institute (NNI) for the clinical support and data collection for system validation.

Finally, I must express my very profound gratitude to my parents and my friends for providing me with unfailing support and continuous encouragement throughout my years of study. This accomplishment would not have been possible without them.

Table of Contents

ACKNOWLEDGMENTS	i
TABLE OF CONTENTS	iii
ABSTRACT	vii
LIST OF FIGURES	x
LIST OF TABLES	xiv
LIST OF ACRONYMS	xviii
1 INTRODUCTION	1
1.1 Motivation	2
1.2 Objectives	4
1.3 Contributions	4
1.4 Organization of the Thesis	6
2 METHODS AND PERFORMANCE INDICES	7
2.1 Machine learning	8
2.1.1 Distance measures	8
2.1.1.1 Euclidean distance (ED)	9
2.1.1.2 Correlation coefficient	9
2.1.1.3 Dynamic Time Warping (DTW)	9
2.1.1.4 Mahalanobis distance	10
2.1.2 Clustering	11
2.1.2.1 K-means clustering	11
2.1.2.2 K-medoids clustering	12
2.1.2.3 Hierarchical clustering	12
2.1.2.4 Fuzzy C-means clustering (FCM)	13
2.1.2.5 Affinity Propagation (AP)	13
2.1.2.6 Clustering evaluation measures	15

2.2	Traditional classifiers	18
2.2.1	Template Matching (TM)	18
2.2.2	Decision tree	18
2.2.3	k -Nearest Neighbors (k -NN)	19
2.2.4	Linear Discriminant Analysis (LDA)	19
2.2.5	Support Vector Machine (SVM)	19
2.2.6	Ensemble of classifiers	20
2.3	Deep learning	20
2.3.1	Recurrent Neural Networks (RNN)	21
2.3.2	Convolutional Neural network (CNN)	21
2.4	Data visualization with t-Distributed Stochastic Neighbor Embedding (t-SNE)	22
2.5	Performance indices	24
3	BACKGROUND ON EEG AND EPILEPSY	26
3.1	Background	27
3.1.1	Electroencephalogram (EEG)	27
3.1.2	Epilepsy	28
3.1.3	Interictal Epileptiform Discharges (IEDs)	29
3.2	Towards quantitative definition of IEDs	30
3.3	IED detection	33
3.3.1	Mimetic	33
3.3.2	Parametric modeling	35
3.3.3	Traditional classifiers	37
3.3.4	Wavelet analysis	37
3.3.5	Neural Networks (NN)	39
3.3.6	Power Spectral Analysis (PSA)	39
3.3.7	Drawbacks and challenges	40
3.4	Commercial EEG software	46
3.5	Epilepsy diagnosis and management through EEG	47
4	SCALP EEG DATA AND PREPROCESSING TECHNIQUES	50
4.1	Clinical EEG data	51
4.1.1	MGH dataset	51
4.1.2	MUSC dataset	54
4.1.3	NUH dataset	55

4.2	Preprocessing	56
4.2.1	Filtering	56
4.2.2	Down-sampling	56
4.2.3	Power-based artifact detection	57
4.2.4	Removal of missing electrode segments	57
4.2.5	EEG montages	57
4.3	IED and non-IED extraction	59
4.4	Quantitative characteristics of IEDs	60
4.4.1	Introduction	61
4.4.2	Methodology	61
4.4.3	Results	65
4.4.4	Discussion	67
4.5	IED statistics	70
4.5.1	IED variability studies	70
4.5.2	IED frequency analysis	71
5	IED DETECTOR	72
5.1	Introduction	73
5.2	Methodology	75
5.2.1	Data preparation and evaluation	75
5.2.2	Background rejection	78
5.2.3	TM IED detector	79
5.2.4	Traditional classifier-based IED detectors	84
5.2.5	CNN IED detector	86
5.3	Results	90
5.3.1	Background rejection	90
5.3.2	TM IED detector	90
5.3.3	Traditional classifier-based IED detectors	99
5.3.4	CNN IED detector	102
5.3.5	Evaluation on MUSC dataset and Persyst	104
5.3.6	Feature evaluation	105
5.3.7	Computational time	106
5.4	Discussion	107

6	EPILEPTIC EEG DETECTOR	109
6.1	Introduction	110
6.2	Methodology	111
6.3	Results	115
6.3.1	Cross-validation results	115
6.3.2	Clinical validation	125
6.4	Discussion	127
7	EXTENDING THE APPROACH FOR SEIZURE DETECTION AND BRAIN-COMPUTER INTERFACES	129
7.1	Seizure detection	130
7.1.1	Introduction	130
7.1.2	Methodology	130
7.1.2.1	Scalp EEG and preprocessing	130
7.1.2.2	Approach	132
7.1.3	Results and discussion	135
7.1.4	Conclusions	138
7.2	Classification for Brain-Computer Interfaces (BCIs)	139
7.2.1	Introduction	139
7.2.2	Methodology	142
7.2.2.1	BCI Dataset and preprocessing	142
7.2.2.2	Approach	142
7.2.3	Results and discussion	146
7.2.4	Conclusions	147
8	CONCLUSIONS AND FUTURE DIRECTIONS	149
8.1	Conclusions	150
8.2	Future work	153
8.2.1	Artifact rejection techniques	153
8.2.2	Development of an adaptive system	154
8.2.3	Validation on multiple datasets	154
8.2.4	Towards multi-observer systems	154
	PUBLICATIONS	160
	REFERENCES	178

Epilepsy Diagnosis from Scalp EEG: A Machine Learning Approach

ABSTRACT

Epilepsy refers to a group of chronic brain disorders characterized by recurrent unprovoked seizures. Interictal Epileptiform Discharges (IEDs) are distinctive biomarkers of epilepsy. According to the International Federation of Societies for Electroencephalography and Clinical Neurophysiology (IFSECN), IEDs are defined as transients distinguishable from background activity, with a characteristic spiky morphology, typically, but neither exclusively nor invariably, found in interictal electroencephalograms (EEGs) of people with epilepsy. Approximately 65 million people are affected worldwide. The early diagnosis of epilepsy is based on the IEDs that appear in the EEG. The EEGs are non-invasive, inexpensive, digitized recordings of brain activity from the scalp. However, there is a lack of experts who can analyze EEGs; moreover, manual interpretation EEG is time-consuming. Moreover, IEDs exhibit huge variation across patients. Besides, a standard definition of interictal epileptiform events is yet to be formulated. This makes the IED detection based diagnosis of epilepsy, tedious, and expert-centered. Experts have a significant disagreement over IED annotation. Consequently, an au-

tomated objective algorithm for detecting IEDs would be beneficial for the diagnosis of epilepsy.

Our study had two primary objectives: to study the underlying characteristics of epileptiform IEDs and to develop an automated epilepsy detection system. First, we quantified the different morphological characteristics of IEDs, and a standard definition of IEDs was formulated. Later, we developed an intelligent automated epileptic EEG identification system based on interictal pattern detection. The system consists of three major parts: the preprocessing module, the interictal pattern detection module, and the EEG classification module. We have evaluated the system on a sizable routine scalp EEG database from multiple centers, namely, Massachusetts General Hospital (MGH), Medical University of South Carolina (MUSC), and National University Hospital (NUH) Singapore. The MGH dataset consists of 545 subjects: 84 epileptic patients with IEDs annotated by two neurologists and 461 non-epileptic subjects. We applied a preprocessing pipeline with the following components: data preparation, data segmentation, filtering, and artifact removal.

Next we employ a convolutional neural network (CNN)-based pattern detection for identifying epileptiform patterns. The system achieved superior performance compared to conventional methods with the highest area-under-curve (AUC) of 0.989 ± 0.006 and a false alarm rate of 0.2 ± 0.12 per minute for a sensitivity threshold of 80%. The proposed system was proved to be

non-inferior with the expert consensus on the MUSC dataset. Further, we developed an EEG classification system based on the output of the interictal pattern detector. Multiple features were derived from the CNN outputs, and the best features were selected based on p -values. The best features were then employed to train the EEG classifier. The proposed system achieved a cross-validation balanced accuracy of 81.61% on the MGH dataset. Later the system, trained on the MGH dataset was validated on 100 EEGs recorded at NUH. The epileptic EEG identification system achieved an accuracy of 88% for clinical validation on the NUH data.

List of figures

2.1.1 DTW distance measure: (a) Non-linear mapping of data points of two sequences; (b) The n -by- n matrix showing the optimal path and Sakoe-chiba band.	10
2.1.2 Message passing between data points in affinity propagation: (a) sending responsibilities; (b) sending availabilities.	15
2.3.1 Basic structure of an LSTM cell.	22
3.1.1 A typical representation of IEDs; (a) a spike segment; (b), (c) polyspikes, and (d) spike followed by a slow-wave.	31
3.2.1 The decomposition of IEDs into three transients by Celesia et al. The duration of the three transients of the spike and slow-wave complex is represented by a , b , and c	32
3.2.2 Six parameters of IEDs defined by Gotman et al. Amplitude and duration of the first and second half-waves are measured between A–C and C–E, respectively. B and D are the mid-points between A–C and C–E, respectively. CB and CD define the slopes.	33
3.3.1 Mimetic parameters described by James et al.	34
3.3.2 The reference diagram for feature extraction presented by Liu et al.	35
3.3.3 The linear prediction model applied in the study performed by Sankar and Natour (1992).	36
3.3.4 Two-level decomposition of a signal by DWT into approximation and detail coefficients.	39
3.3.5 The design of the feedforward NN applied by Carey et al. . .	40
4.1.1 The electrode placement according to the 10-20 International system on a two dimensional human scalp (top-view).	52
4.1.2 Age statistics for the MGH dataset.	53

4.1.3 A ten-second sample from a 19-channel EEG recordings of a subject with epilepsy. The annotated IED is highlighted in the circle.	53
4.1.4 IED event distribution in the MGH dataset.	54
4.2.1 Different preprocessing steps applied on the EEG recordings.	58
4.3.1 Realignment process of IEDs during data extraction.	60
4.3.2 500-millisecond (64-sample) waveforms of (a) IEDs and (b) non-IEDs extracted randomly from the MGH dataset (93 epileptic and 461 non-epileptic EEGs).	60
4.4.1 Peak detection system applied for extracting the morphological characteristics of IEDs.	62
4.4.2 Morphological characteristics of IEDs (a) Amplitude features: peak voltage, rising voltage, and falling voltage; (b) Duration of the IED transients: duration of first half-wave, second half-wave, and the slow-wave; (c) Slope features: rising slope and falling slope.	64
4.4.3 Distribution density plots of different amplitude features of IEDs: (a) Peak voltage; (b) Rising voltage; (c) Falling voltage. The mean values are highlighted in red.	65
4.4.4 Distribution density plots of the duration of the three IED transients: (a) Duration of the first half-wave; (b) Duration of the second half-wave; (c) Duration of the slow-wave. The mean values are highlighted in red.	66
4.4.5 Distribution density plots of the slope values of IEDs: (a) Rising slope values; (b) Falling slope values. The mean values are highlighted in red.	66
4.5.1 The mean correlation value heatmap for intra-patient and inter-patient IED variability.	70
4.5.2 The ratio of IEDs vs. non-IEDs for different patients.	71
5.2.1 Patient age distribution over the different folds. The mean is indicated in red.	76
5.2.2 Block diagram of the proposed IED detector training scheme.	77
5.2.3 Block diagram of the proposed IED detector evaluation scheme.	77
5.2.4 The flow diagram of the proposed method for implemented an efficient TM IED detector.	80

5.2.5 Different steps involved in the selection of a clustering technique.	82
5.2.6 Different steps involved in the selection of a distance measure.	83
5.2.7 Basic structure of the proposed template matching (TM) system. Background templates are integrated to improve the detection accuracy.	83
5.2.8 The proposed CNN architecture with a single convolutional, pooling, and fully connected layer applied in this study.	87
5.2.9 The proposed training and validation steps for trial 1 of the 5-fold cross-validation.	89
5.3.1 T-SNE visualization of IEDs and non-IED waveforms : (a) Density plots of IEDs for ED; (b) Density plots of non-IEDs for ED; (c) Density plots of IEDs for DTW; (d) Density plots of non-IEDs for DTW; (e) Density plots of IEDs for correlation coefficient; (f) Density plots of non-IEDs for correlation coefficient; (g) Density plots of IEDs for Mahalanobis distance; (h) Density plots of non-IEDs for Mahalanobis distance.	94
5.3.2 Density plots of minimum correlation distance values from template library for IEDs and backgrounds based on rule 1.	95
5.3.3 Density plots of minimum correlation distance values from template library for IEDs and backgrounds based on rule 2.	96
5.3.4 Performance curves of the TM system with decision rule 1 (blue) and decision rule 2 (red): (a) PRC; (b) sensitivity and false positive rate curves; (c) ROC.	98
5.3.5 Performance curves of the traditional classifier-based (SVMG in blue and ensemble CART in red) IED detector: (a) PRC; (b) sensitivity and false positive rate curves; (c) ROC.	100
5.3.6 Variation in training and validation indicators (accuracy and loss) with the increase in the number of iterations of CNN training: (a) accuracy; (b) loss. The training termination point is highlighted.	102
5.3.7 Performance curves of the CNN IED detector: (a) PRC; (b) sensitivity and false positive rate curves; (c) ROC.	103

5.3.8 Precision-recall comparison plots of our proposed CNN system with Persyst and traditional classifier systems on the on the MUSC dataset: (a) average annotator consensus of 3; (b) average consensus of 4.	104
5.3.9 Comparison of the proposed CNN IED detector and traditional classifier-based IED detectors based on the precision-sensitivity curve on the MUSC dataset with mean annotator consensus threshold of 4. The performance of experts is indicated as circles (green).	105
5.3.10 Correlation values of the different IED characteristics with the CNN output values.	106
6.2.1 Block diagram of the proposed EEG classification system.	111
6.2.2 Block diagrams for the different EEG classifier systems: (a) Model 1; (b) Model 2, and (c) Model 3.	113
6.2.3 Development of the EEG classifier based on the MGH dataset.	114
6.3.1 Boxplots for the two different features with a CNN threshold of 0.9 for epileptic and non-epileptic EEG: (a) CNN detections; (b) CNN detections per second.	115
6.3.2 Barplots for the two different features with a CNN threshold of 0.9 for epileptic and non-epileptic EEG: (a) CNN detections; (b) CNN detections per second.	116
6.3.3 Performance curves for Model 1 EEG classifier based on CNN detections: (a) ROC; (b) PRC.	120
6.3.4 Performance curves for Model 1 EEG classifier based on CNN detection per second: (a) ROC; (b) PRC.	120
6.3.5 Scatter plot of the two most salient features, i.e., the fraction of CNN outputs between the intervals (0.8-0.9] and (0.9-1] for EEG-level classification based on Model 2. The plot shows that the two classes are reasonably separable.	121
6.3.6 Performance curves for Model 2 EEG classification based on feature set 2: (a) ROC; (b) PRC.	121
6.3.7 Scatter plot of the CNN detections per second and the mean correlation distance from : (a) IED templates; (b) background templates.	123

6.3.8 Performance curves for Model 3 EEG classification based on feature set 3: (a) ROC; (b) PRC.	123
6.3.9 Boxplots of CNN detections per second for MGH and NUH datasets. The epileptic and non-epileptic EEGs are separable for both the datasets, but with different thresholds.	126
6.4.1 Scatter plots for subject age vs. the output of the EEG clas- sification system.	128
6.4.2 Scatter plots for subject gender vs. the output of the EEG classification system.	128
7.1.1 Workflow of the proposed CNN based seizure detection sys- tem implementation.	131
7.1.2 Structure of the implemented CNN network.	134
7.1.3 The sensitivity-precision curves for each fold.	137
7.2.1 The basic model of a BCI study. The training part is shown in red.	140
7.2.2 The block structure of the implemented CNN network. . . .	142
7.2.3 The model of the implemented bidirectional LSTM network. . .	144
8.2.1 Block diagram of the multi-consensus IED detector system. .	157
8.2.2 Block diagram of the proposed multi-observer IED detector system.	157
8.2.3 Boxplots for the RMSE values for the different approaches. .	159

List of tables

3.1.1 Frequency bands present in human EEG.	27
3.2.1 Comparison of IED amplitudes with a mean peak-to-peak amplitude of the 500 ms background waveform immediately preceding an IED. Adapted from the study by Celesia et al.	32
3.3.1 Previous IED detector studies in literature.	41
3.3.2 Previous IED detector studies in literature.	42
3.3.3 Previous IED detector studies in literature.	43
3.3.4 Previous IED detector studies in literature.	44
4.1.1 Annotated IED event distribution of the MGH dataset. . . .	52
4.4.1 Morphological characteristics of IEDs (reported as mean $\pm$ standard deviation).	67
4.4.2 Morphological characteristics of IEDs separated into spikes and sharp waves (reported as mean $\pm$ standard deviation). .	67
4.4.3 Comparison of the amplitude of IED (peak voltage) with the 500 milliseconds background activity (b.a.) preceding the annotated IED.	69
4.4.4 Percentage of IEDs with annotated IEDs in the neighborhood that exhibit an amplitude difference 10% or lesser with the neighborhood waveforms.	69
5.2.1 Distribution of the MGH dataset over different cross-validation folds.	75
5.2.2 Parameter values of the CNN network implemented for back- ground rejection.	79
5.2.3 Decision rules for the different template matching systems. .	84
5.2.4 Different parameter values optimized for the traditional clas- sifiers.	85
5.2.5 Different sets of parameter values evaluated for CNN opti- mization.	88

5.3.1 Results for performing background rejection with a threshold of 0.99.	90
5.3.2 Results of clustering by employing the four validation index values.	91
5.3.3 AUC results for the TM IED systems with different distance measures and different decision rules.	96
5.3.4 5-fold cross-validation results for TM IED detector system with correlation distance as the distance measure.	97
5.3.5 AUC and AUPRC results for the various traditional classifier-based IED detectors.	99
5.3.6 5-fold cross-validation results of traditional classifier systems.	101
5.3.7 5-fold cross-validation results of CNN IED detector system.	102
6.3.1 Results for EEG-level classification based on the waveform-level classification features for Model 1 and based on CNN detections.	118
6.3.2 Results for EEG-level classification based on the waveform-level classification features for Model 1 and based on CNN detections per second.	119
6.3.3 The results for EEG-level classification based on the waveform-level classification features for Model 2.	122
6.3.4 The results for EEG-level classification based on the waveform-level classification features for Model 3.	124
6.3.5 The results for the different models on the NUH data.	126
7.1.1 Details on the CHB-MIT scalp EEG dataset. Gender:female (Fe), male (Ma). Others: not available (NA).	132
7.1.2 Distribution of seizure segments in each folds	133
7.1.3 The different parameters evaluated for CNN implementation	135
7.1.4 Information regarding the dimension of our data structures.	136
7.1.5 CNN seizure detection performance	137
7.2.1 The different parameter values evaluated for traditional classifiers	143
7.2.2 The different sets of parameter values evaluated for CNN optimization	143
7.2.3 Optimized parameters of the CNN network	144

7.2.4 Optimized parameters of the LSTM network	144
7.2.5 LOSO-CV classification accuracies of different classifiers on the SSVEP dataset	148
8.2.1 The accuracy of the different annotators for distinguishing IEDs vs. non-IED waveforms.	156
8.2.2 Consistency score of the different annotators.	156

List of Acronyms

MGH	Massachusetts General Hospital
EEG	Electroencephalogram
IED	Inter-ictal Epileptiform Discharge
CNN	Convolutional Neural Network
ICA	Independent Component Analysis
WT	Wavelet Transform
ES	Expert System
WaT	Walsh Transform
FT	Fourier Transform
HT	Hilbert Transform
NN	Neural Network
ANN	Artificial Neural Network
t-SNE	t-Distributed Stochastic Neighbor Embedding
TM	Template Matching
CWT	Continuous Wavelet Transform
DWT	Discrete Wavelet Transform
DTW	Dynamic Time Warping
AP	Affinity Propagation
DB	Davies-Bouldin
PSD	Power Spectral Density
ED	Euclidean Distance

“Discovery consists of looking at the same thing as everyone else and thinking something different.”

Albert Szent-Gyorgyi

1

Introduction

Epilepsy is one of the prevalent brain disorders in the world. We aim to develop an epilepsy diagnostic system to identify epileptic electroencephalograms (EEGs) expeditiously. The first chapter outlines the premise of our research and the proposed objectives for the development of the epilepsy diagnostic system based on interictal epileptiform pattern recognition. We highlight the motivation of our study in Section 1.1, objectives in Section 1.2, and the achievements of this research in Section 1.3. Further, the structural organization of the thesis is elaborated in Section 1.4.

1.1 MOTIVATION

Epilepsy is a group of chronic brain disorders that are characterized by unpredictable seizures. It is rated as the fourth most common neurological disorder by the Epilepsy Foundation [1]. Approximately 65 million people [2] are affected worldwide. A recent statistics from the USA states that one in every twenty-six is likely to develop epilepsy during their lifetime [3]. Epilepsy and seizures affect all age groups but are common in children and elderly people. The significant causes of epilepsy include brain malformations, tumors, infections in infants, head trauma, and congenital conditions in adults, stroke and Alzheimer’s disease in elderly people. Seizures are unprovoked electrical surges in the human brain that inhibit the normal functioning of the brain. The health risks associated with epilepsy include sleeping problems, osteoporosis, reproductive problems, anxiety, and depression [1]. Electroencephalogram (EEG) is the recording of the electrical signals of the brain acquired by scalp electrodes. Interictal Epileptiform Discharges (IEDs) that appear in the EEG are distinctive biomarkers of epilepsy. IEDs are further categorized into sharp waves, spikes, spike-wave-complexes, and polyspike-wave-complexes [4]. We primarily deal with spikes and spike-wave-complexes in this research.

An important method of epilepsy diagnosis is based on the detection and evaluation of IEDs. Traditionally, the IEDs are visually identified in the EEG recordings by neurologists. This process is tedious and time-consuming. Moreover, the experts are subjective; there is a significant disagreement over EEG interpretation amongst experts [5]. The IEDs show considerable variations in EEG morphology across patients. Besides, a standard or absolute reference on the definition of IEDs is not available [6]. Consequently, there is no uniformity in interpreting the epileptiform events by experts. The reliability of the diagnosis heavily depends on the experience and expertise of the neurologists. An automated IED detection system will increase the uniformity and reliability of EEG interpretation for epilepsy. Several methods and algorithms have been evaluated in the literature to provide a reliable automated IED detection system. The common deficiency of the existing methods is the lack of a sizable IED database to train and validate the IED detection system. The lack of a sufficient number of EEG interpreters is

also a significant concern. Moreover, most methods are offline evaluation systems that prolong the diagnostic process. The complexity of the current detection methods is another significant challenge for clinical deployment.

This research is motivated by the need for an automated and efficient epilepsy diagnostic system. An efficient automated IED detection software could improve the early diagnosis of epilepsy. In order to have a reliable detection system, it is mandatory to study a sizable database of IEDs. We study an IED database extracted from the EEG data of 84 epileptic patients recorded at Massachusetts General Hospital (MGH). The database consists of 18,164 IEDs annotated by two neurologists independently. We also evaluate the systems on data from multiple centers, namely, National University Hospital (NUH) Singapore, and Medical University of South Carolina (MUSC).

Today, fast and efficient pattern detection systems are available that could solve the challenge of processing delays in the diagnostic process. An optimized epileptic EEG identification system based on the interictal pattern recognition is the proposed solution. During the first phase, we focused on EEG data extraction, clinical record translation, and studying the different characteristics of the interictal patterns. A standard quantitative definition for IED is not available in the literature. Therefore, we quantified the different characteristics of the IEDs by analyzing the sizable dataset of IEDs from MGH. In the second phase, we developed a novel Convolutional Neural Network (CNN)-based IED detector that outperformed the traditional systems as well as Persyst. Persyst is a leading EEG software firm which provides epilepsy detection software solutions. Further, we also developed a template matching (TM) system for IED detection. The efficiency of the TM system primarily relies on the effectiveness of the template library. Therefore, to extract an effective IED template library, we applied clustering on the MGH IED database to extract a library of IED templates. In order to optimize this process, we also employed a distance measure along with fine-tuning of the clustering parameters. The template matching system is developed with the extracted template library. Subsequently, we developed an EEG classifier based on the features extracted from the IED detector output. The entire EEG classification system was clinically validated on 100 subjects at NUH Singapore with a prediction accuracy of 88%.

1.2 OBJECTIVES

The objectives of this research are:

- To study the different characteristics of IEDs from the MGH IED database (18,164 IEDs) and provide a quantitative definition of IEDs based on morphological characteristics.
- To develop an efficient template matching IED detector based on a set of high-quality template library.
- To develop an efficient deep learning-based IED detector.
- To develop an automated, efficient EEG classification system based on IED detection.
- To broaden the scope of the system to more applications such as seizure detection and Brain-Computer Interface (BCI) classification.

1.3 CONTRIBUTIONS

The contributions to the research towards an epileptic EEG identification system are:

- **Development of a TM IED detector:** We evaluated different clustering techniques on the MGH IED database and identified the most effective technique based on the clustering validation indices. To enhance the quality of template extraction process, we selected an appropriate distance measure. Next, the parameters of the clustering algorithm were optimized by fine-tuning. Subsequently, we developed the template library by extracting representatives from the IED clusters. We optimized the TM IED detector by incorporating non-IED templates and modifying the decision rules. The performance is comparable to the methods detailed in the literature. However, we have tested the TM IED detector on a database that is significantly larger than those evaluated in the literature. We achieved a mean cross-validation false detection rate of 3.95 ± 2.03 per minute for sensitivity thresholds of 80%.

- **Development of a CNN-based IED detector:** We developed a CNN-based IED detector and achieved mean cross-validation false positive rate of 0.2 ± 0.12 for a sensitivity threshold of 80%. The detector was evaluated on a database of annotated IEDs from MUSC in comparison with the existing market leader in epileptic EEG software, Persyst. The proposed system out-performed Persyst as well as other traditional systems, significantly.
- **Development of an EEG-classification system and Clinical validation:** An EEG classification system was developed based on the IED detector outputs. We extracted different features from the IED detector and selected the best mutually independent set of features. This set was further employed to train the EEG classifier. The EEG classifier system achieved a mean cross-validation accuracy of 81.61% on the MGH dataset. Further, the system trained on the MGH dataset was employed to evaluate the data recorded at NUH Singapore. The proposed system performed with an accuracy of 88% on the NUH data.
- **Extending the system for Seizure Detection and Classification of Brain-Computer Interfaces (BCIs):** We applied the CNN-based approach for seizure detection and for classifying BCI tasks based on EEG recordings. We developed a generalized CNN seizure detection model and achieved a 4-fold cross-validation average sensitivity of 86.29% and false detection rate of 0.74 per hour for an average detection latency of 2.1 sec. For classifying BCI tasks, we achieved superior Leave-One-Subject-Out cross-validation (LOSO-CV) average accuracy of 69.03% for distinguishing stimuli of non-overlapping flickering lights with frequencies 6.66 Hz, 7.5 Hz, 8.5 Hz, 10 Hz, and 12 Hz by analyzing the SSVEP signals. The comparison was made with the different traditional classifiers in combination with sequential forward selection (SFS) [7] of features.

1.4 ORGANIZATION OF THE THESIS

The report is organized as follows:

- Chapter 2 details the different statistical and machine learning techniques investigated in this research.
- Chapter 3 describes the background regarding epilepsy, epileptiform events, electroencephalogram, and the clinical requirements of the epilepsy diagnostic system.
- Chapter 4 highlights the different scalp EEG datasets, preprocessing techniques, and data extraction techniques investigated in our research.
- Chapter 5 describes the development and implementation of the proposed IED detectors (traditional classifier-based and CNN-based).
- Chapter 6 contains detailed information regarding the different feature extraction, feature selection, and classification techniques applied for the development of the epileptic EEG classification system.
- Chapter 7 highlights the extensions of the systems into other domains of seizure detection and BCIs.
- Chapter 8 concludes the thesis with potential ideas for future work.

“You don’t need something more to get something more. That’s what emergence means.”

Murray Gell-Mann

2

Methods and Performance Indices

We have applied different statistical and machine learning techniques for data preparation, extraction, analysis, and implementation of the epilepsy diagnostic system. In this chapter, we present the methodology of different methods and materials applied to extract knowledge, visualize, and analyze data throughout this thesis. We have outlined basic machine learning techniques in Section 2.1, traditional classifiers in Section 2.2, deep learning techniques in Section 2.3, and the data visualization techniques in Section 2.4. The different performance evaluation measures employed to evaluate the proposed system is presented in Section 2.5.

2.1 MACHINE LEARNING

Machine learning is a subsection of artificial intelligence that deals with identifying knowledge from data. Arthur Samuel coined the term machine learning in 1959 [8]. In the modern world, machine learning is applied to many aspects of technology: automated systems, personalized suggestions, advertising and marketing, robotics, to mention a few. Machine learning is subdivided into mainly three categories: supervised learning, unsupervised learning, and reinforcement learning.

- **Supervised Learning:** In supervised algorithms, the output classes are pre-determined, and we identify or model a relation that maps the features to the corresponding labels. Examples: regression, decision tree, k -NN (k -Nearest Neighbors), logistic regression, etc.
- **Unsupervised Learning:** In unsupervised learning, the target label is unknown, and the algorithm estimates the structures from data. Examples: Apriori algorithm, K-means clustering, etc.
- **Reinforcement Learning:** Here a system is designed to make specific decisions. Later it is exposed to more data, where it learns independently based on trails and error method. It learns from past experiences to make a better decision in the future. Example: Markov Decision Process

In the subsequent sections, we detail the different machine learning techniques applied in this study.

2.1.1 DISTANCE MEASURES

Distance measures play a vital role in machine learning. Most of the algorithms or techniques employ various distance measures. In our study, we investigated four different distance measures: Euclidean distance (ED), Pearson's correlation coefficient [9], Dynamic Time Warping (DTW) [10], and Mahalanobis distance [11].

2.1.1.1 EUCLIDEAN DISTANCE (ED)

The ED d between $x = (x_1, x_2, \dots, x_n)$ and $y = (y_1, y_2, \dots, y_n)$ is defined as:

$$d(x, y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \dots + (x_n - y_n)^2}. \quad (2.1)$$

2.1.1.2 CORRELATION COEFFICIENT

Pearson's correlation coefficient measures the extent of a linear relationship between two vectors [9]. The correlation coefficient value lies in the interval -1 to +1, with + and - signs indicating positive and negative correlation. The correlation coefficient ρ between two vectors x and y is given by:

$$\rho_{x,y} = \frac{\mathbf{E}[(x - \mu_x)(y - \mu_y)]}{\sigma_x \sigma_y} \quad (2.2)$$

where μ_x and μ_y denotes the mean values of x and y , σ_x and σ_y are the standard deviations of x and y , and $\mathbf{E}[x]$ is the expectation of x . Correlation distance is defined as $1 - \rho$. The correlation output can be interpreted as:

- **Positive correlation:** The value of $\rho_{x,y}$ is positive. This indicates a proportional relation between x and y , i.e., if the values of x increases, values of y also increases.
- **Negative correlation:** The value of $\rho_{x,y}$ is negative. This indicates a inverse proportional relation between x and y , i.e., if the values of x increases, values of y decreases.
- **No correlation:** The value of $\rho_{x,y}$ is close to zero. This indicates no or a weak linear relation between x and y .
- **Perfect correlation:** The value of $\rho_{x,y}$ is ± 1 .

2.1.1.3 DYNAMIC TIME WARPING (DTW)

DTW is a non-linear distance measure that estimates the shortest distance between two signals [10]. First, the optimal matching of points between the two signals is determined and later the distance between them is calculated. Specifically, the DTW distance between $x = (x_1, x_2, \dots, x_n)$ and $y = (y_1, y_2, \dots, y_n)$ is calculated as follows. First an n -by- n matrix is created

with entries (i, j) equal to $(x_i - y_j)^2$. The Sakoe-Chiba Band [12] is applied to limit the warping window size. Next all the possible cumulative distances between the two sequences are estimated for every possible path, and the minimal one is chosen as the DTW distance:

$$\text{DTW}(x, y) = \min_{w \in P} \sqrt{\sum_{k=1}^K d_{w_k}}, \quad (2.3)$$

where P is the set of all available warping paths, w_k is the position (i, j) at the k^{th} observation of a warping path, d is the distance metric, and K is the length of the warping path. We apply the optimized DTW implementation with lower bounding, and early abandoning provided in [13]. We set the Sakoe-Chiba band as 0.1. An illustration of DTW calculation between two temporal sequences S_1 and S_2 is given in Fig. 2.1.1.

Figure 2.1.1: DTW distance measure: (a) Non-linear mapping of data points of two sequences; (b) The n -by- n matrix showing the optimal path and Sakoe-chiba band).

2.1.1.4 MAHALANOBIS DISTANCE

The Mahalanobis distance is an extension of ED that considers covariance amongst the datapoints. The covariance of two vectors x and y is defined as:

$$\text{cov}(x, y) = \mathbf{E}[(x - \mu_x)(y - \mu_y)], \quad (2.4)$$

where μ_x and μ_y denotes the mean values of x and y . The Mahalanobis distance d_M between x and y is calculated as:

$$d_M = \sqrt{(x - y)' C^{-1} (x - y)}, \quad (2.5)$$

where C is the covariance matrix.

2.1.2 CLUSTERING

Clustering is a technique of categorizing different observations or data points into groups or clusters, where the observations belonging to the same cluster exhibit certain similarity. In our study, we employ clustering for Interictal Epileptiform Discharges (IEDs) data extraction and the development of the Template Matching (TM) IED detector. We have investigated five different clustering techniques: K-means clustering [14], K-medoids clustering, agglomerative hierarchical clustering [15], Fuzzy C-means clustering (FCM) [16], and Affinity Propagation (AP) [17].

2.1.2.1 K-MEANS CLUSTERING

K-means clustering is one of the traditional clustering algorithms that groups the data observations into a pre-specified number of clusters. The grouping is performed by minimizing the sum of squares within clusters [14]. Let $x = (x_1, x_2, \dots, x_n)$ be a set of n observations. X is to be clustered into k (with $k \leq n$) sets $S = (S_1, S_2, \dots, S_k)$. The objective is to minimize the objective function:

$$J = \sum_{i=1}^k \sum_{j=1}^n \|x_j - c_i\|^2, \quad (2.6)$$

where c_i is the cluster center and $\|x_j - c_i\|^2$ is the distance between the data point x_j and c_i . Initially, the set of cluster centers is chosen randomly, and then it is updated repeatedly until the squared error is below a specific threshold. The cluster center c_i is determined as the mean of the points in set S_i . The main shortcoming of the algorithm is the requirement for pre-specifying the number of clusters as the natural number of clusters is unknown in many applications. Moreover, different random initialization can lead to distinct outcomes.

2.1.2.2 K-MEDOIDS CLUSTERING

The K-medoids clustering algorithm is a partitioning method similar to the K-means algorithm, where the primary objective is to divide the set of observations into clusters to minimize the sum of distances between the cluster centers and the cluster members. The significant difference in K-medoids is that the data points are chosen as the cluster centers or exemplars, rather than the mean. In comparison with K-means, it is quite robust to outliers and noise resistant. The K-medoids algorithm is efficient in partitioning categorical data, where the mean is difficult to interpret. The number of clusters should be specified apriori.

2.1.2.3 HIERARCHICAL CLUSTERING

Hierarchical clustering is a method of building nested groups by repeated merging or splitting [15]. The two types of hierarchical clustering are:

- **Agglomerative clustering** (bottom-up approach): each observations starts as an independent cluster and later merge until convergence.
- **Divisive clustering** (top-down approach): all the observations belong to a single cluster and later split up until convergence.

The splitting or merging of a cluster is performed based on a distance metric and a linkage method. The commonly applied distance metrics are ED, Minkowski distance, correlation, and cosine. Single, complete, average, and Ward's are the conventional linkage methods. Ward's linkage method has been shown to provide better grouping than other linkages according to [18]. It is less affected by noise and outliers. Ward's distance of two clusters is the difference of within-cluster variance of two clusters separately and the within-cluster variance of the cluster generated by merging the two clusters. The specified number of clusters is set as the convergence criteria, and the cluster center is set as the mean of the cluster members.

2.1.2.4 FUZZY C-MEANS CLUSTERING (FCM)

FCM is a clustering algorithm which groups the observations or data points into multiple clusters with varying degrees of membership specified by an objective function [16]. The aim of the algorithm is to minimize the objective function:

$$J_m = \sum_{i=1}^D \sum_{j=1}^N \mu_{ij}^m \|x_i - c_j\|^2, \quad (2.7)$$

where D is the number of data points, N is the number of clusters, m is the fuzzy partition matrix exponent that determines the degree of fuzzy overlap (with $m > 1$), x_i is the i^{th} data point, c_j is the center of the j^{th} cluster, and μ_{ij} is the degree of membership of x_i in the j^{th} cluster. For a given data point x_i , the sum of membership values is one (i.e., $\sum_{j=1}^N \mu_{ij} = 1$). The exponent for the fuzzy partition matrix is set as 2, the maximum number of iterations as 100, and the minimum improvement in objective function between two consecutive iterations as 10^{-5} . At convergence, each data point is assigned to the cluster to which it has the highest membership value, and the remaining empty clusters are discarded.

2.1.2.5 AFFINITY PROPAGATION (AP)

AP is a clustering method based on an exchange of real-valued messages between the data points [17]. The inputs to the algorithm are the pairwise similarities between the data points and the initial priorities. The initial priorities specify the chances for a data point to be chosen as a cluster center or an exemplar. Unlike the traditional algorithms like K-means, the affinity propagation algorithm estimates the number of clusters based on the input similarity and priority values. Information is transferred between the data points as two distinct messages, namely responsibility, and availability (see Fig. 2.1.2).

Responsibility $r(i, k)$ is transmitted from the data point i to the potential cluster center k . This message states the degree of appropriateness for the point k is to function as the cluster center for the data point i , taking into consideration all the other potential points. Availability $a(i, k)$ is sent from a cluster center k to the data point i . This message states how appropriate it

is for the data point to join the cluster represented by the particular cluster center [17]. The responsibilities are revised as:

$$r(i, k) \leftarrow s(i, k) - \max_{k' \text{ s.t. } k' \neq k} \{a(i, k') + s(i, k')\}. \quad (2.8)$$

The availabilities are revised as:

$$a(i, k) \leftarrow \min \left\{ 0, r(k, k) + \sum_{i' \text{ s.t. } i' \notin \{i, k\}} \max\{0, r(i', k)\} \right\}. \quad (2.9)$$

Self-availability $a(k, k)$, which reflects to what extent a point k will be chosen as a cluster center, is updated as:

$$a(k, k) \leftarrow \sum_{i' \text{ s.t. } i' \neq k} \max\{0, r(i', k)\}. \quad (2.10)$$

Initially, the availabilities are set to zero. As a result, in the first iteration responsibility $r(i, k)$ is set to the difference of input similarity between point i and point k as its cluster center and the maximum of similarity between point i and other potential cluster centers (see (2.8)). This competitive update is purely data-driven. In the next iterations when some points become attached to certain cluster centers, their availabilities will drop below zero as defined in (2.9). Next, these availabilities will decrease the responsibility values, thereby removing the competition from the corresponding candidate cluster centers. For $k = i$, the responsibility $r(k, k)$ is defined as the input choice that point k be selected as a cluster center $s(k, k)$, minus the largest of similarities between point i and all other candidate cluster centers. The higher the value of self-responsibility, the higher will be the likelihood to be chosen as a cluster center.

The availability $a(i, k)$ is set to the sum of self- responsibility $r(k, k)$ and the sum of the positive responsibilities candidate cluster center k acquired from other data points. Only the positive fractions of the inbound responsibilities are considered as it is vital for a good cluster center to describe a few data points well (positive responsibilities), irrespective of how poorly it describes other data points (negative responsibilities). In case the self-responsibility $r(k, k)$ is negative (indicating at the moment that point k is better associated with another cluster rather than being a cluster center itself), the availability of point k as a cluster center can be increased if some

Figure 2.1.2: Message passing between data points in affinity propagation: (a) sending responsibilities; (b) sending availabilities (adapted from [17]).

other points have positive responsibilities for point k being their cluster center. To inhibit the influence of strong incoming positive responsibilities, the total sum is thresholded in order to prevent it from attaining a value higher than zero. Next, the self-availabilities that reflect the accumulated evidence of point k being a cluster are computed (see (2.10)).

The algorithm iteratively updates the messages until a convergence criterion or until a specified number of iterations. At any iteration, the combination of availability and responsibility values determine the clusters and their centers. For a point i , the value of k which maximizes $a(i, k) + r(i, k)$ indicates the cluster center for i . If $k = i$, point i is a cluster center itself. A damping factor λ ($\lambda \in (0, 1)$) is introduced to prevent oscillations. Each update is set to the sum of λ times the previous value and $1 - \lambda$ times the new value [17]. We set the damping parameter as 0.9 to prevent oscillations. The initial priorities are set to the median of the similarity values.

2.1.2.6 CLUSTERING EVALUATION MEASURES

Clustering evaluation measures determine the efficiency of clustering algorithms. We apply four different clustering evaluation indices to evaluate the clustering efficiency during the implementation of the TM IED detector. The four measures are: S_Dbw validity index [19], Davies-Bouldin (DB) index [20], silhouette value [21], and Mean Squared Error (MSE). A similar selection of evaluation of measures was explored in [22].

2.1.2.6.1 S_DBW INDEX: The S_Dbw validity index takes both inter-cluster compactness and inter-cluster separation into account. Smaller values of S_Dbw index indicate better clustering. S_Dbw index is defined as the sum of average scattering of clusters and inter-cluster density:

$$S_Dbw = Scatt + Dens_bw. \quad (2.11)$$

The average scattering of clusters is defined as:

$$Scatt = \frac{1}{n} \sum_{i=1}^n \frac{\|\sigma(v_i)\|}{\|\sigma(x)\|}, \quad (2.12)$$

where n is the number of clusters, $\sigma(x)$ is the variance vector of the dataset, and $\sigma(v_i)$ is the variance vector of the i^{th} cluster. The inter-cluster density is defined as:

$$Dens_bw = \frac{1}{n(n-1)} \sum_{i=1}^n \left(\sum_{j=1, i \neq j}^n \frac{\text{density}(u_{ij})}{\max\{\text{density}(c_i), \text{density}(c_j)\}} \right), \quad (2.13)$$

where n is the number of clusters, u_{ij} is the midpoint of the line segment defined by cluster centers, c_i and c_j . The term density around a point is defined as the number of points in the hyper-sphere with radius as the average standard deviation of clusters. The average standard deviation of clusters is defined as:

$$\text{stdev} = \frac{1}{n} \sqrt{\sum_{i=1}^n \|\sigma(v_i)\|}, \quad (2.14)$$

where n is the number of clusters and $\sigma(v_i)$ is the variance vector of the i^{th} cluster.

2.1.2.6.2 DAVIES-BOULDIN (DB) INDEX: The DB index is based on the ratio of inter-cluster similarity to intra-cluster similarity. Lower values indicate better clustering. The dispersion measure within a cluster is defined

as:

$$s_i = \frac{1}{n_i} \sum_{x \in i} d(x, c_i), \quad (2.15)$$

where n_i is the number of members in the i^{th} cluster and c_i the cluster center of the i^{th} cluster. The similarity measure between clusters is defined as:

$$R_{ij} = \frac{s_i + s_j}{d_{ij}}, \quad (2.16)$$

$$R_i = \max_{\substack{j=1 \dots n, i \neq j \\ i=1 \dots n}} (R_{ij}). \quad (2.17)$$

where s_i is the dispersion measure of the i^{th} cluster and d_{ij} the dissimilarity measure between the i^{th} and j^{th} clusters. The DB index is defined as:

$$\text{DB index} = \frac{1}{n} \sum_{i=1}^n R_i, \quad (2.18)$$

where n is the number of clusters and R the similarity measure of clusters.

2.1.2.6.3 SILHOUETTE VALUE: The silhouette value measures the similarity between the cluster members compared to the other data points. The silhouette value for a data point i is defined as:

$$S_i = \frac{b_i - a_i}{\max(a_i, b_i)}, \quad (2.19)$$

where a_i is the mean distance from the i^{th} point to the other members in the same cluster and b_i is the minimum mean distance from the i^{th} point to the members in different clusters. The value ranges between -1 to +1, and a high positive value indicates that the cluster members are well matched to other points in the same cluster. The silhouette value is generally defined for a single data point. We consider the mean silhouette value for comparison of clustering techniques.

2.1.2.6.4 MEAN SQUARED ERROR (MSE): The MSE is the mean distance between the cluster members and the cluster centers. MSE is defined

as:

$$\text{MSE} = \sum_{i=1}^K \sum_{j=1}^{n_i} \|x_j^{(i)} - c_i\|^2, \quad (2.20)$$

where K is the number of clusters, n_i is the number of members in the i^{th} cluster, $x_j^{(i)}$ member data point of the i^{th} cluster, and c_i cluster center of the i^{th} cluster. A low value of MSE indicates better affinity inside the cluster.

2.2 TRADITIONAL CLASSIFIERS

Machine learning classifiers are algorithms employed to perform prediction or classification (into different classes) tasks based on a previous set of labeled observations. The different traditional classifiers investigated in our study are: Template Matching (TM) [23], decision trees [24], k -NN [25], Linear Discriminant Analysis (LDA) [26], SVM [27], and AdaBoost [28].

2.2.1 TEMPLATE MATCHING (TM)

TM [23] is one of the most commonly applied for classification and regression. The objective of the method is to identify the waveforms that are similar to the waveforms present in the template library. The template library is the most crucial element of a template matching system. This is because the effectiveness of the system primarily depends on the quality of the templates in the library. The given waveforms and the templates are matched based on a prescribed distance measure. The matching between the given waveforms and the templates is based on a prescribed distance measure. The correlation coefficient, ED, and Sum of Squared Error (SSE) are the common distance measures adopted for TM. The major problem associated with TM is the quality of the templates. The choice of templates primarily determines the efficiency of the TM system.

2.2.2 DECISION TREE

A decision tree [24] is a popular and powerful tool for classification and prediction. It is capable of generating lucid and comprehensible rules that can be applied in knowledge systems. While decision trees are fast in generating decisions, error propagation exists as a significant challenge associated with

decision trees. An error at a decision node amplifies along the path originating from the node. Overfitting and underfitting are also prominent problems in decision trees. Overfitting results in a decision tree that is more complex than necessary. Pruning techniques are adopted to overcome the challenge of overfitting. Here we implement the Classification and Regression Tree (CART) [24] algorithm.

2.2.3 k -NEAREST NEIGHBORS (k -NN)

k -NN is a popular, traditional non-parametric classification method that performs grouping based on proximity [25]. We apply ED as proximity. The prediction is made based on the K training examples that are closest to the test input, and the most common class is assigned as the test label. We evaluate k -NN for different values of K and adopt the best system with the best performance.

2.2.4 LINEAR DISCRIMINANT ANALYSIS (LDA)

LDA [26] is a statistical model applied for pattern detection and machine learning. It is also referred to as normal discriminant analysis (NDA), or discriminant function analysis. LDA is a generalization of Fisher’s linear discriminant, applied to find a linear combination of features that discriminate or characterizes two or more classes of observations. The resulting feature combination could be employed for linear classification or dimensionality reduction before classification.

2.2.5 SUPPORT VECTOR MACHINE (SVM)

SVM is one of the most widely applied classifiers for EEG studies [27]. SVM aims to find the optimum hyperplane which separates different classes to the maximum. It can perform a multi-class classification. Kernel-based SVMs can be considered for non-linear classes. The major advantage of SVM is that it is independent of data dimensions. We implement two different types of SVMs: SVM with linear kernel (SVML) and SVM with Gaussian kernel (SVMG). For SVML, we optimize C (soft margin cost function parameter) and for SVMG, we optimize C and γ (free parameter of the Gaussian kernel). The higher the box-constraint or the regularization parameter (C), the higher

the cost of the misclassified points, thus leading to a more strict separation of the data resulting in a complex decision boundary. Similarly, in the SVMG model, the γ that represents the width of the Gaussian kernel determines the strength of the data points near the separation margin.

2.2.6 ENSEMBLE OF CLASSIFIERS

We also evaluate AdaBoost [28], which is a popular boosting algorithm for classification. In many cases, a single training run of a classifier might not provide the best result because the training process can get trapped in a local error minimum. In AdaBoost, a set of identical classifiers are trained sequentially, and later the different sub-system decisions are combined to produce a combined decision. We evaluate AdaBoost classifiers based on k -NN, CART, and LDA.

2.3 DEEP LEARNING

Neural networks are computational models resembling the biological nervous systems. They contain numerous interconnected processing units (neurons) working together. Neural networks can be trained to perform classification, clustering, and predictive analysis. The deep neural network consists of several stacked layers of neurons [29]. Most importantly, these networks can detect hidden features with unmarked, unstructured data with no or minimal supervision, contrary to most traditional machine learning methods. Here each layer trains on a different combination of features in concordance with the output of the preceding layer. As we advance further with layers, more complex features are trained upon from the aggregated output from the previous layer. This is referred to as feature hierarchy; the complexity and abstraction increase with the growing number of layers. Also, it performs unsupervised automatic feature extraction without human intervention. There are various types of deep neural networks. We investigate Convolutional Neural Networks (CNN) [29], Recurrent Neural Networks (RNN) with Long-Short-Term Memory (LSTM) architecture [30] in our study.

2.3.1 RECURRENT NEURAL NETWORKS (RNN)

RNN consists of units where the interconnections form a directed loop [31]. This allows the network to demonstrate dynamic temporal behavior. Contrary to feedforward neural networks, RNNs can utilize their internal memory to evaluate arbitrary sets of inputs. The human brain does not restart each time it encounters a new problem. Instead, it considers the sum of knowledge it has gained while solving past issues to address the issue at hand. NNs before RNN does not consider this persistence of memory issue. The purpose of the loop in RNNs is to allow information to persist [31]. Unrolling the loop will make the RNN a series of layers and will help to understand the classifier further. An LSTM solves the problem of long-term dependency found in RNNs. The back-propagation algorithm is not suitable in cases of large contexts and long temporal intervals. Thus, in cases with a large number of contexts, where the gap between the present state and the considered context is very high, RNNs become incapable of accurate training. LSTMs have multiple hidden layers instead of the single *tanh* layer found in RNNs, which solves this problem.

LSTM is an RNN architecture with the ability to learn from observations biased by arbitrary long time lags between relevant events [30]. This gives an added advantage for LSTM over Hidden Markov Models (HMMs), other RNNs, and other sequence learning methods for a variety of applications. The basic structure of an LSTM cell is illustrated in Fig. 2.3.1.

2.3.2 CONVOLUTIONAL NEURAL NETWORK (CNN)

CNN is a multilayered feedforward neural network [29]. The weights of the network are revised by error back-propagation. It is implemented with a combination of the following layers: the input layer, convolutional layer, Rectified Linear Units (ReLU) layer, pooling layer, and fully connected layer.

- *Input layer*: The input feature vectors or observations are presented at this layer.
- *Convolution layer*: This layer is connected directly to the input neurons. Different convolutional copies of the input are generated in this layer.

Figure 2.3.1: Basic structure of an LSTM cell.

- *ReLU layer*: This layer introduces non-linearity by applying an element-wise activation function, such as thresholding at zero.
- *Pooling layer*: This layer performs downsampling operations along the spatial dimensions.
- *Fully connected layer*: This layer computes the class scores and provides the output.

2.4 DATA VISUALIZATION WITH T-DISTRIBUTED STOCHASTIC NEIGHBOR EMBEDDING (T-SNE)

Data visualization is quite significant for data analysis. We employ t-SNE, a popular dimensionality reduction technique, to visualize the data points in two or three dimensions. T-SNE converts high dimensional data into two or three-dimensional map, preserving the similarities in the parent space. The input to t-SNE is the distance matrix between the data points. It transforms the multidimensional distances between pairs of data observations into conditional probabilities that indicate the similarity between them.

Let $x = (x_1, x_2, \dots, x_n)$ be a dataset of n observations. A relatively high conditional probability $P_{i|j}$ between two data points x_i and x_j indicates that there is a high chance that x_i and x_j would be neighbors. Similarly, a low conditional probability indicates that the points will be well separated in the lower dimensional space. Mathematically, conditional probability of data point x_j to data point x_i is given by:

$$p_{j|i} = \frac{\exp(-\|x_i - x_j\|^2 / 2\sigma_i^2)}{\sum_{k \neq i} \exp(-\|x_i - x_k\|^2 / 2\sigma_i^2)}, \quad (2.21)$$

where σ_i is the variance of the Gaussian that is centered on data point x_i . The parameter σ_i is selected based on the perplexity value specified by the user. The algorithm identifies the value of σ_i that produces a P_i with the specified perplexity through a binary search. The perplexity can be explained as the measure of the effective number of neighbors. The perplexity is calculated as:

$$\text{Perp}(P_i) = 2^{H(P_i)}, \quad (2.22)$$

when $H(P_i)$ is the Shannon entropy of P_i measured in bits.

$$H(P_i) = - \sum_j p_{j|i} \log_2 p_{j|i}. \quad (2.23)$$

The two-dimensional plot of IED waveforms and non-IED waveforms are compared with t-SNE with different distance measures. The choice of distance measure is made based on the grouping of IEDs and the separation of IED and non-IEDs. For this analysis, we consider the following parameter values:

- Perplexity Perp: Half the number of IEDs in the input dataset.
- Number of iterations T : 1000
- Learning rate η : 500
- Momentum $\alpha(t)$: 0.5 for $T < 250$, 0.8 for $T \geq 250$

2.5 PERFORMANCE INDICES

The selection of performance indicators or measures is critical while implementing systems, especially while dealing with health-care diagnostic systems. We report the performance of our system in terms of the various state-of-the-art measures described in the literature. The various statistical measures employed to compute the performance of the detection system are:

- True positives (TP): The number of true IEDs predicted as ‘IEDs’.
- False negatives (FN): The number of true IEDs predicted as ‘non-IEDs’.
- True negatives (TN): The number of true non-IEDs predicted as ‘non-IEDs’.
- False positives (FP): The number of true non-IEDs predicted as ‘IEDs’.

The performance of the classifiers is evaluated based on the following metrics:

$$\text{Sensitivity} = \frac{TP}{TP + FN}. \quad (2.24)$$

$$\text{Specificity} = \frac{TN}{TN + FP}. \quad (2.25)$$

$$\text{Selectivity} = \frac{TP}{TP + FP}. \quad (2.26)$$

$$\text{Accuracy} = \frac{TN + TP}{TN + TP + FN + FP}. \quad (2.27)$$

- False detection rate: Number of FP’s during a period. It is reported by the hour or minute.
- Receiver Operating Characteristic (ROC) curve: The ROC curve is the graphical representation between sensitivity (true positive rate) and 1-specificity (false positive rate).

- Area Under ROC Curve (AUROC): The area under the ROC curve is referred to as AUROC or AUC. It gives the probability that the system identifies a true IED over a randomly chosen non-IED as an IED. A high AUC value indicates better system performance.
- Precision-recall (PR) curve: PR curve is the graphical plot between sensitivity against the precision for various thresholds. A PR curve is more reliable than ROC in a scenario of highly class-imbalanced data.
- Area Under PR Curve (AUPRC): The area under the PR curve is referred to as AUPRC.

“Equipped with his five senses, man explores the universe around him and calls the adventure Science.”

Edwin Powell Hubble

3

Background on EEG and Epilepsy

A comprehensive literature analysis is critical for research in order to identify potential requirements and challenges. In this chapter, we review the literature on electroencephalogram (EEG), epilepsy, and epileptiform events in Section 3.1. We also highlight the need for an optimal technology-assisted epilepsy detection system. The studies related to defining an Interictal Epileptiform Discharge (IED) based on the morphological characteristics are detailed in Section 3.2. The different IED detector studies in the literature are detailed in Section 3.3 and the commercial IED detector solutions are described in Section 3.4. Furthermore, the clinical significance of identifying the epileptic features from the EEG in the diagnosis and management of the epilepsy is investigated in Section 3.5.

3.1 BACKGROUND

3.1.1 ELECTROENCEPHALOGRAM (EEG)

EEG is the electrical activity of the human brain recorded from the scalp. Discovered in 1929 by Hans Berger, a German psychiatrist, the routine scalp EEG is widely explored as a fundamental medical test for diagnosis of brain disorders such as epilepsy, coma, sleep disorders, etc. It consists of 20-30 minute recordings from scalp electrodes. The EEG is non-invasive, relatively inexpensive, and portable. Moreover, it also provides a high temporal resolution. Consequently, it is employed for a wide variety of applications, including emotion prediction, robotics, Brain-Computer Interfaces (BCIs), lie detector tests, and prosthetic devices [32–34]. EEG is recorded by placing electrodes over the scalp. The International 10-20 system is the most widely applied electrode placement system. The amplitude of a typical human EEG is in the order of 10 microvolts (μV s). There are few limitations while applying EEG as a medical diagnostic tool. Since the amplitude of an EEG recording is low, it is highly susceptible to interferences such as electrical noise, physical artifacts, environmental noises, etc. Also, since the recording is performed on the surface (scalp), the source localization could be challenging as the signals get mixed up by the time they reach the surface of the brain. Efficient localization is crucial in the scenario of surgical resections. The standard frequency bands present in the EEG are detailed in Table 3.1.1.

Table 3.1.1: Frequency bands present in human EEG [35].

Frequency band	Frequency (Hz)
Delta	< 4
Theta	4 - 8
Alpha	8 - 12
Beta	12 - 31
Gamma	> 30

3.1.2 EPILEPSY

Epilepsy is a neurological disorder characterized by unprovoked recurrent seizures. It is prevalent globally, i.e., it affects the population irrespective of gender, race, ethnic backgrounds, geographical differences, and age [36]. Epilepsy generally caused by the malfunctioning of the brain and thereby can interfere with any process of the human body. The signs and symptoms of epilepsy vary significantly from person to person. The common symptoms include temporary confusion, uncontrollable body movements, anxiety, fear, and loss of consciousness. The exact cause of epilepsy is unknown for almost half of the affected population [36]. However, some of the common causes of epilepsy include genetics, prenatal or developmental abnormalities, brain tumors, strokes, and trauma. The most life-threatening complications of epilepsy are status epilepticus and Sudden unexpected death in epilepsy (SUDEP) [36]. Status epilepticus denotes a state of persistent or continuous seizure occurrences without a recovery period. This may lead to death or brain damage. As the name suggests, SUDEP denoted sudden unexpected death due to epilepsy. Other complications due to seizures include loss of balance, drowning, car accidents, pregnancy complications, etc., mainly attributed to the loss of control over the body due to the brain malfunction.

The International League Against Epilepsy (ILAE) defines epilepsy as a disease of the brain defined by any of the following conditions: (1) at least two unprovoked (or reflex) seizures occurring greater than 24 hours apart; (2) one unprovoked (or reflex) seizure and a probability of further seizures similar to the general recurrence risk (at least 60%) after two unprovoked seizures, occurring over the next 10 years; (3) diagnosis of an epilepsy syndrome (ILAE, 14 APR 2014) [37]. Typically, seizures are divided into three phases: Aura, the first symptom of a seizure; Ictus, the middle phase; and the postictal phase, the recovery period after the seizure. The clinical manifestations of epileptic seizures include the impairment of motor, sensory, memory, and/or cognition of the patient [38]. Seizures are infrequent and alarming. Moreover, the symptoms of seizures are highly variable amongst subjects, and as a result, these could be misinterpreted. This makes the diagnosis of epilepsy based on seizures, challenging.

There are various interictal EEG events associated with epilepsy. These events are quite frequent in comparison with seizures. However, reliable

identification of these patterns is challenging. The four main interictal patterns associated with epilepsy are periodic patterns, High-frequency Oscillations (HFOs), slowing, and Interictal Epileptiform Discharges (IEDs). In critical patients, periodic EEG patterns contain discharges, which is epileptiform in appearance, that happen at even interval periods. They are generally classified as periodic lateralized epileptiform discharges (PLEDs), bilateral independent PLEDs or BIPLEDs, generalized periodic epileptiform discharges (GPEDs), and triphasic waves [39]. Periodic discharges reveal a high possibility for seizures with convulsive or nonconvulsive status epilepticus. Sometimes, the periodic patterns are not epileptiform in appearance nor epileptogenic in nature [40, 41]. The clinical implication of the periodic EEG patterns needs to be explored, as it can reveal the nature and characteristics of a variety of etiologies.

In recent years, HFOs, i.e., events with a frequency higher than 80 Hz, have been shown to have more relevant information and useful biomarker for epilepsy [42, 43]. A minimum sampling frequency of 2000 Hz is required to capture the HFOs. HFOs are present between the start and end of seizures, at seizure event onset, and during a seizure event. Interictal HFOs typically happen during slow-wave sleep [44, 45]. The HFO activity can be displayed by filtering low frequencies and quantifying the amount of high-frequency activity. HFOs are shown to have the potential to improve surgical outcomes and to accelerate intracranial EEG investigations. The investigations on HFOs are still ongoing. EEG can provide evidence for underlying cerebral impairment through background slowing activity. There are two types of slowing: focal slowing and generalized slowing [46]. Slowing could potentially identify dysfunctions but are not unique to epilepsy. Paroxysmal abnormalities in EEG or spikes (IEDs in general) that appear in the interictal stage are unique biomarkers of epilepsy [47]. Consequently, IEDs are applied clinically to diagnose epilepsy based on EEG. Therefore, in our study, we implement the epilepsy diagnostic system based on the identification of IEDs.

3.1.3 INTERICTAL EPILEPTIFORM DISCHARGES (IEDs)

According to the International Federation of Societies for Electroencephalography and Clinical Neurophysiology (IFSECN), IEDs are defined as tran-

sients distinguishable from background activity with a characteristic spiky morphology, typically, but neither exclusively nor invariably found in interictal EEGs of people with epilepsy [48]. These epileptiform transients play a crucial role in the diagnosis and treatment of epilepsy. An IED is defined based on its morphological characteristics: amplitude, duration, sharpness, and distinction from its background. The IEDs can be categorized into sharp waves, spikes, spike-wave-complexes, and polyspike-wave-complexes, defined by IFSCEN [48] as:

1. **Sharp wave:** Transient, clearly distinguished from background activity, with pointed peak at conventional paper speeds and a duration of 70 ± 200 milliseconds(ms).
2. **Spike:** Transient, clearly distinguished from background activity, with pointed peak at conventional paper speeds and a duration of 20 to less than 70 ms.
3. **Spike-wave-complex:** A pattern consisting of a spike followed by a slow wave. A wave with a duration longer than the alpha wave, i.e., over 1/8 seconds(s), is referred to as a slow wave.
4. **Polyspike-wave-complex:** A sequence of two or more spikes associated with one or more slow waves.

A typical illustration of the different IEDs is presented in Fig. 3.1.1. In our research, we focus predominantly on spikes and spike-wave-complexes.

3.2 TOWARDS QUANTITATIVE DEFINITION OF IEDS

The epileptic IED morphology has the potential to be applied in determining various neuropathologies over time. Experienced neurologists evaluate most of these abnormalities in the EEG by visual examination. However, there is a gap in defining the quantitative metrics that correspond to the IED features and their abnormalities owing to the nature and variability of different IEDs across patients. The IFSCEN has provided a clinical definition of IEDs, but a standard quantified definition is required to develop an automated, reliable system to identify IEDs. Here we review the different studies aimed towards providing a quantitative definition of IEDs.

Figure 3.1.1: A typical representation of IEDs; (a) a spike segment; (b), (c) polyspikes, and (d) spike followed by a slow-wave.

In the process of identifying distinct IED characteristics, Kooi et al. [49] conducted slope measurements for IEDs. The measurements were carried out manually for a single patient. In [50], Gloor et al. defined an IED as (A) a restricted triangular transient clearly distinguishable from background activity and having an amplitude of at least twice that of the preceding five seconds of background activity in any channel of EEG, (B) having a duration of ≤ 200 ms, and (C) including the presence of a field as defined by involvement of a second adjacent electrode [51]. Celesia et al. attempted to quantify IEDs based on amplitude and duration measurements [47]. The IED was decomposed into three transients (see Fig. 3.2.1), the corresponding amplitudes and duration were calculated. Next the amplitudes were compared to the mean peak-to-peak amplitude of the 500 ms background waveform immediately preceding the IED. The main observations were: (1) IEDs are closely associated with slow waves; (2) The amplitude of the IEDs are significantly higher than that of background (see Table 3.2.1); (3) The IED duration varied from 9 - 200 ms with a mean of 45.06 ms. This study was performed on a set of 600 IEDs. In this thesis, we conduct a similar amplitude comparison with the MGH IED database (see Chapter 4).

Figure 3.2.1: The decomposition of spikes into three transients by Celesia et al. [47]. The duration of the three transients of the spike and slow-wave complex is represented by a , b , and c .

Total number of IEDs	Smaller or 0-30% higher than b.a.	31%-50% higher than b.a.	51%-100% higher than b.a.	101%-300% higher than b.a.	301%-500% higher than b.a.	501% higher than b.a.
600	15.5%	9.8%	21.2%	42.4%	11%	0.8%

Table 3.2.1: Comparison of IED amplitudes with a mean peak-to-peak amplitude of the 500 ms background waveform immediately preceding an IED. Adapted from the study by Celesia et al. [47].

Further, Ktonas et al. [52] modified the slope parameters and devised a digitized technique to quantify IEDs based on six parameters (amplitude, duration, and slopes of the first and second half-waves). The study was limited to three patients and 16 IEDs. Incorporating all these parameters, Gotman et al. [53] quantified the asymmetry in amplitude, duration, and slope between rising and falling phases of IEDs (see Fig. 3.2.2). The asymmetry was compared between groups of channels-unilateral, bilateral primary side, and bilateral secondary side. The results showed that IEDs had a significant asymmetry in their amplitude and duration in different channels. This study was performed on a set of 60 recordings. Further, few more attempts towards extracting the morphological characteristics of IEDs have been performed [54–56]. However, a standard quantitative definition for an IED has not been yet formulated.

Figure 3.2.2: Six parameters of IEDs defined by Gotman et al [53]. Amplitude and duration of the first and second half-waves are measured between A–C and C–E, respectively. B and D are the midpoints between A–C and C–E, respectively. CB and CD define the slopes.

3.3 IED DETECTION

In this section, we review the main IED detection methods described in the literature. The different types of IED detection methods available are mimetic, parametric, traditional classifier-based, Neural Network (NN) based, and wavelet-based. The performance of the various methods is reported in terms of four performance measures: sensitivity, specificity, selectivity, and false-positive rate. Sensitivity is the proportion of IEDs that are correctly identified from a set of IED and non-IED waveforms. Specificity is the proportion of the non-IED or background waveforms that are correctly classified. Selectivity is the proportion of true IEDs in the entire set of waveforms that are predicted as IEDs. The false-positive rate (false detections reported by the hour or minute) is the rate at which non-IEDs are misclassified as IEDs. The performance measures are mathematically described in Section 2.5.

3.3.1 MIMETIC

The mimetic approach is one of the most applied techniques for IED detection [54, 55, 57–77]. This approach tries to analyze the IED waveform in terms of morphological characteristics, as observed by an expert. The IED waveform is decomposed into different transients and the features, namely,

amplitude, duration, and slope, are extracted. Mimetic method has been applied independently [72, 73, 75–77] or in conjunction with other methods [62, 64, 65, 70].

James et al. in [54] implemented a combined system of mimetic, Artificial Neural Network (ANN), and fuzzy logic. The mimetic parameters extracted are amplitude (A_p), the duration of half-waves (D_1, D_2, D_3), rising and falling slopes (S_1, S_2), and sharpness (S_p). Three contextual parameters: average amplitude ($\hat{A}$), average duration ($\hat{D}$), and average slope ($\hat{S}$) are also extracted. The features are illustrated in Fig. 3.3.1. The dataset consisted of 8 patients with 65 certain and 125 questionable epileptiform events. The mimetic stage achieved an overall sensitivity of 91.1% coupled with a very low selectivity of 0.8% and a false detection rate of 6796/hr.

Figure 3.3.1: Mimetic parameters described by James et al. in [54].

Boos et al. [55] applied a combination of 7 morphological feature categories: amplitude, vertex angle of peaks, initial region of epoch, final region of epoch, statistical indices, and entropy. The ANN was trained with a set of 45 descriptors to classify epileptiform events from artifacts and background EEG. The analysis was conducted on a set of 77 epileptiform events (47 for training and 30 for testing). The system was reported to have achieved a sensitivity and specificity of 90% and 87% respectively. Liu et al. defined a model-based IED detector in [56]. The study was conducted on 142 IEDs from 12 patients, annotated by two neurologists. Four categories of features were extracted: duration, amplitude, slope, and area. The IED model was proposed based on nine features, and the effectiveness was compared with the conventional model (6 features) and a model with 13 features. The refer-

ence diagram for feature extraction is illustrated in Fig. 3.3.2. The proposed system achieved a sensitivity of 95.5% and a specificity of 92.4%.

Figure 3.3.2: The reference diagram for feature extraction presented by Liu et al. in [56].

The major problem with the mimetic method is that there is no standard quantitative definition for IEDs. Consequently, the mimetic parameters are subjective, and there is no uniformity in defining the IED parameters. Many other EEG transients such as artifacts, sleep EEG patterns (K-complexes, spindles, vertex waves), etc., exhibit similar morphological features as IEDs. Such waveforms may be labeled as IEDs. Therefore, they cause interference with the detection process. Some new extensions of the mimetic method by applying Fourier Transform (FT), Walsh Transform (WaT), etc., have also been implemented in [58, 59, 62, 64].

3.3.2 PARAMETRIC MODELING

Parametric modeling is an alternative approach to fit mathematical models to EEG signals. It is following the hypothesis that the EEG signal is stationary, i.e., the mean, variance, and autocorrelation function of the signal stay constant over time. A non-stationary occurrence indicates epileptiform events. Sankar et al. modeled an EEG signal as a weighted linear combination of its past values and the present value of a hypothetical input (u_m) to the filter whose output is the EEG signal (s_n). The linear model is illustrated in Fig. 3.3.3.

Figure 3.3.3: The linear prediction model applied in the study performed by Sankar et al. in [78].

The time series representation of the EEG is given by:

$$S_n = - \sum_{k=1}^p a_k S_{n-k} + G u_n, \quad (3.1)$$

where parameters a_k are the linear prediction coefficients, p is the prediction or model order, and G is the gain. The residual error is calculated as e_n :

$$e_n = s_n - \hat{s}_n = G u_n. \quad (3.2)$$

The model parameters a_k 's are obtained by minimizing the total squared error by applying least square analysis:

$$\sum_{k=1}^p a_k \sum_{n=0}^{N-1-i} s_{n-k} s_{n-i} = \sum_{n=0}^{N-1-i} s_n s_{n-i}. \quad (3.3)$$

The auto-correlation function R is defined as:

$$R(i) = \sum_{n=0}^{N-1-i} s_n s_{n+i}, \quad (3.4)$$

where $i \geq 0$ and N is the number of samples in the observation window.

Applying autocorrelation function, (3.3) reduces to

$$\sum_{k=1}^p a_k R(i - k) = -R(i), \quad (3.5)$$

where R is the auto-correlation function and $1 \leq i \leq p$. The predictor coefficients are estimated by solving (3.5). Once the parameters are evaluated, the output values are calculated for each input waveforms, and the final decision is made based on it. The main problem associated with the parametric modeling approach is that the epileptiform events tend to be less stationary compared to other EEG events such as the artifacts.

3.3.3 TRADITIONAL CLASSIFIERS

Different traditional classifiers (see Section 2.2) have been applied with and without feature extraction for IED detection. Lodder et al. devised a self-adapting Template Matching (TM) system based on a set of smart templates in [79]. The study was conducted on a set of 23 patients with an IED count of 723. The sensitivity was reported as 95% and Fp/minute as 6.6, after 15 iterations. Nonclercq et al. applied temporal clustering to a set of IEDs from a set of 3 patients (2500 IEDs) to develop a template library. Subsequently, a system which adapts to interpatient and inpatient variation in IED morphology was devised based on the template library. The sensitivity and selectivity of the system were reported as 93.7% and 93.7% respectively [80]. The performance of a TM system primarily depends on the quality of the template library. It is challenging to identify the best template library in many cases. Zacharaki et al. have employed a Support Vector Machine (SVM)-based system for IED pattern recognition in [81]. The analysis was performed on data from a single patient with 101 IEDs, thereby making it unreliable.

3.3.4 WAVELET ANALYSIS

Wavelet Transform (WT) is an effective method for signal analysis similar to the Fourier transform (FT). FT decomposes a temporal waveform into a series of sine waves of different frequencies, whereas WT decomposes the signal into wavelets, scaled and shifted versions of the mother wavelet. WT is a multi-resolution approach. It allows localization in both time domain and

frequency domain utilizing translations and scaling, respectively. WT has been applied to identify various trends, discontinuities, and patterns in EEG recordings. Many recent algorithms have demonstrated the effectiveness of WT in IED detection [82–85]. The two types of WT are Continuous Wavelet Transform (CWT) and Discrete Wavelet Transform (DWT). CWT provides the similarity between the signal and an analyzing function utilizing the inner product. The analyzing function is a wavelet ψ . The CWT compares the signal with the shifted and compressed or stretched version of a wavelet. The continuous wavelet transform of a function $f(t)$ is given by:

$$C(a, b) = \frac{1}{\sqrt{b}} \int_{-\infty}^{\infty} f(t) \psi\left(\frac{t-a}{b}\right) dt, \quad (3.6)$$

where $\psi(t)$ is the mother wavelet function, a is the translation or shift parameter, and b the scaling parameter. By continuously varying the values of a and b , we obtain the CWT coefficients $C(a,b)$. Mexican Hat and Morlet are the common CWTs.

DWT decomposes the signal into mutual orthogonal sets of wavelets. It is implemented by a series of high-pass and low-pass filters. The signal is convolved with the filters followed by a down-sampling step to obtain the coefficients. A low-pass filter convolution provides the approximation coefficients while the detail coefficients are obtained by convolving the signal with a high-pass filter. The entire decomposition of a discrete signal $f[n]$ is illustrated in Fig. 3.3.4. IED detection is performed by extracting the DWT coefficients as IED features. Daubechies, Coifman, and Haar wavelets are some of the common DWTs.

Indiradevi et al. applied a 6-level DWT (Daubechies 4) decomposition with an adaptive threshold on sub-bands 4 and 5, to automatically identify of epileptiform activity [83] in EEG signals. The study was conducted on EEG data from 22 patients (684 IEDs) annotated by two experts. The sensitivity and specificity of the system were reported as 91.70% and 89.30% respectively. Chavakula et al. devised an IED quantification approach by applying a double threshold wavelet transform in [82]. The system was reported to exhibit a sensitivity of 78% and a specificity of 96%. In wavelet analysis, the choice of optimum scale parameters and wavelet basis functions is still an open research topic.

Figure 3.3.4: Two-level decomposition of a signal by DWT into approximation and detail coefficients.

3.3.5 NEURAL NETWORKS (NN)

Several NN-based approaches have been explored for IED detection in the literature [5, 84, 86–88]. Wilson et al. developed a multiple monotonic NN-based IED detection method in [66]. The study was conducted on a set of 50 epileptic patients (IED count of 2400) and ten control subjects. The comparison between the performance of the proposed NN-based method with traditional visual inspection and rule-based method is evaluated. The proposed method achieved a sensitivity of 89.9% and specificity of 99.63%.

Carey et al. employed a standard feedforward NN with the backpropagation algorithm for IED detection. The NN contained a single hidden layer with ten neurons. The output layer consisted of one binary neuron, and the input layer contained 120 neurons. The design of the NN is illustrated in Fig. 3.3.5. The network was trained with features extracted from the IED and non-IED waveforms. The study was conducted with a set of 6 patients (the number of IEDs is not reported). The sensitivity and selectivity were reported as 82.68% and 72.67% (specificity is not stated) [86].

3.3.6 POWER SPECTRAL ANALYSIS (PSA)

PSA is a classical method of defining an EEG signal in terms of frequency power quantified among different frequency bands. The major types of power spectral analysis explored in detecting epileptiform events are FT, Hilbert transform (HT), and Walsh Transform (WaT) [59, 61, 89]. The primary lim-

Figure 3.3.5: The design of the feedforward NN applied by Carey et al. in [86].

itation of this method is the fixed time-frequency resolution. Consequently, this method has been substituted by wavelet analysis.

3.3.7 DRAWBACKS AND CHALLENGES

In this section, we provide the comparison between the methods that have been explored for IED detection. The methods are summarized chronologically in Tables 3.3.1, 3.3.2, 3.3.3, and 3.3.4.

Table 3.3.1: Previous IED detector studies in literature.

Ref.	Dataset count	IED Count	Algorithm	Performance Measures			
				Sensitivity	Specificity	Precision	FP/min
[90]	100	1815	2D-CNN	47.4%	98-99%	-	0.03-0.6
[91]	19	1491	DBN	87.35%	97.89%	-	-
[92]	8	469	Eigen value analysis Clustering	70.8%	-	39%	-
[93]	12	5582	EpiSpike 1.5 ANN	34.6% 44.1%	-	58.5% 56.2%	0.33 0.33
[94]	50	8929	AP - TM	90%	85.9%	66.8%	-
[95]	17	2157	WT, RF	62%	-	26%	-
[96]	40	5474	Human experts Persyst 13	51.5% 43.9%	-	-	1.99 1.65
[97]	5	7500	CNN		AUC =0.947		
[86]	6	NR	NN	82.68%	-	72.67%	-
[98]	NR+3	NR+1308	WT, Fuzzy Logic	80%	70%	-	-
[81]	1	101	SVM	97%	63	-	0.1
[79]	8+15	482+241	Self-adapting TM	92%	-	-	6
[99]	8+15	482+241	TM	90%	-	-	2.36
[56]	15	142	Adaboost classifier	95.5%	92.4%	-	-
[5]	100	2571	ANN WT	65.3%	80.6%	-	-

Key: CNN: Convolutional Neural networks, ANN: Artificial Neural Networks
 DBF: Deep Belief Networks, TM: Template Matching, WT: Wavelet Transform
 AP:Affinity Propagation, RF: Random Forest, SVM: Support Vector Machine
 GA: Genetic Algorithm, NN: Neural Networks, '+': Multiple sets, NR: Not Reported.

Table 3.3.2: Previous IED detector studies in literature.

Ref.	Dataset count	IED Count	Algorithm	Performance Measures			
				Sensitivity	Specificity	Precision	FP/min
[100]	7	621	ICA Quadratic classifier	65%	86%	-	-
[83]	22	684	WT	91.7%	89.3%	78.10%	-
[101]	8	1625	Dipole analysis	92%	-	77%	-
[57]	5+3	43+15	Clustering NN	60.4%	98.7%	81.8%	-
[58]	25	274	Mimetic, NN CLustering, ES	89.2%	91.1%	82.5%	-
[59]	25	274	FT, Mimetic	86.1%	91.7%	83.1%	-
[84]	7	6721	WT, NN, ES	70.9%	99.99%	-	0.13
[60]	19+10	216+93	Mimetic, NN	89.1%	-	85.9%	0.125
[61]	9+9	47+139	Mimetic, WaT	79%	-	85%	-
[62]	10+21	319	Mimetic, WaT	82%	93%	92%	-
[102]	4	340	WT	70%	-	67%	-
[63]	6+20	17302+23619	Mimetic Dipole Analysis	-	-	-	-
[64]	11	89	FT, Mimetic	37%	-	-	16
[85]	81	6048	Parametric, WT NN, ES	90%	-	93.9%	-

Key: ICA: Independent Component Analysis, WT: Wavelet Transform
 FT: Fourier Transform, ES: Expert System, WaT: Walsh Transform
 NN: Neural Networks, NR: Not Reported, NA: Not Available, '+': Multiple sets.

Table 3.3.3: Previous IED detector studies in literature.

Ref.	Dataset count	IED Count	Algorithm	Performance Measures			
				Sensitivity	Specificity	Precision	FP/min
[103]	4	54+NR	NN	62.3%	-	61%	-
[104]	20	300	NN	-	-	-	-
[105]	11	298	WT	84%	-	12%	-
[65]	521	NR	Mimetic, ES	-	-	-	0.0068
[66]	50	1952	Mimetic, NN	47%	99.63%	-	2.5
[67]	6+18	NR+982	Mimetic, ES	96%	-	-	0.4
[106]	10	166	WT	70%	-	52%	1.28
[66]	50	1952	Mimetic, NN	47%	99.63%	-	2.5
[67]	6+18	NR+982	Mimetic, ES	96%	-	-	0.4
[106]	10	166	WT	70%	-	52%	1.28
[54]	35+8	3096+190	Mimetic, NN	55.3%	-	82%	0.12
[107]	5	NR	NN	93%	94%	-	-
[108]	32	NA	WT, NN, ES	97%	-	89.5%	-
[89]	4	1509	HT, NN	89.9%	-	84.6%	1.83
[68]	137	NR	Mimetic	-	-	-	-
[109]	1	982	Parametric, WT	86%	-	-	6.8

Key: HT: Hilbert Transform, WT: Wavelet Transform, ES: Expert System

Key: FT: Fourier Transform, NN: Neural Networks

NR: Not Reported, NA: Not Available '+' : Multiple sets.

Table 3.3.4: Previous IED detector studies in literature.

Ref.	Dataset count	IED Count	Algorithm	Performance Measures			
				Sensitivity	Specificity	Precision	FP/min
[110]	5	1200+2414	WT, NN	90.8%	93.2 %	-	-
[69]	5+5	927	Mimetic, NN	73.6%	-	73.6%	6.1
[70]	6	NR	FT, Mimetic	31%	33%	-	-
[71]	11	462	Mimetic, ES	67%	100%	-	0
[78]	11	NR	Parametric	-	-	17.5%	
[72]	5	1393	Mimetic	76 %	-	89%	5.2
[111]	5	652	NN	94.2%	-	79.1%	1.5
[73]	20	NR	Mimetic State analysis	-	-	72%	0.79
[74]	14+6	NR	State analysis Mimetic	-	-	-	0.25
[75]	5+5	NR	Mimetic	65%	-	-	-
[112]	10	341	Parametric Template matching	73%	-	-	1
[76]	24	NR	Mimetic	-	-	-	0.11
[77]	93	NR	Mimetic	-	-	-	0.33

Key: ES: Expert System, TM: Template Matching
 NN: Neural Networks, NR: Not Reported, '+': Multiple sets.

The following are the significant issues and challenges gathered after reviewing the literature.

1. **Lack of a sizable database:** Many methods described in the literature are trained and tested on small databases with fewer IEDs (less than 1000). It is evident from the literature that IEDs exhibit considerable variability in morphology across patients. Therefore, the methods evaluated on small databases are unreliable.
2. **Lack of a standard IED database:** The various methods available in the literature are tested on different datasets. Therefore, the comparison between different studies is challenging.
3. **Limited number of interpreters:** In most studies, the number of EEG interpreters is limited (one or two). Consequently, the studies tend to depend primarily on the experience and expertise of the experts. The limited number of interpreters is mainly attributed to reducing the budget of the study.
4. **Correlated testing and training datasets:** Most of the studies do not employ independent datasets for training and testing. This leads to skewed results that are difficult to interpret.
5. **Diverse IED feature definitions:** The definitions of IED features proposed are highly diverse. This makes the interpretation of methods, specifically mimetic and parametric modeling, difficult.
6. **Lack of artifact rejection mechanisms:** The artifacts such as eye blinks, muscular movements, etc., are similar to the epileptiform events in many aspects. They pose a significant problem for methods such as parametric modeling and mimetic. Most studies do not consider artifacts into account.
7. **Diverse performance evaluation metrics:** The performance metrics reported tends to be subjective. For instance, most of the studies report high values of sensitivity and specificity but do not report false-positive rates (false detection per hour or minute). High sensitivity and specificity can be correlated with high false-positive rates, especially when the IED count is high.

8. **Computational complexity:** The computational complexity of the system is also a concern. Most studies achieved good performance in terms of IED detection for smaller datasets, but the computational aspects such as the processing time, memory requirement are not reported. The feasibility of implementing the methods in real-time applications is also not indicated. Such information is essential for deploying the system into clinical settings.
9. **Lack of clinical validation:** Most of the methods in the literature are not validated clinically. Only a very few systems are currently available in the market such as Persyst [96] and B-Alert.

Consequently, it is difficult to compare the different methods in the literature due to the highly diverse datasets and performance evaluation measures. Most studies that reported good detection performances are only validated with a smaller database of IEDs. Besides, the computational complexity of the methods is not reported. However, better performance was reported for template matching systems in comparison with the other methods. The development of a robust and reliable, automated, or semi-automated IED detection system remains an open challenge.

3.4 COMMERCIAL EEG SOFTWARE

We identified Persyst as the current market leader in EEG software solutions [96]. In this section, briefly review Persyst. Scheuer et al. have proved that Persyst 13 was non-inferior to the human experts in [96]. However, a recent study by Halford et al. presents the inferior performance of the software disproving Scheuer et al. [113]. The Persyst IED detector is produced by the Persyst Development Corporation, San Diego.

The P13 IED detector employs the International 10–20 system EEG recording in a common average referential montage as input. The performance of the system is calibrated based on the power spectrum (delta and alpha activity) of the initial three minutes of the recording such that the visual coordinates can be applied to identify an IED, similar to that performed by neurologists. Each EEG channel is then expressed by a stream of curvature extrema and inflection points. These are later applied to identify

the alternating rhythmic half-wave segments. Finally, a six half-wave model is employed to represent an IED.

The first two half-waves estimates the activity immediately preceding the IED, the middle two describes the rising/falling deflections, and the final two half-waves details the presence of slow-wave discharges. Each half-wave is characterized by its amplitude, duration, and curvature. These features are employed to distinguish the IEDs from the background activity. Later the IED detections from the individual channels are concatenated to denote the IED event field.

The non-cerebral activities such as electrocardiogram (ECG), muscle, eye artifacts, etc., are identified and removed. Later multiple feedforward neural network rules (NN-rule) are employed to describe an IED. Each NN-rule consists of an ensemble of about 40 individual NNs. In order to expedite the process of non-IED event removal, three pre-specified feature thresholds were applied. Later the identified events were assigned perception values, with a higher value indicating an IED. The uncertain events were assigned values near 0.5 [96].

3.5 EPILEPSY DIAGNOSIS AND MANAGEMENT THROUGH EEG

In this section, we analyze the potential and practicality of the proposed idea based on the available statistics. Epileptiform activity based on EEG is specific, but not sensitive. EEG has a comparatively low sensitivity to diagnose epilepsy (25–56%) [114]. Although with EEG, better specificity can be achieved, it ranges between 78–98% [114]. The relationship between various EEG patterns and epilepsy is incidence specific, and only IEDs have exhibited adequately high clinical significance with epilepsy. The incidence of epileptiform events in the routine EEG recording of healthy adults with no past incidence seizures was 0.5% [115].

Numerous reasons determine whether the epileptic patients exhibit IED in the EEG recording. IED can be seen in some epilepsy groups and seizure types. Besides, the epileptogenic location also plays an essential role in IED identification. Majority of the temporal lobe epilepsy patients exhibit IEDs, while in patients with epileptic foci in mesial or basal cortical regions, it is challenging to identify IEDs based on scalp recordings. It is also reported that patients with persistent (at least one per month) seizures are

more prone to exhibit IEDs in comparison with those having rare (one per year) occurrences [116]. The EEG recording period may be relevant: it is reported that the EEG examination within 24 hours of a seizure incidence has a higher probability of IED occurrence (51%) than those who had it later (34%) [117]. Medications could also be potentially relevant, especially those that minimize seizure threshold or those that induce epileptiform activity themselves. A combination of wake and sleep records has shown an identification rate of 80% in epileptic patients [118].

Approximately 10% of the epileptic patients never display epileptiform discharges [114]. Moreover, the presence of IED in EEG does not confirm seizure disorder, as IEDs can be observed in healthy subjects to some extent. Also, IEDs may also be observed in patients with different neurological abnormalities [114]. This could be resolved by analyzing a large amount of data and learn the characteristic of the epileptic conditions. From the above studies, it has been shown that multiple factors influence the nature of the epileptic activity. Thus accurate detection of epilepsy is both challenging and necessary to effectively diagnose the chronic condition. Summarizing, EEG could be applied for epilepsy as follows [114]:

- Diagnosis
 - Diagnosis of paroxysmal neurological events
 - Syndrome-specific changes identification
 - Detection of photosensitivity
 - Differentiate between a focal and generalized seizure disorder
- Management
 - Evaluating the risk of re-occurrence after seizure episodes
 - Identifying the proper antiepileptic treatment
 - Determining the effects of medication
 - Determining the epileptogenic zone for epilepsy surgery
 - Examination of cognitive decline
 - Identification of nonconvulsive status
 - Observing in convulsive status

Long-term monitoring generates enormous amounts of data for analysis. The current gold standard of IED detection is the visual inspection of EEG carried out by experts. This makes the process time-consuming, tedious, and subjective. Our research aims to provide an alternative to the current manual detection. The objective of our study is to develop a standardized method build on large amounts of data with deep learning techniques to evaluate and detect IEDs quantitatively, that would aid the neurologists to make a reliable and faster diagnosis. We propose two IED detectors: TM-based and CNN-based. A TM system is a simple model where the IED waveforms are directly applied for learning and recognition without any significant processing. This could potentially provide a clinically acceptable IED detector as the IRA is quite low amongst the neurologists. However, TM systems are quite susceptible to disturbances. On the other hand, CNN is capable of providing robust, reliable detection performance compared to the various systems in the literature. They are capable of extracting highly complex features from raw data with no or minimal supervision. Finally, combining both the techniques, we develop a clinically validated epileptic EEG identification system. The system was successfully clinically validated at National University Hospital (NUH), Singapore. We achieved an accuracy of 88% for classifying EEGs as epileptic and non-epileptic on a dataset of 100 subjects.

“Facts are the air of scientists. Without them, you can never fly.”

Linus Pauling

4

Scalp EEG Data and Preprocessing Techniques

A reliable, sizable dataset is essential for developing efficient machine learning models. Moreover, the data preprocessing, extraction, and evaluation methodology significantly influence the system performance measures. In this chapter, we present a detailed overview of the EEG datasets and the various data preparation techniques investigated in this thesis. We employ multiple datasets from different hospitals in our study. The different datasets are described in Section 4.1. Following this, the data preprocessing techniques are presented in Section 4.2 and data extraction techniques in Section 4.3. Further, we have formulated a standard quantitative definition of IEDs based on the morphological characteristics in Section 4.4.

4.1 CLINICAL EEG DATA

In our research, we analyze 30-minute routine scalp EEG recordings from multiple centers. All the data has been recorded according to the 10-20 electrode system (see Fig. 4.1.1) with varying sampling frequencies. The different centers are:

- Massachusetts General Hospital (MGH), Boston, USA
- Medical University of South Carolina (MUSC), South Carolina, USA
- National University Hospital (NUH), Singapore

In the MGH dataset [119], we have the annotations for individual IEDs as well as EEG-level annotations, i.e., whether an EEG is from an epileptic subject following the clinical records. The MUSC dataset [5] contains only annotated IED segments, whereas the NUH dataset [120] consists of only EEG-level annotations. Consequently, the development of our system is performed exclusively on the MGH dataset. The datasets from other centers are employed to evaluate the performance and validate our system. Along with the labels, we have also recorded different information regarding the subjects, namely, age, gender, and state of the subject while recording the EEG (sleep/awake). This information could be applied as additional features for improving the performance of the epilepsy diagnostic system.

4.1.1 MGH DATASET

The MGH dataset [119] consists of 93 recordings from 84 epileptic patients and 461 IED-free or non-epileptic EEGs. The additional seven recordings are collected from the existing patients but recorded longitudinally over time. The data were recorded at multiple sampling frequencies of 200 Hz and 250 Hz according to clinical settings. The epileptic dataset contains data from 46 males (age 34.58 ± 27.22 years) and 47 females (age 35.40 ± 28.36 years). The age-gender statistics are illustrated in Fig. 4.1.2. A total of 18,164 individual IEDs instances are annotated by two neurologists, independently. We consider the waveforms with majority agreement (both the annotators) as IEDs. The IED events are marked for each time instant with the channel information on which the most significant IED pattern appears (see Fig. 4.1.3).

Figure 4.1.1: The electrode placement according to the 10-20 International system on a two dimensional human scalp (top-view).

There are multiple transients annotated at the same time-instant; we have 14,170-time events that are marked as epileptiform events. The inter-IED annotation time window was set as 250 milliseconds (32 samples), i.e., only annotated IEDs with a separation of 32 samples are considered for the analysis. Efforts had been made to annotate all the IEDs present in the EEGs. However, there could be overlooked epileptiform events. The distribution of the annotated IEDs are presented in Table 4.1.1 and illustrated in Fig. 4.1.4.

Table 4.1.1: Annotated IED event distribution of the MGH dataset.

IED events range	Number of patients
(0,15]	48
(15,100]	18
(100,500]	17
501+	10

Figure 4.1.2: Age statistics for the MGH dataset.

Figure 4.1.3: A ten-second sample from a 19-channel EEG recordings of a subject with epilepsy. The annotated IED is highlighted in the circle.

Figure 4.1.4: IED event distribution in the MGH dataset.

4.1.2 MUSC DATASET

The MUSC dataset [5] consists of 200 30-second EEG segments with 235 IEDs cross-annotated by thirty-five EEG scorers (EEGers). The 200 segments were extracted from approximately 1,000 scalp EEGs performed in the Medical University of South Carolina (MUSC) Neurophysiology Laboratory for clinical purposes during the years 2010 to 2011. The scoring was conducted in three phases with the EEGnet software [5]. The data was recorded at multiple sampling frequencies of 200 Hz, 250 Hz, and 512 Hz. The 200 segments were extracted from 4 different subject groups: 50 from EEGs containing IEDs, 50 from normal EEGs, 50 from EEGs containing difficult-to-interpret IEDs, and 50 from EEGs with benign paroxysmal activity (e.g., exaggerated alpha activity, wicket spikes, and small sharp spikes) that could be easily misinterpreted as IEDs. The MUSC dataset annotation consisted of two phases. In Phase 1, 35 annotators analyze the EEG segments and indicate the presence of IEDs as ‘1’ for different time instants. A total of 573 IED instants were annotated by atleast one expert. In Phase 2, 235 selected IEDs were re-examined by 18 experts on a scale of 1 to 5, with 5 indicating a deterministic IED. The database is further described by Halford et al. in [121]. In our analysis, we employ this dataset to evaluate the

performance of our proposed system as well as compare it with the existing software in the market.

4.1.3 NUH DATASET

The NUH dataset [120] consists of 150 routine scalp EEGs recorded at National University Hospital (NUH), Singapore during the period 2016-2017. The data was collected from NUH Singapore under the supervision of Dr. Rahul Rathakrishnan, Senior Consultant with the Division of Neurology. This study was conducted under the National Health Innovation Center (NHIC) grant (Ref. no. : NHIC-I2D- 1608138) titled “Clinical development and validation of an automated EEG data interpretation system”. The data were recorded at a sampling frequency of 250 Hz. The data collection was performed in two phases: Phase 1: 50 EEGs (25 epileptic and 25 non-epileptic) with labels for calibrating the proposed EEG classification system; Phase 2: 100 EEGs (50 epileptic and 50 non-epileptic) without labels. The phase 1 data non-epileptic EEG set consisted of 17 males (age 47.82 ± 18.71 years) and eight females (age 58.75 ± 18.38 years). The phase 1 epileptic EEG set consisted of 14 males (age 53.42 ± 20.02 years) and 11 females (age 55.54 ± 20.89 years). The phase 2 data epileptic EEG set consisted of 29 males (age 59.03 ± 21.45 years) and 21 females (age 63.85 ± 17.76 years). The phase 2 non-epileptic EEG set consisted of 29 males (age 44.34 ± 17.06 years) and 21 females (age 50.80 ± 21.98 years).

Initially, the phase 1 data was transferred with labels to calibrate the EEG classification system. Later the predicted labels for phase 2 EEGs was sent to the hospital in order to compare it with the clinician’s report. We employ the NUH dataset for evaluating our epilepsy diagnostic system. Since the evaluation was performed without the labels (phase 2), these results provide clinical validation for our system.

4.2 PREPROCESSING

The amplitude of the EEG recording is in the range of microvolts. Subsequently, it is susceptible to various disturbances such as power-line interference, environmental noise, equipment related-noise, etc. An appropriate preprocessing routine is necessary to extract information from the recordings. The different preprocessing methodologies adopted in our research are highlighted in this section. We performed the following preprocessing steps:

- Filtering
- Down-sampling
- Power-based artifact removal
- Removal of low information segments
- EEG montage

The flow diagram of the proposed preprocessing steps is illustrated in Fig. 4.2.1.

4.2.1 FILTERING

We applied two filters: an Infinite Impulse Response (IIR) filter (order 4) to remove the electrical interference, and a 1 Hz high-pass (order 4) filter (HPF) to remove the Direct Current (DC) shift and baseline fluctuations. The notch filter cutoff frequency was set as 60 Hz for MGH and MUSC datasets, and 50 Hz for the NUH dataset, following the respective power line standards. The HPF also helps to remove the perspiration artifact, which usually manifests as low amplitude waves with a duration greater than 2 seconds.

4.2.2 DOWN-SAMPLING

The original data was recorded at multiple sampling frequencies: 200 Hz, 250 Hz, and 512 Hz. The data is down-sampled to 128 Hz after applying an anti-aliasing Finite Impulse Response (FIR) low-pass filter (LPF). We have confirmed the epileptiform events are not distorted at 128 Hz. This also helps us in reducing the implementation complexity of the system.

4.2.3 POWER-BASED ARTIFACT DETECTION

We performed an artifact rejection based on noise statistics. First, we divided the data from each channel into consecutive segments of 1 second and computed the root mean square (RMS) value for each segment. For each channel data, the segments that exhibited a higher RMS value than a threshold was identified as a noisy segment. This mechanism helps us to remove localized artifacts such as electrode popping, faulty electrode contacts, undesired glitches, etc. The threshold was set as $median \pm 4 \times standard\ deviation$ of the RMS values. In this step, we are not removing the noisy segments since the noise is localized. The other EEG channels might contain epileptiform patterns.

4.2.4 REMOVAL OF MISSING ELECTRODE SEGMENTS

The different channels of the EEG are evaluated, and the missing electrode segments are removed. We consider the segments with 6 or more channel data missing or identified as noisy by the artifact rejection step as segments with inadequate information.

4.2.5 EEG MONTAGES

The Common Average Referential (CAR) montage is applied since it is more effective for identifying focal activity by removing the general trend from the EEG. It also helps to minimize the effect of Electrocardiogram (ECG) induced artifacts due to the ECG's equipotential field.

Figure 4.2.1: Different preprocessing steps applied on the EEG recordings.

4.3 IED AND NON-IED EXTRACTION

In this section, we detail the various steps involved in the data extraction process. We extract the IED segments and non-IED segments to train the waveform-level pattern identification module. After applying the preprocessing module, each annotated IED was extracted as a 500-millisecond (64 samples) waveform. The IED is annotated at random locations by the neurologists. Often the IEDs are annotated at the onset, but sometimes there could be a shift of few sample points. To make sure that the 500-millisecond window covers the IED and the associated events, we employ a clustering mechanism to realign the IEDs. We apply K-means clustering with the similarity measure as squared Euclidean distance (ED). We randomly set the number of clusters, K as 1000. We fix the onset of IEDs to be approximately at sample point 16 (with a tolerance range of ± 2 samples). We identify the different peaks of the IED by applying ‘findpeaks’ command in MATLAB [122]. Now we determine the actual onset of the IED clusters as the valley immediately preceding the peak of an IED. We compute the difference or shift between the annotated IED cluster onset and the specified IED onset (sample point 16). Shift s for a cluster i is defined as $s = \text{Annotated}_{IED \text{ onset}} - \text{Actual}_{IED \text{ onset}} + 16$. Later the individual IED onsets are set as the annotated onset minus the calculated shift for the cluster it belongs to. The IEDs are re-extracted according to the new onset values. This realignment would make it easier for identifying the morphological characteristics of IEDs. The realignment process is illustrated in Fig. 4.3.1. This realignment process is reciprocated until all the IEDs are appropriately re-aligned.

The background waveforms or the non-IEDs were also extracted as 500-millisecond waveforms with an overlap of 75% from the non-epileptic EEG. We considered only non-epileptic EEG as backgrounds since there is a possibility of overlooked epileptiform events in the epileptic EEGs. There is a high probability that an IED at a particular channel could spread into the neighboring channels. Consequently, we applied all the annotated IEDs for training, while the testing performance was evaluated based on identifying IED time instants. Each fold contains about 90 non-epileptic EEGs; consequently, each testing set contains approximately 1.25 million background

events. The testing class imbalance ratio is approximately 1:500, which represents the real-world scenario. Some typical IEDs and backgrounds are illustrated in Fig 4.3.2.

Figure 4.3.1: Realignment process of IEDs during data extraction.

Figure 4.3.2: 500-millisecond (64-sample) waveforms of (a) IEDs and (b) non-IEDs extracted randomly from the MGH dataset (93 epileptic and 461 non-epileptic EEGs).

4.4 QUANTITATIVE CHARACTERISTICS OF IEDs

In this section, we explain how we perform the analysis of the different morphological characteristics of IEDs. We quantize the different IED characteristics, thereby attempting to provide a standard quantitative definition of IEDs. This definition could be further applied for identifying and eval-

uating new IEDs, thus advocating a generic IED definition to reduce the subjectivity amongst experts in visually identifying IEDs.

4.4.1 INTRODUCTION

Earlier attempts in the literature [47, 53] towards providing a quantitative definition is not reliable due to lack of adequately large IED database. Beyond its subjectivity, expert interpretation is often inefficient, particularly when EEG recordings last for hours or days at a time, opening the possibility of fatigue and missed events. In many problems studied to date, with larger datasets, it is often possible for “unreasonable effectiveness of data” to allow even relatively simple methods to substantially outperform more sophisticated methods [123]. Here we quantify the different characteristics of epileptics IEDs by analyzing the MGH dataset of 18,164 IEDs that were cross-annotated by two neurologists. We apply a semi-automated peak detection system to detect the different peaks of the IEDs 4.3. The different characteristics are estimated from the detected peaks. We have quantified the probability for the presence of slow waves associated with IEDs. These IED characteristics can be further applied for automated detection systems that would promote faster and reliable diagnosis of epilepsy.

4.4.2 METHODOLOGY

We identified the different transients of the IEDs with a peak detection system described in section 4.3. The peak detection system identifies the different onset and end points of the IED transients. First, we apply a low-pass filter (LPF) of 50 Hz in order to remove the high-frequency variations as well as to facilitate efficient peak detection. We set the cutoff as 50 Hz as an IED is defined as a transient with a duration of atleast 20 milliseconds by IFSCEN [48]. Next we identify the most prominent peak in the EEG segment along with troughs immediately preceding and succeeding the peak. We compute the duration of the transient and verify whether it lies in the range 20-200 milliseconds. If the duration is less than 200 ms (according to IFSCEN definition [48]), it is identified as the ‘IED’. If the duration is greater than 200 ms, it is identified as the ‘slow-wave’ discharge associated with the IED. Once we identify the prominent peak, the onset is set as the immediate ‘valley’ preceding the prominent peak and the endpoint is set as

the immediate ‘valley’ succeeding the prominent peak. The different characteristics of the IED are quantized based on these points. We also detect the presence of slow waves in the segment. This is performed computing the duration of all the peaks in the segment and examining if any of the duration’s are greater than 200 ms. If a slow wave is identified, the onset as well the endpoints are computed similarly. In case our peak detection system failed to identify the different points (onset, endpoint), the particular waveform was visually re-examined. If IEDs are still present in the segment, the onset and endpoint are identified manually. If we could not identify any presence of IEDs, the particular EEG segment was discarded from the analysis. An illustration of the peak detection system is presented in Fig. 4.4.1.

Figure 4.4.1: Peak detection system applied for extracting the morphological characteristics of IEDs.

We investigate the following morphological features of epileptic IEDs. We quantify the three sets of features: amplitude, duration, and slope features (see Fig. 4.4.2) [53]. The amplitude features are the peak voltage, rising voltage, and the falling voltage:

- **Peak voltage** is the maximum overshoot from the reference line.
- **Falling voltage** is the maximum peak to peak voltage fluctuation as a result of the IED.

- **Rising voltage** is the maximum amplitude fluctuation, during the onset of an IED.

The IED can be divided into three transients:

- **First half-wave** is the segment between the onset of the IED to the maximum overshoot peak.
- **Second half-wave** is the segment between the crest to the trough of the IED.
- **Slow wave** is the slow segment that follows an IED.

The two slope features are defined as:

- **Rising slope** is the slope of the first half-wave.
- **Falling slope** is the slope of the second half-wave.

Once the features are extracted, we remove the extreme outliers from the dataset by applying a threshold on the feature vector. We set the threshold as $mean \pm 3 \times standard\ deviation$ of the feature values in order to accommodate the IED variability. Further, we also evaluate the differences between spikes and sharp waves. According to the definition provided by IFSCEN [48], spikes and sharp waves differ in the duration of the pointed peaky-segment. The duration of spikes is between 20 to 70 ms, whereas the sharp waves exhibit a duration of 70 ± 200 ms 3.1.3. We divide the IEDs into spikes and sharp waves, and investigate their features.

Figure 4.4.2: Morphological characteristics of IEDs (a) Amplitude features: peak voltage, rising voltage, and falling voltage; (b) Duration of the IED transients: duration of first half-wave, second half-wave, and the slow-wave; (c) Slope features: rising slope and falling slope.

4.4.3 RESULTS

We extracted 500 millisecond IED waveforms after applying the clustering-based realignment process. The various morphological characteristics were extracted with a semi-automated peak detection system. The MGH IED database with a total of 14,170 annotated IED events consisted of 6.8% spikes and 93.2% sharp waves. We separated spikes and sharp waves based on the sum of the duration of the two half-waves (first and second half-waves). The histograms plots for the various morphological features of CAR montage IEDs are illustrated in Figs. 4.4.3 (amplitude features), 4.4.4 (duration features), and 4.4.5 (slope features). The range of quantified values for the different features is illustrated in Table 4.4.1. Further, the characteristics of spikes and sharp waves are presented in Table 4.4.2.

Figure 4.4.3: Distribution density plots of different amplitude features of IEDs: (a) Peak voltage; (b) Rising voltage; (c) Falling voltage. The mean values are highlighted in red.

Figure 4.4.4: Distribution density plots of the duration of the three IED transients: (a) Duration of the first half-wave; (b) Duration of the second half-wave; (c) Duration of the slow-wave. The mean values are highlighted in red.

Figure 4.4.5: Distribution density plots of the slope values of IEDs: (a) Rising slope values; (b) Falling slope values. The mean values are highlighted in red.

Table 4.4.1: Morphological characteristics of IEDs (reported as mean $\pm$ standard deviation).

Morphological Characteristics	Values
Peak voltage	$71.13 \pm 38.75 \mu\text{V}$
Rising voltage	$118.25 \pm 64.84 \mu\text{V}$
Falling voltage	$129.19 \pm 64.07 \mu\text{V}$
Duration of first half-wave	$45.64 \pm 14.23 \text{ ms}$
Duration of second half-wave	$62.38 \pm 17.18 \text{ ms}$
Duration of slow wave	$235.89 \pm 56.94 \text{ ms}$
Rising slope	$2.74 \pm 1.62 \mu\text{V/ms}$
Falling slope	$2.20 \pm 1.26 \mu\text{V/ms}$

Table 4.4.2: Morphological characteristics of IEDs separated into spikes and sharp waves (reported as mean $\pm$ standard deviation).

Morphological Characteristics	Spikes	Sharp waves
Peak voltage	$73.25 \pm 60.28 \mu\text{V}$	$71.20 \pm 37.48 \mu\text{V}$
Rising voltage	$98.72 \pm 80.47 \mu\text{V}$	$119.76 \pm 63.77 \mu\text{V}$
Falling voltage	$151.47 \pm 104.41 \mu\text{V}$	$128 \pm 61.72 \mu\text{V}$
Duration of first half-wave	$25.41 \pm 5.53 \text{ ms}$	$47.14 \pm 13.50 \text{ ms}$
Duration of second half-wave	$29.34 \pm 5.54 \text{ ms}$	$64.82 \pm 15.10 \text{ ms}$
Duration of slow wave	$274.54 \pm 50.71 \text{ ms}$	$235.06 \pm 56.57 \text{ ms}$
Rising slope	$4.12 \pm 3.41 \mu\text{V/ms}$	$2.69 \pm 1.54 \mu\text{V/ms}$
Falling slope	$5.27 \pm 3.74 \mu\text{V/ms}$	$2.06 \pm 1.08 \mu\text{V/ms}$

4.4.4 DISCUSSION

The following observations can be derived from the observed morphological characteristic results of IEDs:

- The falling amplitude of IEDs are greater than the rising amplitudes.
- The first transient of IEDs are generally shorter than the second transient.
- The rising slope is generally higher than the falling slope for sharp waves, whereas the falling slope is generally higher than rising slope for spikes.

We have also quantified the presence of slow-wave discharges associated with IEDs. We identified 44.44% of slow-wave discharges associated with IEDs (36.88% of spikes and 44.98% of sharp waves). We compared the amplitude of the slow-wave discharges with that of IEDs (peak voltage) and observed that only 30.32% of the slow-wave discharges are exhibiting higher amplitudes than IEDs. We compared the IEDs with the background immediately preceding and succeeding the IEDs. This would provide us with an estimate of how much an IED 'stands out' from the other EEG background waveforms.

We performed the comparison of amplitudes of IEDs (peak voltage) and the corresponding immediate backgrounds (maximum voltage). While considering the immediate preceding background waveform, we applied an offset of 250 milliseconds (32 samples) in order to prevent the capturing of IED duplicates, due to spatial spreading. We observed that 69.04% (in comparison with preceding background waveform) and 62.98% (in comparison with the succeeding background waveform) of the IED waveforms significantly stand out (with a threshold of 10%) from the background activity. The detailed results are presented in Table 4.4.3. This is the adapted comparison from Celesia et al. [47]. They conducted the study based on 600 IEDs and reported that 84.5% of the IED waveforms were significantly higher than the background activity (preceding the IED) with a percentage threshold of 30% (see Table 3.2.1). We have kept the inter-IED separation as 32 samples or 250 milliseconds (see Section 4.1.1). Consequently, there is a possibility that the preceding background waveforms may contain annotated IEDs. We have removed the IED instances that correspond to the annotated neighborhood IED. Table 4.4.4 tabulates the percent of IEDs that have been marked as similar to the background activity (the percentage threshold at 10%), that is having an annotated IED at its neighborhood. After removing the observations with neighborhood annotated IEDs and setting the percentage threshold as 30%, 63.5% of the IED instances stand out from the background waveforms immediately preceding the IEDs. This result corroborates with the previous observation by Celesia et al. [47].

Table 4.4.3: Comparison of the amplitude of IED (peak voltage) with the 500 milliseconds background activity (b.a.) preceding the annotated IED.

Background activity	Smaller or 10% higher than b.a.	11%-50% higher than b.a.	51%-100% higher than b.a.	101%-300% higher than b.a.	301%-500% higher than b.a.	501% higher than b.a.
Comparison with b.a. preceding the annotated IED	30.96%	16.98%	15.31%	28.51%	5.94%	2.3%
Comparison with b.a. succeeding the annotated IED	37.02%	17.34%	15.17%	24.07%	5.01%	1.39%

Table 4.4.4: Percentage of IEDs with annotated IEDs in the neighborhood that exhibit an amplitude difference 10% or lesser with the neighborhood waveforms.

Background activity (b.a.)	Percentage of IEDs with amplitude difference of 10% or lesser in comparison with b.a.	Percentage of IEDs with neighborhood annotated IEDs
Preceding b.a.	30.96%	7.9%
Succeeding b.a.	37.02%	24.52%

One potential limitation of this analysis is that even though the experts have annotated the segments as IEDs, they have not provided any distinction in labels between spikes and sharp waves. They are separated based on the duration of the pointed segment. However, only the presence and location of IEDs are relevant from a clinical perspective. Moreover, we have observed that spikes and sharp waves exhibit similar trends in morphological characteristics, except for the slope of the segments.

4.5 IED STATISTICS

4.5.1 IED VARIABILITY STUDIES

In order to develop an efficient system to identify IEDs, it is necessary to understand the variability of IEDs within a patient as well as across patients. We performed an intra-patient and inter-patient IED variability study based on the correlation. In the intra-patient study, the correlation values were computed for each of the IEDs, and the mean and standard deviation of the correlation values were recorded. Next for inter-patient studies, the correlation values were calculated across the IED sets of different subjects. The mean correlation heatmap for intra-patient and inter-patient variability are illustrated in Fig. 4.5.1. We observe that the IEDs from the same patients are moderately similar with a correlation value of 0.48 ± 0.28 . There are a few patients with a single IED annotation. These are removed from the analysis in order to provide a better approximation of intra-patient IED variability. Subsequently, the mean correlation values for intra-patient similarity are 0.43 ± 0.24 . On the contrary, the IEDs from different subjects are significantly variable with the mean correlation value of 0.21 ± 0.23 . This makes IED detection quite challenging.

Figure 4.5.1: The mean correlation value heatmap for intra-patient and inter-patient IED variability.

4.5.2 IED FREQUENCY ANALYSIS

Here we try to quantize the frequency or the rarity of the IEDs in comparison with the non-IED waveforms. We have attempted to annotate entire IED events during the evaluation procedure. There could have a few overlooked IEDs, but this evaluation would provide an estimate of the rarity of IEDs. Since we are considering an IED event as a 500-millisecond waveform, we divided the remaining EEGs portions into 500-millisecond segments. Next we compute the ratio of IEDs to the number of background segments for each patient. The mean ratio was observed to be approximately 1:650. The ratio for individual patients is illustrated in Fig. 4.5.2.

Figure 4.5.2: The ratio of IEDs vs. non-IEDs for different patients.

We have preprocessed, extracted, and analyzed the IEDs from the MGH dataset. A standard quantitative definition of the IED morphological characteristics was formulated. Further, we describe the different strategies for developing the epilepsy diagnostic system in the subsequent chapters.

“What we find changes who we become.”

Peter Morville

5

IED Detector

Identification of interictal epileptiform events is the foundation of our epilepsy diagnostic system. We have applied different machine learning and deep learning approaches for implementing a fast, reliable IED detection system. In this chapter, we describe the different steps in the development of the IED detectors, namely, data preparation and evaluation (Section 5.2.1), background rejection technique (Section 5.2.2), template matching (TM) IED detector (Section 5.2.3), other traditional classifier-based IED detectors (Section 5.2.4), and Convolutional Neural Network (CNN)-based IED detector (Section 5.2.5). Later the performance of different IED detectors is evaluated on the MUSC dataset along with Persyst in Section 5.3.5.

5.1 INTRODUCTION

Automated IED detection would enable faster and reliable epilepsy diagnostics and thereby a more effective treatment plan. Several methods [124] have been tested to develop reliable automated IED detection systems. Unfortunately, these methods have not been validated on a sizable dataset and consequently, are not universally accepted. We explore different machine learning systems available in the literature to develop an efficient IED detector. We explore the simplest template matching (TM) systems up to the latest deep learning techniques.

Today, fast and efficient TM systems [125] are available that can solve the problem of processing delays in the diagnostic process. The efficiency of the TM system primarily depends on the quality of the template library. To extract an effective IED template library, we applied clustering on the MGH IED database to extract a library of IED templates. In order to optimize this process, we also employed a distance measure along with fine-tuning of the clustering parameters. The TM system is developed with the extracted template library. Besides TM, we evaluate various traditional classification algorithms, namely, k -Nearest Neighbors (k -NN) [25], decision trees [126], Support Vector Machine (SVM) [27], Linear Discriminant Analysis (LDA) [26], and AdaBoost [28] (a popular boosting algorithm based on LDA, k -NN, and decision trees) for IED detection. Further, deep learning techniques are evaluated.

Deep learning techniques [29] are shown to provide higher classification efficiency in the literature. Moreover, deep networks can detect underlying structures or features from the raw data. This would help us to reduce the dependency on the feature extraction part. We analyze the performance of traditional classification algorithms with the newer, sophisticated deep learning methods of Convolutional Neural Network (CNN) [29]. CNN has been applied for various EEG-based studies: navigation in virtual environment (VE) [127], liquid crystal display (LCD) flickering frequency identification [128], classification of Steady-State Visual Evoked Potentials (SSVEPs) [129], and other Brain-Computer Interface (BCI) studies like P300 detection [130]. The studies investigating deep learning technique of CNN for IED detection is limited. Johansen et al. developed a CNN-based IED

detector in [97]. Le et al. have proposed a Deep Belief Network (DBN) for detecting IEDs [91]. Tjepkema-Cloostermans et al. had developed a CNN based IED detector which is shown to have achieved a sensitivity of 47.4% for a sensitivity range of 98-99% on a dataset of 100 subjects (with 1815 annotated IEDs) [90]. All the systems have been evaluated only on a small dataset. Therefore, the performance results are unreliable. Moreover, no clinical validation on multiple datasets has been performed.

In the literature, there is a mention of a relatively new background rejection technique. One major issue in evaluating EEG is the bulkiness of the data. This makes the analysis, tedious and time-consuming. This technique aims to reduce the bulk of the EEG dataset by eliminating the non-IED waveforms. After the rejection, sophisticated and highly accurate techniques are applied to the remaining miniature dataset. Bagheri et al. has demonstrated the method of background rejection in [131]. The feature set consisted of morphological features, Non-Linear Energy Operator (NLEO) values, Discrete Wavelet Transform (DWT) coefficients, Continuous Wavelet Transform (CWT) coefficients, and frequency band features. The testing was carried out on 100 patients by ten selected features, by thresholding. The results reported indicated that 98.65% of the background was rejected while preserving 90.6% of the IEDs. The fact that the system was tested on the same dataset used for feature extraction is a concern. Moreover, 9.4% of IEDs lost might be a problem in patients with low IED count. Inspired by the above work, we have designed a CNN-based background rejection system which is aimed towards improving the efficiency of our IED detector.

We analyze the accuracy of the various classification algorithms by performing 5-fold cross-validation on the MGH dataset of 545 subjects (84 epileptic EEGs with 18,164 annotated IEDs and 461 non-epileptic EEGs) recorded at Massachusetts General Hospital (MGH), Boston. Various performance matrices are evaluated: receiver operating characteristics (ROC) curve, Area under ROC curve (AUC), precision-recall (PR) curve, Area under PR curve (AUPR), and false-positive (FP) rate.

5.2 METHODOLOGY

5.2.1 DATA PREPARATION AND EVALUATION

The different IED detectors are trained and evaluated at the waveform-level, i.e., each IED detector is trained to identify if a single-channel EEG segment is an IED or background. We set the segment size as 500 milliseconds (1×64 samples) as detailed in Section 4.3. The set of extracted IEDs and backgrounds are applied to train the classifier-based IED system. We perform 5-fold cross-validation to evaluate data invariance of our system. Cross-validation or out-of-sample testing is a standard technique for evaluating the generalization performance of a system. It is performed by training several models on subsets of the available input data and evaluating them on the residual subset of the data. For example, consider k -fold cross-validation. Initially, the data is divided into k subsets, $k - 1$ subsets are applied for training, and the remaining subset is employed for testing. Later the performance indices are averaged over the k folds. We divide the entire dataset into 5-folds by keeping the IED distribution consistent over the different folds. The details regarding the different folds are presented in Table 5.2.1.

Table 5.2.1: Distribution of the MGH dataset over different cross-validation folds.

Fold no.	No. of Epileptic EEGs	No. of IEDs events	No. of annotated IEDs	No. of non-epileptic EEGs	No. of background event segments ($\times 10^6$)
1	18	2920	4077	93	1.34
2	19	2757	3571	92	1.47
3	18	2831	3207	92	1.47
4	19	2781	4021	92	1.37
5	19	2881	3288	92	1.23
Total	93	14170	18164	461	6.88

We also divided the dataset with the age distribution (see Fig. 5.2.1) and gender (male-female) ratio approximately the same over the different folds. A two-sample t -test was performed on the age distribution across the five folds with the null hypothesis that the two vectors came from independent

random samples of a normal distribution with equal means and equal but unknown variances. The null hypothesis was accepted across the different combination of folds with 5% significance level. This suggests that the different folds are a representation of the entire database. We have also made sure that the different recordings from the same patient are accommodated in the same fold, to prevent skewing.

Figure 5.2.1: Patient age distribution over the different folds. The mean is indicated in red.

Once the classifiers are trained at the waveform-level, we obtain 19 outputs (different EEG channels) for a single time-instant. Next we combine these 19 outputs to produce a single time-instant output. We combine the outputs of the different channels by computing the maximum output. When this maximum output is above a threshold, an IED is detected at the time instance at hand. This scheme is a simple way to combine information from all channels to detect IEDs. Thereby the system predicts the IED time-events. The entire block diagram of the proposed approach is illustrated in Fig. 5.2.2 (training) and Fig. 5.2.3 (evaluation).

Figure 5.2.2: Block diagram of the proposed IED detector training scheme.

Figure 5.2.3: Block diagram of the proposed IED detector evaluation scheme.

The comparison between different IED detectors is evaluated based on the performance measures listed in Section 2.5. The performance of the different classifiers is initially performed based on area-related measures (AUC and AUPRC) since the different classes are heavily imbalanced. The typical IED to background ratio is around 1:650 (see Section 4.5.2). At times, area-related measures also could be misleading. For class-imbalanced classification, the area on the left quadrant has more relevance compared to the area on the right quadrant. Consequently, we evaluate the precision values for fixed sensitivity values along with AUC and AUPRC. We validate the effectiveness of our IED detectors on the MUSC dataset. The MUSC dataset contains IEDs that are cross-annotated by 18 annotators, making it ideal for comparing the performance of different IED detector. We evaluate the traditional classifier-based IED detectors, proposed CNN-based IED detector, and Persyst [96]. We apply the Persyst annotations employed by Halford et al. in [113]. Persyst is one of the clinically validated, commercially available EEG software solutions for identifying IEDs, seizures, and other EEG trends. We also compare the performance of the proposed system in comparison with the consensus of the experts with the MUSC dataset. The traditional classifier-based IED detectors are evaluated using MATLAB 2017b [122] and the CNN IED detector is implemented with Tensorflow 1.2 [132], Python v3.5 on a K80 Tesla GPU. We applied Tensorflow as it has more deployment or production support.

5.2.2 BACKGROUND REJECTION

In our study, we develop an efficient background rejection system based on CNN. It is evident from the literature that a high percentage of the EEG are background patterns with very few rare or interesting patterns. Most of the background waveforms are predictable or easily distinguishable from the IEDs. This system would help to reduce the misclassification rates as well as help to provide a challenging or efficient background set for training the classifiers such as CNN where the training depends on backpropagation. We applied CNN as it has shown superior performance in the literature. We applied a simple architecture for the background rejection CNN (see Table 5.2.2). The training was performed in batches containing an equal number of IED segments and background segments randomly chosen from

the extracted MGH dataset. Each batch was alternated after every ten iterations. The training was performed until a training accuracy of 99.99% was achieved so that the system is typically overfitted on the predictable background waveforms. Later the entire MGH background set is evaluated with the system, and the challenging False Positive (FP) background waveforms are extracted by setting a threshold of 0.99. The FP background set was further applied for CNN IED detector optimization.

Table 5.2.2: Parameter values of the CNN network implemented for background rejection.

Parameters	Value
Number of convolution layers	1
Number of pooling layers	1
Number of fully connected layers	1
Number of convolution filters	32
Dimension of convolution filters	1×5 each
Number of hidden layer neurons	100
Drop probability	0.5
Activation function	ReLu

5.2.3 TM IED DETECTOR

In order to make IED detection faster and reliable, we proposed the development of a TM system. The performance of a TM system primarily depends on the quality of the template library it maintains. Developing a highly effective template library is crucial. We cluster the IED waveforms based on the raw temporal characteristics to extract the suitable templates. To choose the optimal clustering algorithm, we explore different clustering techniques with the MGH IED database, and the selection is performed based on the clustering evaluation indices. We investigate five different clustering algorithms: K-means clustering [14], K-medoids clustering, agglomerative hierarchical clustering [15], Fuzzy C-means clustering (FCM) [16], and affinity propagation (AP) [17]. AP is an efficient clustering technique that has been applied to traffic network analysis, audio pattern recognition, gene pattern detection, etc. All the algorithms except AP have been previously investigated in the literature for grouping IED [57, 66, 80, 133]. The clustering results are compared using four measures: S_Dbw validity index [19], Davies-Bouldin

(DB) index [20], Silhouette Value [21], and Mean Squared Error (MSE). A similar selection of evaluation of measures was explored in [22]. The selection of a distance measure, capturing the similarity between the IED waveforms appropriately, is crucial for the selection of proper templates. We investigated four different distance measures: Euclidean distance (ED), Pearson’s correlation coefficient [9], Dynamic Time Warping (DTW) [10], and Mahalanobis distance [11]. DTW is a robust non-linear method to find the optimal match between temporal sequences. DTW has been applied for face detection [134], TM [135], and many other applications. DTW has also been demonstrated to enhance the performance of various clustering algorithms such as K-means clustering, hierarchical clustering, and fuzzy clustering [136–138]. The other three measures have been traditionally applied to find the similarity amongst EEG waveforms [99, 139]. We evaluated the effectiveness of the distance measures by applying t-Distributed Stochastic Neighbor Embedding (t-SNE). After the selection process, we optimize the combination of the clustering technique and the distance measure by fine-tuning the parameters. We apply the optimized combination to cluster the MGH IED database. Once the IED clusters are obtained, we pool the cluster centers together to form the template library. Therefore, the template library will provide a representation of the entire IED database. Finally, we apply the template library as the look-up table for the proposed template matching system. The entire approach is summarized in Fig. 5.2.4.

Figure 5.2.4: The flow diagram of the proposed method for implemented an efficient TM IED detector.

We evaluated the performance of the five clustering techniques on the MGH IED database. Each IED is considered as a 64-point vector, and the entire IED database is considered as the input dataset for clustering. The steps involved in selecting the clustering technique (see Fig. 5.2.5) are as follows:

1. We apply AP to the MGH IED database. The similarity measure is set as the negative of ED, and the priority values are set to the median of the similarity values. The number of clusters K is determined by the algorithm based on the density of the dataset.
2. We evaluate the other clustering algorithms with ED as the distance measure and with the same number of clusters K that we obtained with AP.
3. For K-means, FCM, and agglomerative clustering the cluster center is generally chosen as the mean of the member data points rather than a member data point. We map the cluster centers to the nearest cluster member in order to have a fair comparison with the other clustering techniques. In AP and K-medoids the cluster center is chosen from the data points.
4. We iterate the clustering process 100 times for each of the techniques, and the evaluation metric values are averaged.

We evaluate the distance measures based on their ability to capture the similarity amongst IEDs. DTW is a non-linear measure, and it causes deformation in the input space to capture the better correlation between the cluster members. Moreover, DTW does not obey triangular inequality [140]. Therefore, the traditional indices like the DB index, S_Dbw index cannot be employed for comparing DTW with other distance measures. We analyze the performance of the four distance measures with the visualization technique, t-SNE. A subset of IEDs and non-IEDs are taken, and 2-dimensional density plots are constructed by applying t-SNE with the four distance measures. The grouping of IEDs and their separation from non-IEDs are investigated. The entire steps are illustrated in Fig. 5.2.6. We optimize the combination of the clustering technique with the distance measure by fine-tuning the parameters.

Figure 5.2.5: Different steps involved in the selection of a clustering technique.

The final evaluation of the templates is carried out by testing them for detecting IEDs on the entire MGH dataset. We perform 5-fold cross-validation on the MGH dataset. For each step, the number of clusters K was derived from the AP algorithm based on the density of the data. For this, we set the initial priorities as the median of the similarity values. In traditional TM systems, only the templates of the patterns of interest are employed. In other words, only IED templates are employed in conventional template-based IED detectors. Here, we apply a combination of IED and background templates, as illustrated in Fig. 5.2.7.

The final decision is based on the minimum distance from the combined set of templates, following certain decision-making rules. Consider a waveform x to be classified. Let S_i and B_i be the IED and background template sets, respectively. The decision rules for each system is described in Table 5.2.3. Decision rule 1 is the traditional template matching rule, based purely on the IED templates. In decision rule 2, we incorporate the background templates. In rule 2, the distance from the background templates is integrated as an additive term. The background waveforms would be more similar to the background templates in comparison with the IEDs. Consequently, the background template set further helps to increase the separa-

Figure 5.2.6: Different steps involved in the selection of a distance measure.

Figure 5.2.7: Basic structure of the proposed template matching (TM) system. Background templates are integrated to improve the detection accuracy.

tion between the IEDs and backgrounds. Here β are the weight parameters and $d(x, S_t)$ is the distance measure between the given waveform and the IED templates. We set β to 0.5 for ED and 0.25 for correlation values in our analysis. These thresholds could be optimized in the future. We develop two different systems based on decision rules. The different systems are

evaluated based on performance indices.

Table 5.2.3: Decision rules for the different template matching systems.

Decision rule 1	$\min\{d(x, S_t)\}$
Decision rule 2	$\min\{d(x, S_t)\} - \beta \times \min\{d(x, B_t)\}$

5.2.4 TRADITIONAL CLASSIFIER-BASED IED DETECTORS

We evaluate various traditional classification algorithms, namely, decision trees (Classification and Regression Tree (CART)) [24], k -NN [25], LDA [26], SVM [27], and AdaBoost [28] (a popular boosting algorithm based on LDA, k -NN, and decision trees) for IED detection. The different parameters of the classifiers, CART, K-NN, LDA, SVM, and ensemble methods were optimized by applying Bayesian optimization [141]. Bayesian optimization attempts to minimize an objective function in a confined range of parameters. Here the objective function is the loss or error produced by a classifier. The hyper-parameters of the classifier are the input set applied for optimization. The loss for different parameter sets is evaluated by applying 5-fold nested cross-validation on the training set. The best set of parameters are chosen based on the loss values. The Bayesian optimization algorithm can be stochastic, i.e., in each run, the optimized parameter values could be slightly different. Consequently, there could be a slight variation for the IED detector outputs for the different iterations. While training the IED detector, we applied balanced training, i.e., the IED to background ratio is maintained as 1:1. This was applied to prevent negative class overfitting due to the class-imbalance. The different classifiers and the range of parameter values optimized are detailed in Table 5.2.4.

Table 5.2.4: Different parameter values optimized for the traditional classifiers.

Key: N : Number of observations

Classifier	Parameters	Range
Classification and Regression Tree (CART)	Minimum number of leaf node observations Maximal number of decision splits Split criterion	$[1, \max(2, \lfloor \frac{N}{2} \rfloor)]$ $[1, \max(2, N - 1)]$ Gini's diversity index, Cross entropy
K-Nearest Neighbors (K-NN)	Distance measure Minkowski distance exponent Number of neighbors k	Euclidean, Minkowski $[0.5, 3]$ $[3, 100]$
Linear Discriminant Analysis (LDA)	Discriminant Type Delta Gamma	Linear $[10^{-6}, 10^3]$ $[0, 1]$
Ensemble of classifiers	Ensemble aggregation method Weak learners Number of weak learners Learning rate	AdaBoost K-NN, LDA, Decision Tree (CART) $[10, 500]$ $[10^{-3}, 1]$
Support Vector Machine (SVM)	Kernel Box constraint Kernel scale parameter	Linear, Gaussian $[10^{-3}, 10^3]$ $[10^{-3}, 10^3]$

5.2.5 CNN IED DETECTOR

The CNN was also evaluated according to the 5-fold cross-validation (see Section 5.2.1). Similar to the traditional classifiers, we apply balanced training. We also optimized the IED training process by improving the quality of the background set applied for training. The entire background set contains approximately 99% of easily separable waveforms. Even though the challenging background waveform proportion is relatively small, it accounts for approximately 100 false detection instances per 30-minute EEG recordings. This makes the identification of epileptic patterns in EEG challenging, especially the EEGs with few IED count. While training the IED detector, since we are applying balanced training, the challenging background waveforms always become a minority. This results in a system capable of addressing only predictable backgrounds. Prior investigations, however, disregard this fact and have reported favorable IED detection results. Instead we apply the FP backgrounds extracted by the background rejection system to improve the performance of the CNN IED detector.

We also employ different techniques to optimize the CNN network and to prevent overfitting, namely, dropout and batch processing. In dropout, the weights of a certain percentage of the fully connected layer were dropped at each iteration to prevent overfitting. We set the dropout probability as 0.5. The two classes, IEDs and backgrounds events, are heavily skewed. The typical IED to background ratio is typically 1:500 at the waveform-level. In batch processing the ratio is maintained as 1:1, but the training set is each time replaced after a particular number of iterations. This training scheme yields better sensitivity and precision for detecting IEDs. At each step, we randomly choose a subset of IEDs and backgrounds for training. The output of the CNN-based was mapped into $[0,1]$, i.e., for each 500-millisecond input, the CNN produces an output between $[0,1]$ with a higher value indicating the spikiness of the waveform. The flow diagram of the proposed CNN approach is illustrated in Fig. 5.2.8.

Figure 5.2.8: The proposed CNN architecture with a single convolutional, pooling, and fully connected layer applied in this study. N_f indicates the number of convolutional filters, D_f the dimension of convolutional filters, and N_{fc} the number of neurons in the fully connected layer.

The hyperparameters of CNN and the training termination criteria were optimized in a brute force manner with nested cross-validation on the training dataset: 75% for training and 25% for validation. The CNN training termination criteria are dependent on the validation error. As a result, we need a separate validation set in order to evaluate the different sets of hyperparameters. Consequently, we divide the validation set into two: One for training termination criteria and the other for the hyperparameter selection. For each fold, the system was trained until there is no significant improvement in the validation error (by evaluating validation set 1) for at least five iteration steps. Later we evaluate the performance of the trained CNN on the validation set 2. This process is repeated for the different set of CNN hyperparameters and the set that produces the best performance is chosen as the optimized hyperparameter set. Later the entire training data is applied for training the final CNN network. Here the CNN training is performed until the maximum number of iterations. At each step, the trained network is saved. Finally, the network is retrieved that has produced the lowest validation error. The testing is performed by applying this network on the testing data fold. The proposed training and validation flow are illustrated in Fig. 5.2.9 and the different parameters of the CNN optimized are detailed in Table 5.2.5.

Table 5.2.5: Different sets of parameter values evaluated for CNN optimization.
Key: n_s : Number of IEDs

Parameters	Values
Number of convolution layers	1, 2, 3, 4
Number of pooling layers	1, 2, 3
Number of fully connected layers	1, 2, 3
Number of convolution filters	4, 8, 16, 32, 64
Dimension of convolution filters	1×3 , 1×4 , 1×5 , 1×6 , 1×7 , 1×8 each
Number of hidden layer neurons	100, 500, 1000
Activation functions	ReLu, tanh, sigmoid
Dropout probability	0.5
Size of the batch processing	$2n_s$, n_s , $\frac{n_s}{2}$, $\frac{n_s}{4}$
Maxium number of iterations	20000

Figure 5.2.9: The proposed training and validation steps for trial 1 of the 5-fold cross-validation.

5.3 RESULTS

5.3.1 BACKGROUND REJECTION

We successfully developed the CNN-based background rejection system. It was trained on the MGH IED set, and backgrounds randomly sampled from the background set. The results for background rejection with the threshold of 0.99 are illustrated in Table 5.3.1. The background rejection system separated approximately 1% of the background waveforms, with less than 1% loss in sensitivity. The performance of the background rejection system is superior to the different methods suggested in the literature [131]. Later the extracted FP backgrounds were applied for improving the performance of the various IED detectors.

Table 5.3.1: Results for performing background rejection with a threshold of 0.99.

Waveforms	Total number	Percentage
Background waveforms	19 channels * 6.88 million	-
IED waveforms	18164	-
IED waveforms with CNN prediction greater than 0.99	18031	99.27%
Background waveforms with CNN prediction greater than 0.99	0.4 million	0.003%

5.3.2 TM IED DETECTOR

Our first objective was to determine the best algorithm for clustering epileptic IEDs. Five clustering algorithms (K-means, K-medoids, AP, FCM, and agglomerative clustering) were tested with the MGH IED database. Initially, we implemented affinity propagation on the MGH IED database with the negative of the ED as the similarity measure and the input priorities set as the median of the similarity values. The number of clusters K was obtained as 1381. With ED as the distance measure and $K = 1381$, the other clustering algorithms were evaluated on the MGH IED database. The results of cluster validation indices are shown in Table 5.3.2. In FCM, as we assigned

each data point to the cluster to which it has the highest membership value. In this process, we obtained a few empty clusters (about 10%) that we discarded. The positive silhouette values for the clustering techniques suggests the cohesion amongst the cluster members. The mean silhouette values are positive and high for all the clustering algorithms, except for FCM. The negative silhouette value FCM is mainly attributed to the empty clusters produced while mapping of the data points to the cluster to which it has the highest membership value. The MSE and S_Dbw index values are the lowest for affinity propagation. AP also has the second smallest value of DB index. This establishes the superiority of AP over other clustering techniques. This suggests that AP has effectively grouped the IEDs into clusters. Additionally, the number of clusters is also determined by the algorithm based on the density of the dataset.

Table 5.3.2: Results of clustering by employing the four validation index values.

Validation criteria	AP	K-means	K-medoids	FCM	Agglomerative clustering
Mean Silhouette value	0.0219	0.0755	0.0648	-0.626	0.0662
Davies-Bouldin index	1.287	1.395	1.129	3.742	1.306
S_dbw	0.1609	.2635	.2050	.1903	.2145
MSE ($\times 10^4$)	1.543	1.884	1.642	15.268	1.795

Next the four distance measures (ED, DTW, correlation coefficient, and Mahalanobis distance) were tested and evaluated based on t-SNE density plots. The perplexity parameter of t-SNE gives an effective number of neighbors. We fixed the perplexity parameter to half the number of input IEDs while evaluating the distance measures. Fig. 5.3.1 shows the density plots of t-SNE between IEDs and non-IED waveforms based on the four different distance measures. It is observed that the IEDs are well separated from the non-IEDs or the background waveforms. We can see that the four different distance measures have been observed to group the IEDs effectively. The density plot of IEDs exhibits high concentration at a particular region for all the four measures, whereas in the case of ED and Mahalanobis distance, the IED nuclei are slightly scattered (see Figs. 5.3.1a, 5.3.1c, 5.3.1g, and 5.3.1h).

While comparing the separability of the IEDs and backgrounds, there is a significant overlap between the plots for Mahalanobis distance (see Figs. 5.3.1g and 5.3.1h). DTW also has exhibited more overlap between IED and background waveforms in comparison with ED and correlation coefficient (see Figs. 5.3.1c and 5.3.1d). This is a legitimate challenge for DTW: while it identifies patterns more efficiently, it could account for more false positives as well. ED and correlation coefficient has demonstrated the superior separation between the IEDs and background waveforms. Therefore we apply and compare both the correlation coefficient and ED as the distance measure in the TM matching system. Since we already have the annotated IEDs, it is justified to apply DTW as the distance measure while applying clustering to extract the template library. This would help us to eliminate redundant patterns more efficiently. A similar comparison study with the four different distance measures has been implemented in [142] for clustering building energy patterns by applying clustering balance and clustered-vector balance techniques. DTW distance has been shown to define the “important clusters” effectively than other distance measures [142].

(a)

(b)

(c)

(d)

Figure 5.3.1: T-SNE visualization of IEDs and non-IED waveforms : (a) Density plots of IEDs for ED; (b) Density plots of non-IEDs for ED; (c) Density plots of IEDs for DTW; (d) Density plots of non-IEDs for DTW; (e) Density plots of IEDs for correlation coefficient; (f) Density plots of non-IEDs for correlation coefficient; (g) Density plots of IEDs for Mahalanobis distance; (h) Density plots of non-IEDs for Mahalanobis distance.

We developed an effective template library by applying AP-DTW on the MGH IED database. We implemented a basic TM system with the extracted template library as the look-up table. The system was tested on the whole MGH database. The decision of classifying a waveform as an IED was made based on the minimum distance from all the templates in the library. We compared the correlation coefficient and the ED as the distance measures for the TM system along with different decision rules. The comparison of the different distance measures and decision rules based on AUC are presented in Table 5.3.3. Correlation distance was observed to be the effective distance measure in comparison with ED based on AUC and AUPRC values. The decision rule 2 was established to be the superior to decision rule 1, for both the distance measures. The density of correlation distances (i.e., one minus correlation values) for IEDs and backgrounds are illustrated in Fig. 5.3.2 and Fig. 5.3.3 for TM systems with decision rules 1 and 2, respectively. It is observed that the two different classes are moderately separable. The overlap between the two classes was less for the system with decision rule 2 in comparison with the system with decision rule 1. This proves the effectiveness of the integration of background templates in the template matching system.

Figure 5.3.2: Density plots of minimum correlation distance values from template library for IEDs and backgrounds based on rule 1.

Figure 5.3.3: Density plots of minimum correlation distance values from template library for IEDs and backgrounds based on rule 2.

Table 5.3.3: AUC results for the TM IED systems with different distance measures and different decision rules.

Fold number	AUC values for Decision Rule 1	AUC values for Decision Rule 2
Distance measure: ED		
1	0.628	0.859
2	0.290	0.363
3	0.231	0.451
4	0.143	0.445
5	0.325	0.610
Mean	0.323 ± 0.187	0.545 ± 0.197
Distance measure: Correlation coefficient		
1	0.970	0.979
2	0.947	0.955
3	0.959	0.955
4	0.947	0.952
5	0.968	0.977
Mean	0.958 ± 0.011	0.963 ± 0.013

We perform 5-fold cross-validation in order to assess the TM IED detector. The results are illustrated as ROC (see Fig. 5.3.4c), PRC (see Fig. 5.3.4a), and sensitivity-false positive rate curves (see Fig. 5.3.4b). The 5-fold cross results for the optimized TM system are summarized in Table 5.3.4. The TM system was found to be useful in identifying IEDs with an average false detection rate of 11.10 per minute and 3.95 per minute for sensitivity thresholds of 90% and 80%, respectively.

Table 5.3.4: 5-fold cross-validation results for TM IED detector system with correlation distance as the distance measure.

Fold No.	No. of IED templates	No. of background templates	AUC	AUPRC	FP rate (per minute) for sensitivity of 90%	FP rate (per minute) for sensitivity of 80%
1	664	1160	0.979	0.575	5.9	1.7
2	659	1227	0.955	0.342	13.7	5.2
3	686	1171	0.955	0.471	15	6.3
4	675	1157	0.952	0.464	15.4	4.6
5	642	1212	0.977	0.440	5.5	1.95
Mean	-	-	0.963	0.458	11.10	3.95
Std	-	-	0.013	0.083	4.97	2.03

Figure 5.3.4: Performance curves of the TM system with decision rule 1 (blue) and decision rule 2 (red): (a) PRC; (b) sensitivity and false positive rate curves; (c) ROC.

5.3.3 TRADITIONAL CLASSIFIER-BASED IED DETECTORS

The AUC and AUPRC results for 5- fold cross-validation for the different classifiers are presented in Table 5.3.5. The AUC values are quite high for all the classifiers. This is mainly attributed towards the class imbalance. Moreover, there could be slight variations in the AUC values depending on the threshold width we apply to calculate the ROC. In such scenarios, the area under the PR curve is more meaningful compared to AUC values. Consequently, we discard the classifiers that have achieved inferior AUPRC values (value less than 0.5) from further analysis. The IED detectors based on Ensemble CART and SVM with Gaussian kernel provided the best performance indices. We explored the other performance measures of these two IED systems, and the results are detailed in Table 5.3.6. The SVMG system and the ensemble CART system performed superior to the other traditional classifier systems, with an average false detection rate of 4.72 per minute and 2.55 per minute, respectively, for a sensitivity threshold of 80%.

Table 5.3.5: AUC and AUPRC results for the various traditional classifier-based IED detectors.

Method		AUC Mean±Standard Deviation	AUPRC Mean±Standard Deviation
SVMG		0.988±0.007	0.615±0.19
SVML		0.954±0.04	0.177±0.10
CART		0.856±0.026	0.376±0.08
<i>k</i> -NN		0.925±0.034	0.499±0.01
LDA		0.973±0.0121	0.159±0.07
ENSEMBLE	CART	0.991±0.0006	0.730±0.103
	<i>k</i> -NN	0.908±0.034	0.122±0.01
	LDA	0.971±0.008	0.431±0.13

Figure 5.3.5: Performance curves of the traditional classifier-based (SVMG in blue and ensemble CART in red) IED detector: (a) PRC; (b) sensitivity and false positive rate curves; (c) ROC.

Table 5.3.6: 5-fold cross-validation results of traditional classifier systems.

Fold No.	AUC	AUPRC	Precision for sensitivity of 90%	Precision for sensitivity of 80%	FP rate (per minute) for sensitivity of 80%	FP rate (per minute) for sensitivity of 90%
Traditional classifier: SVMG						
1	0.982	0.768	0.166	0.482	0.958	5.042
2	0.988	0.8394	0.054	0.092	7.64	15.17
3	0.981	0.541	0.025	0.054	12.91	31.57
4	0.995	0.520	0.149	0.281	1.83	4.57
5	0.997	0.855	0.382	0.767	0.25	1.50
Mean	0.988	0.615	0.155	0.335	4.72	11.57
Std	0.007	0.19	0.140	0.295	5.43	12.30
Traditional classifier: Ensemble CART						
1	0.984	0.799	0.250	0.705	0.372	3.00
2	0.993	0.618	0.096	0.218	2.78	8.21
3	0.984	0.642	0.029	0.079	8.52	27.20
4	0.997	0.729	0.219	0.449	0.878	2.86
5	0.997	0.862	0.342	0.783	0.229	1.79
Mean	0.991	0.730	0.187	0.447	2.55	8.61
Std	0.006	0.103	0.124	0.303	3.48	10.68

5.3.4 CNN IED DETECTOR

We developed a CNN based IED detectors and evaluated it based on the 5-fold cross-validation on the MGH dataset. The illustration of CNN training and validation is presented in Fig. 5.3.6. The results for 5-fold cross-validation for CNN are presented in Table 5.3.7. The different performance curves of the different folds are illustrated in Fig. 5.3.7. The system has achieved a superior average false detection rate of 0.92 per minute and 0.2 per minute for sensitivity thresholds of 90% and 80%, respectively.

Figure 5.3.6: Variation in training and validation indicators (accuracy and loss) with the increase in the number of iterations of CNN training: (a) accuracy; (b) loss. The training termination point is highlighted.

Table 5.3.7: 5-fold cross-validation results of CNN IED detector system.

Fold No.	AUC	AUPRC	Precision for sensitivity of 90%	Precision for sensitivity of 80%	FP rate (per minute) for sensitivity of 80%	FP rate (per minute) for sensitivity of 90%
1	0.999	0.771	0.623	0.691	0.4	0.62
2	0.990	0.878	0.634	0.923	0.085	0.52
3	0.993	0.869	0.531	0.863	0.13	0.73
4	0.993	0.833	0.702	0.79	0.19	0.34
5	0.978	0.843	0.279	0.84	0.18	2.4
Mean	0.989	0.838	0.55	0.82	0.2	0.92
Std	0.006	0.040	0.16	0.08	0.12	0.83

Figure 5.3.7: Performance curves of the CNN IED detector: (a) PRC; (b) sensitivity and false positive rate curves; (c) ROC.

5.3.5 EVALUATION ON MUSC DATASET AND PERSYST

We evaluated the different IED detector on the MUSC dataset. In terms of annotations, the MUSC dataset has the highest number of cross-annotated IEDs. However, the number of IEDs are quite low, considering the fact that IED morphology is highly variable across patients. The results for the evaluation of our proposed system on the MUSC dataset (Phase 2) are presented in Fig. 5.3.8. In MUSC dataset, each neurologist has annotated the IEDs on a scale from 1 to 5, with 5 indicating more confidence that the segment is an IED. We have evaluated the results for average IED annotator consensus (average of 18 annotators) of 3 and 4 (see Fig. 5.3.8(a) and Fig. 5.3.8(b), respectively). The TM and the traditional classified based systems were observed to be highly inferior to the proposed CNN system and Persyst. The proposed CNN system outperforms Persyst according to the AUPRC plots.

Figure 5.3.8: Precision-recall comparison plots of our proposed CNN system with Persyst and traditional classifier systems on the MUSC dataset: (a) average annotator consensus of 3; (b) average consensus of 4.

We also compared the three IED detectors (CNN, SVMG, and Ensemble CART) with the consensus of the annotators on the MUSC dataset. Here the ground truth is set as annotator consensus of at least 4 on the Phase 2 MUSC dataset annotation. The performance of each expert is computed based on the Phase 1 MUSC annotation. The results for the evaluation with mean annotator consensus of 4 is presented in Fig. 5.3.9 along with the individual annotator performance. The proposed IED detector was observed to be non-inferior to most of the annotators.

Figure 5.3.9: Comparison of the proposed CNN IED detector and traditional classifier-based IED detectors based on the precision-sensitivity curve on the MUSC dataset with mean annotator consensus threshold of 4. The performance of experts is indicated as circles (green).

5.3.6 FEATURE EVALUATION

Finally, we evaluate the correlation of IED features with the CNN outputs. CNN's have been proven to learn sophisticated features without supervision. Since we are applying the raw time-series EEG waveforms as the features, this would provide a machine perspective of IEDs. We investigate the common morphological features of IEDs, i.e., amplitude, duration, and slope features[53]. We applied the Pearson correlation test and verified that the correlation values are significant with a confidence level of 1%. The correlation value plot for the different features is illustrated in Fig. 5.3.10. The correlation values for amplitude and slope features are positive. This suggests that the output of the CNN IED detector will be higher for waveforms with high amplitudes and the steep rise and fall slopes. The correlation values for the duration features were negative, i.e., IEDs are generally of shorter duration. We observe that the rising slope is the most consistent feature of an IED, as it has the most substantial average correlation value and the lowest associated standard deviation.

Figure 5.3.10: Correlation values of the different IED characteristics with the CNN output values.

5.3.7 COMPUTATIONAL TIME

Our proposed CNN IED detector takes approximately 10-15 seconds for evaluating a single 30-minute pre-processed and sampled 19-channel routine EEG recording sampled at 128 Hz. The preprocessing and sampling modules take up about 60-70 seconds for preparing the data for the IED detector. As a result, the CNN IED detector needs approximately 85 seconds for evaluating a single 30-minute routine EEG recording. An expert takes approximately 10 minutes to review a 30 minute routine EEG with approximately 19 electrodes. Comparing the time cost values, the proposed IED detector performs superior to the experts in both time-saved and making the diagnostic process faster. The evaluation was performed with an Intel (R) Xeon(R) CPU E5-2630 v4 @2.2 GHz CPU using Python v3.5. We are yet to apply the optimization of the system code. Moreover, the system could be employed in GPU settings, instead of CPU. This would make the system capable of processing a massive amount of data expeditiously.

5.4 DISCUSSION

We have developed an efficient, automated IED detection system employing CNN. We also developed a TM IED detector based on a high-quality template library. The template library was developed by clustering a sizable database of IEDs with affinity propagation and negative of the DTW distance as the similarity measure. Amongst the traditional classifiers, SVM with Gaussian kernel and AdaBoost CART performed better than other classifiers. The different IED detector systems were evaluated by performing 5-fold cross-validation with the MGH dataset of 545 subjects. The mean cross-validation results show the superiority of CNN-based system over the other systems. It has achieved a false positive rate of 0.2 and 0.92 per minute for the sensitivity threshold of 80% and 90%, respectively. We also compared AUC and AUPRC values for the different systems. Due to the high imbalance nature of the two classes, the AUC values are quite inconclusive, and we opted AUPRC values for comparison. The proposed CNN achieved superior AUPRC compared to the other classifiers.

We evaluated the various systems on the MUSC database. MUSC dataset is a high-quality dataset with 18 expert cross-annotations. On this dataset, we also compared our system with Persyst, the current market leader in EEG software. The comparison results corroborated with the 5-fold cross-validation results. The CNN system exhibited superior performance in comparison with the other IED detection systems, including Persyst. However, Persyst requires a three minute EEG segment for calibration [34]. Since the MUSC dataset consists of only 30-second EEG segments, the calibration may not be proper. Further, we have evaluated the proposed system with the consensus of the experts. The system was found to be non-inferior to the agreement of the experts.

Our system has achieved superior performance in comparison with the different methods proposed in the literature. We compare the studies that have at least an IED count of 1000, with all the performance indices reported. Le et al. have proposed a Deep Belief Network (DBN) for detecting IEDs [91]. This study is performed on data from 19 subjects with 1,491 annotated events. The study employs a feature extraction method in combination with a DBN and has achieved a mean leave-one-subject-out cross-validation

(LOSO-CV) sensitivity of 76% for the specificity of 97.24%. The results are quite low for clinical application even though the testing is performed on an IED vs. background ratio of 1:2 [91]. Tjepkema-Cloostermans et al. had developed a CNN based IED detector which is shown to have achieved a sensitivity of 47.4% for a specificity range of 98-99% on a dataset of 100 subjects (with 1815 annotated IEDs) [90]. The evaluation was reported for a single run, i.e., cross-validation was not performed. Also, the study does not report any benchmark testing on the traditional classifiers. In [93], Ganglberger et al. has evaluated Persyst on a dataset of 5,474 IEDs and achieved a sensitivity of 43.9% for an FP rate of 1.65 per minute. In the above studies, all the performance indices have not been reported. In our analysis, our proposed system has achieved a sensitivity of 80% for an FP rate of 0.2 per minute. Moreover, in our tests on the MUSC data, we have shown that the proposed CNN system outperforms the traditional classifier systems as well as Persyst.

The final objective of the study is to develop an automated system to identify epileptic EEG, rather than identifying individual epileptiform patterns. From a clinical perspective, only a few clear IEDs are necessary to conclude whether the EEG is epileptic. However, identification of interictal patterns is challenging due to the presence of artifacts and due to the rarity of the IED events. Therefore, an efficient IED detection system is mandatory. In the next chapter, we describe the development of the EEG classification system based on the output from the proposed CNN-based IED detector.

“Science is simply common sense at its best, that is, rigidly accurate in observation, and merciless to fallacy in logic.”

Thomas Huxley

6

Epileptic EEG Detector

An EEG-level prediction based on EEG has more clinical significance than accurate identification of individual epileptic transients. In this chapter, we describe the development of an EEG classification system based on the IED detector output. It includes the following parts: feature extraction, selection, and classification techniques. We detail the methodology adopted for developing the EEG classification system and the feature definitions in Section 6.2. We have described the cross-validation results on the MGH dataset in Section 6.3.1 and the clinical validation results at NUH in Section 6.3.2.

6.1 INTRODUCTION

An automated IED detection system is highly beneficial for the clinical assessment and treatment of epilepsy. Often, the neurologists are not concerned about the number of IEDs identified to diagnose epilepsy. Consequently, the EEG-level classification (whether it is epileptic or non-epileptic) is more significant in comparison with detecting individual IEDs. We propose an automated EEG classification system to detect epileptic EEGs based on IEDs. The system comprises of three major modules: preprocessing, waveform-level classification, and EEG-level classification. The preprocessing module performs the necessary filtering, normalization, and configures the EEG montage. The waveform-level classification module detects the various IED patterns in the EEG signal. We apply the preprocessing steps highlighted in Section 4.2 for the preprocessing module and the IED detectors developed in Chapter 5 for waveform-level classification.

The EEG-level classification module is developed based on the output of the waveform-level classification to identify epileptic EEGs. Multiple features are extracted from the waveform-level classification output and are applied for classifying EEGs. We apply three different approaches to develop the EEG classification system based on a combination of feature sets and methods. In Model 1, we employ simple thresholding of features extracted from the waveform-level classification output. In Model 2, instead of simple thresholding, we employ a traditional classifier for EEG classification. Here we employ Support Vector Machine (SVM) [27] with Gaussian kernel. SVMs have been shown to perform well amongst the traditional classifiers. Considering the small input dataset size for EEG-level classification, we employ an SVM classifier. In both Model 1 and 2, only CNN system is applied at the waveform-level classification. In Model 3, we employ an ensemble of classifiers in parallel to boost the IED detection as well as generate multiple features. We employ an SVM for EEG classification, similar to Model 2.

To the best of the authors' knowledge, no research work in the literature have tried to classify entire EEGs as IED-free vs. containing IEDs. Most studies that claim to have achieved 'EEG classification', had only tested on segments of EEG rather than a whole EEG. Here we have performed the evaluation based on clinical data from MGH and NUH. The MGH data was

applied for the training of the system, and later it was applied to the data from NUH for clinical validation. We achieved a mean 5-fold classification balanced accuracy (BAC) of 81.61% for identifying EEGs with and without IEDs on the MGH data. We have achieved promising results on NUH data with an accuracy of 88%.

6.2 METHODOLOGY

We propose the development of a robust, fast, and efficient EEG classification system with three modules: preprocessing, waveform-level classification, and EEG-level classification. The block diagram of the proposed system is illustrated in Fig. 6.2.1.

Figure 6.2.1: Block diagram of the proposed EEG classification system.

The EEG-level classification system predicts whether the EEG contains IEDs by analyzing the output of waveform-level classification. We develop three different Models for EEG classification (see Fig. 6.2.2):

1. **Model 1:** Here we employ simple thresholding to the waveform-level output, and the EEG is classified based on the number of IEDs identified (see Fig. 6.2.2a). The CNN-based IED detector is employed at the waveform-level classification.
2. **Model 2:** Here we extract different features from the waveform-level output and employ a traditional classifier for EEG classification (see Fig. 6.2.2b). The extracted features are applied as the input to the EEG classifiers. Similar to Model 1, only the CNN-based IED detector as applied at the waveform-level classification. An SVM with Gaussian kernel is employed for EEG classification.

3. **Model 3:** Here we employ multiple IED detectors in parallel at the waveform-level classification. We employ the CNN and TM IED detector in parallel for feature extraction. We apply a threshold of 0.9 for CNN. The TM detector is applied to the EEGs separately, and statistical features are extracted from the TM IED detector output. Finally, an SVM is trained based on the extracted features (see Fig. 6.2.2c).

The different parameters of the SVM were optimized by applying Bayesian optimization [141] with 5-fold nested cross-validation on the training set (see Section 5.2.4). We employ the same 5-fold cross-validation steps detailed in Section 5.2.1 for evaluating the EEG classification system. We applied four folds for training: three folds for waveform-level classification, one fold for feature extraction and EEG classifier training. The final fold was employed for testing. The entire process was repeated 20 times, and the performance indices are averaged. We extract various features from the waveform-level classification system. The different feature sets are:

- **Feature set 1:**
 - *CNN detections with a specified threshold:* It quantizes the number of detections or IED identifications by the IED detector. This feature is similar to the visual inspection performed by the experts, where the decision is made based on a few IED occurrences.
 - *CNN detection per second for a specified threshold:* Here we consider the detection rate rather than the number of detections. This would help us remove the dependency on the duration of EEG recordings.
- **Feature set 2:** *Fraction of CNN outputs in a specified threshold range:* The CNN output range $[0,1]$ is divided into 10 equal ranges of width 0.1: $[0,0.1)$, $[0.1,0.2)$, ..., $[0.9,1]$. Each feature is defined as the fraction of time-instant CNN outputs that belongs to a particular threshold range.
- **Feature set 3:** *Mean of correlation distance values from the IED and background template libraries:* This feature is employed in Model 3 while applying the TM IED detector ensemble.

Figure 6.2.2: Block diagrams for the different EEG classifier systems: (a) Model 1; (b) Model 2, and (c) Model 3.

Model 1 is simple and the most straightforward to implement. We investigate feature set 1 with Model 1. The challenge while applying simple thresholding is that we will not be able to translate the system into a clinical product. The baseline of EEG is quite variable across the population. Moreover, the environmental factors, equipment factors, would affect the threshold significantly. As a result, we apply feature sets 2 and 3, and make a decision based on the relative significance and separability of the features. As a result, Models 2 and 3 would be more viable for practical applications. Further, we evaluate our system on the NUH dataset. We employ the same 5-fold data split for this evaluation: three folds for IED detector training and two folds for EEG-level classifier training. First we select the best IED detector based on the waveform-level classification results. Then the EEG-level classifier is trained with the remaining two folds. Fig. 6.2.3 illustrates the development of the EEG classifier based on the MGH dataset. We also calibrate the EEG-level classifier threshold based on the NUH phase 1 dataset (25 epileptic and 25 non-epileptic labeled EEGs). The calibration is performed by adjusting the decision boundary of the EEG classifier based on the phase 1 dataset evaluation. Next we evaluate the NUH phase 2 data (100 EEGs) with the calibrated system. The EEG training process is reiterated 100 times in order to substantiate the results.

Figure 6.2.3: Development of the EEG classifier based on the MGH dataset.

6.3 RESULTS

6.3.1 CROSS-VALIDATION RESULTS

The EEG classification system was developed based on three modules: pre-processing modules, waveform-level classification module, and the EEG classification module. The EEG classification module was developed based on the output of the waveform-level classification module. We applied three different approaches with different features for the EEG classification system.

We first evaluate the feature set 1 with a CNN threshold of 0.9. The boxplots for the number of cnn detections and the number of cnn detections per second are presented in Fig. 6.3.1. The barplot for the number of detections for epileptic and non-epileptic EEGs is illustrated in Fig. 6.3.2a. The comparison between epileptic and non-epileptic EEGs for the feature, CNN detections per second is illustrated in Fig. 6.3.2b. For both the features, it can be observed that the two different classes are separable based on the identification of IEDs. The time-normalized features have shown a superior capability to separate epileptic and non-epileptic EEGs.

Figure 6.3.1: Boxplots for the two different features with a CNN threshold of 0.9 for epileptic and non-epileptic EEG: (a) CNN detections; (b) CNN detections per second.

(a)

(b)

Figure 6.3.2: Barplots for the two different features with a CNN threshold of 0.9 for epileptic and non-epileptic EEG: (a) CNN detections; (b) CNN detections per second.

Model 1 was developed by applying simple thresholding to the CNN outputs. The threshold is determined based on the train data. We chose the threshold that best separates the epileptic and non-epileptic EEGs. We applied two different features for Model 1, namely, the number of CNN detections and the number of CNN detections per second. The latter is invariant to the duration of the EEG recording. We also performed a two-sample t -test to determine the ability of the two features to distinguish the two different classes, i.e., epileptic and non-epileptic. The t -test was performed with the null hypothesis that the two vectors came from independent random samples of a normal distribution with equal means and equal but unknown variances. The null hypothesis was rejected across the different combination of folds with 1% significance level for both the features and the CNN detections per second achieved lower p -value in comparison with the CNN detections feature. This suggests that both the features are capable of separating the two classes, with CNN detections per second being the more efficient feature. The results for Model 1 with different features are detailed in Tables 6.3.1 and 6.3.2. The system based on CNN detections per second feature performed with superior performance values compared the number of detections feature. Model 1 has achieved a sensitivity of $78.92 \pm 0.16\%$ and specificity of $84.30 \pm 0.12\%$ with the best feature. The performance curves for Model 1 with the different features are illustrated in Figs. 6.3.3 and 6.3.4.

Table 6.3.1: Results for EEG-level classification based on the waveform-level classification features for Model 1 and based on CNN detections.

Trial no.	CNN training set	SVM training set	Testing set	AUC	TP	TN
1	Fold 1, 2, 3	Fold 4	Fold 5	0.866	7/18	78/93
2	Fold 1, 2, 3	Fold 5	Fold 4	0.724	14/19	72/92
3	Fold 1, 2, 4	Fold 3	Fold 5	0.706	16/18	65/92
4	Fold 1, 2, 4	Fold 5	Fold 3	0.907	6/18	81/93
5	Fold 1, 2, 5	Fold 4	Fold 3	0.917	16/18	82/92
6	Fold 1, 2, 5	Fold 3	Fold 4	0.976	13/19	90/92
7	Fold 1, 3, 4	Fold 5	Fold 2	0.756	17/19	72/92
8	Fold 1, 3, 4	Fold 2	Fold 5	0.951	10/18	72/93
9	Fold 1, 3, 5	Fold 2	Fold 4	0.933	19/19	81/92
10	Fold 1, 3, 5	Fold 4	Fold 2	0.974	16/19	83/92
11	Fold 1, 4, 5	Fold 3	Fold 2	0.924	16/19	71/92
12	Fold 1, 4, 5	Fold 2	Fold 3	0.906	15/18	81/92
13	Fold 2, 3, 4	Fold 5	Fold 1	0.695	14/19	91/92
14	Fold 2, 3, 4	Fold 1	Fold 5	0.969	14/18	37/93
15	Fold 2, 3, 5	Fold 1	Fold 4	0.934	14/19	92/92
16	Fold 2, 3, 5	Fold 4	Fold 1	0.972	17/19	71/92
17	Fold 2, 4, 5	Fold 3	Fold 1	0.945	13/19	92/92
18	Fold 2, 4, 5	Fold 1	Fold 3	0.991	16/18	72/92
19	Fold 3, 4, 5	Fold 1	Fold 2	0.974	13/19	92/92
20	Fold 3, 4, 5	Fold 2	Fold 1	0.963	19/19	69/92
Mean	Fold 3, 4, 5	Fold 2	Fold 1	0.899	-	-
Std	Fold 3, 4, 5	Fold 2	Fold 1	0.09	-	-
Mean Sensitivity				76.46±0.17%		
Mean Specificity				83.76±0.14%		
Balanced accuracy				80.11%		

Table 6.3.2: Results for EEG-level classification based on the waveform-level classification features for Model 1 and based on CNN detections per second.

Trial no.	CNN training set	SVM training set	Testing set	AUC	TP	TN
1	Fold 1, 2, 3	Fold 4	Fold 5	0.869	7/18	81/93
2	Fold 1, 2, 3	Fold 5	Fold 4	0.720	13/19	82/92
3	Fold 1, 2, 4	Fold 3	Fold 5	0.712	16/18	66/92
4	Fold 1, 2, 4	Fold 5	Fold 3	0.915	8/18	80/93
5	Fold 1, 2, 5	Fold 4	Fold 3	0.921	16/18	83/92
6	Fold 1, 2, 5	Fold 3	Fold 4	0.978	17/19	75/92
7	Fold 1, 3, 4	Fold 5	Fold 2	0.759	19/19	75/92
8	Fold 1, 3, 4	Fold 2	Fold 5	0.966	14/18	59/93
9	Fold 1, 3, 5	Fold 2	Fold 4	0.937	19/19	80/92
10	Fold 1, 3, 5	Fold 4	Fold 2	0.980	13/19	89/92
11	Fold 1, 4, 5	Fold 3	Fold 2	0.939	16/19	73/92
12	Fold 1, 4, 5	Fold 2	Fold 3	0.919	15/18	81/92
13	Fold 2, 3, 4	Fold 5	Fold 1	0.698	14/19	91/92
14	Fold 2, 3, 4	Fold 1	Fold 5	0.953	14/18	45/93
15	Fold 2, 3, 5	Fold 1	Fold 4	0.936	15/19	91/92
16	Fold 2, 3, 5	Fold 4	Fold 1	0.981	16/19	72/92
17	Fold 2, 4, 5	Fold 3	Fold 1	0.966	15/19	92/92
18	Fold 2, 4, 5	Fold 1	Fold 3	0.972	16/18	77/92
19	Fold 3, 4, 5	Fold 1	Fold 2	0.980	12/19	92/92
20	Fold 3, 4, 5	Fold 2	Fold 1	0.969	19/19	70/92
Mean	Fold 3, 4, 5	Fold 2	Fold 1	0.903	-	-
Std	Fold 3, 4, 5	Fold 2	Fold 1	0.09	-	-
Mean Sensitivity				78.92±0.16%		
Mean Specificity				84.30±0.12%		
Balanced accuracy				81.61%

Figure 6.3.3: Performance curves for Model 1 EEG classifier based on CNN detections: (a) ROC; (b) PRC.

Figure 6.3.4: Performance curves for Model 1 EEG classifier based on CNN detection per second: (a) ROC; (b) PRC.

In Fig. 6.3.5, we illustrate the scatter plot for the best features for the feature set 2, namely, the fraction of CNN detections in the range (0.8-0.9] and the fraction of CNN detections in the range (0.9-1]. The features were evaluated based on p -values. The plot indicates that the two classes, EEGs with and without IEDs, are reasonably separable by the feature set 2. We developed Model 2 based on the optimized feature set. It achieved an AUC of 0.795, mean sensitivity of 76.33%, and specificity of 76.30% (see Table 6.3.3). The performance curves for Model 2 is illustrated in Fig. 6.3.6.

Figure 6.3.5: Scatter plot of the two most salient features, i.e., the fraction of CNN outputs between the intervals (0.8-0.9] and (0.9-1] for EEG-level classification based on Model 2. The plot shows that the two classes are reasonably separable.

Figure 6.3.6: Performance curves for Model 2 EEG classification based on feature set 2: (a) ROC; (b) PRC.

Table 6.3.3: The results for EEG-level classification based on the waveform-level classification features for Model 2.

Trial no.	CNN training set	SVM training set	Testing set	AUC	TP	TN
1	Fold 1, 2, 3	Fold 4	Fold 5	0.667	6/19	77/92
2	Fold 1, 2, 3	Fold 5	Fold 4	0.820	12/18	81/93
3	Fold 1, 2, 4	Fold 3	Fold 5	0.788	16/18	58/93
4	Fold 1, 2, 4	Fold 5	Fold 3	0.713	12/18	45/92
5	Fold 1, 2, 5	Fold 3	Fold 4	0.953	16/19	84/92
6	Fold 1, 2, 5	Fold 4	Fold 3	0.907	14/18	85/92
7	Fold 1, 3, 4	Fold 2	Fold 5	0.834	17/18	68/93
8	Fold 1, 3, 4	Fold 5	Fold 2	0.647	12/19	53/92
9	Fold 1, 3, 5	Fold 2	Fold 4	0.873	16/19	76/92
10	Fold 1, 3, 5	Fold 4	Fold 2	0.867	17/19	62/92
11	Fold 1, 4, 5	Fold 2	Fold 3	0.720	9/18	88/92
12	Fold 1, 4, 5	Fold 3	Fold 2	0.791	16/19	68/92
13	Fold 2, 3, 4	Fold 1	Fold 5	0.753	17/18	65/93
14	Fold 2, 3, 4	Fold 5	Fold 1	0.641	11/19	60/92
15	Fold 2, 3, 5	Fold 1	Fold 4	0.772	13/19	84/92
16	Fold 2, 3, 5	Fold 4	Fold 1	0.697	15/19	58/92
17	Fold 2, 4, 5	Fold 1	Fold 3	0.847	12/18	92/92
18	Fold 2, 4, 5	Fold 3	Fold 1	0.917	18/19	61/92
19	Fold 3, 4, 5	Fold 1	Fold 2	0.964	15/19	85/92
20	Fold 3, 4, 5	Fold 2	Fold 1	0.722	19/19	57/92
Mean	Fold 3, 4, 5	Fold 2	Fold 1	0.795	-	-
Std	Fold 3, 4, 5	Fold 2	Fold 1	0.10	-	-
Mean Sensitivity				76.33±0.17%		
Mean Specificity				76.30±0.14%		
Balanced accuracy				76.31%		

We developed Model 3 based on features extracted from an ensemble of IED detectors. The different feature combination scatterplots are illustrated in Figs. 6.3.7a and Fig. 6.3.7b. We observe that the additional features help to distinguish the two different classes more effectively. Model 3 achieved an AUC of 0.868, mean sensitivity of 82.22%, and specificity of 76.44%. The ROC and PRC of Model 3 are illustrated in Fig. 6.3.8.

Figure 6.3.7: Scatter plot of the CNN detections per second and the mean correlation distance from : (a) IED templates; (b) background templates.

Figure 6.3.8: Performance curves for Model 3 EEG classification based on feature set 3: (a) ROC; (b) PRC.

Table 6.3.4: The results for EEG-level classification based on the waveform-level classification features for Model 3.

Trial no.	CNN training set	SVM training set	Testing set	AUC	TP	TN
1	Fold 1, 2, 3	Fold 4	Fold 5	0.758	13/19	61/92
2	Fold 1, 2, 3	Fold 5	Fold 4	0.799	15/18	70/93
3	Fold 1, 2, 4	Fold 3	Fold 5	0.939	18/18	75/93
4	Fold 1, 2, 4	Fold 5	Fold 3	0.842	11/18	68/92
5	Fold 1, 2, 5	Fold 3	Fold 4	0.939	14/19	85/92
6	Fold 1, 2, 5	Fold 4	Fold 3	0.818	12/18	81/92
7	Fold 1, 3, 4	Fold 2	Fold 5	0.916	16/18	83/93
8	Fold 1, 3, 4	Fold 5	Fold 2	0.748	13/19	56/92
9	Fold 1, 3, 5	Fold 2	Fold 4	0.938	18/19	74/92
10	Fold 1, 3, 5	Fold 4	Fold 2	0.834	17/19	80/92
11	Fold 1, 4, 5	Fold 2	Fold 3	0.888	9/18	88/92
12	Fold 1, 4, 5	Fold 3	Fold 2	0.960	18/19	67/92
13	Fold 2, 3, 4	Fold 1	Fold 5	0.870	14/18	89/93
14	Fold 2, 3, 4	Fold 5	Fold 1	0.759	16/19	47/92
15	Fold 2, 3, 5	Fold 1	Fold 4	0.867	14/19	89/92
16	Fold 2, 3, 5	Fold 4	Fold 1	0.908	19/19	37/92
17	Fold 2, 4, 5	Fold 1	Fold 3	0.851	13/18	89/92
18	Fold 2, 4, 5	Fold 3	Fold 1	0.936	18/19	51/92
19	Fold 3, 4, 5	Fold 1	Fold 2	0.867	14/19	87/92
20	Fold 3, 4, 5	Fold 2	Fold 1	0.933	18/19	53/92
Mean	Fold 3, 4, 5	Fold 2	Fold 1	0.868	-	-
Std	Fold 3, 4, 5	Fold 2	Fold 1	0.06	-	-
Mean Sensitivity				82.22±0.12%		
Mean Specificity				76.44±0.16%		
Balanced accuracy				79.33%		

The MGH dataset consists of 93 recordings from 84 patients, with 7 patients having multiple EEG recordings over time. Since the EEGs are recorded independently, longitudinally over time, it is justified to consider them as separate for EEG classification analysis. For a fair comparison, we have added the recordings of the same patient in the same fold (see Section 5.2.1). Here we have computed the results considering the multiple recordings from the same patient as independent EEG records. Further, the optimization of the SVM parameters also induce some variation in the results: we observed a variation of $\pm 2\text{-}3\%$ in the performance measures at each run. However, this issue does not affect our evaluation of the NUH dataset.

6.3.2 CLINICAL VALIDATION

We performed the clinical validation of our EEG classification system on 100 EEGs (50 epileptic and 50 non-epileptic EEGs) from NUH, Singapore. The evaluation was performed in the supervision of Dr. Rahul Rathakrishnan, Senior Consultant with the Division of Neurology at the NUH. Initially, a dataset of 50 labeled EEGs (25 epileptic and 25 non-epileptic) was provided to calibrate our system. Later the 100 evaluation EEG files were provided without the labels. Finally, the predicted labels were forwarded to NUH in order to compare them with the clinician’s report. Our system achieved an accuracy of 88% (sensitivity: 43/50, specificity: 45/50) in identifying epileptic and non-epileptic EEGs. Later the original labels were shared, and further analysis on the dataset was performed. We initially evaluated the CNN detection per second for the NUH dataset. The comparison boxplots between MGH and NUH are illustrated in Fig. 6.3.9. Further, we evaluated the three different models on NUH data. We applied the 50 EEGs for calibration of the various models, similar to that of the clinical validation procedure. We trained the different models on the entire MGH dataset and later evaluated the NUH 100 EEGs. Since the number of CNN detection were lower for NUH data, we apply a CNN threshold of 0.5 while extracting the features for NUH analysis. The EEG classifier was trained multiple times (100 times) in order to substantiate the results. The results for the three different models are presented in Table 6.3.5.

Figure 6.3.9: Boxplots of CNN detections per second for MGH and NUH datasets. The epileptic and non-epileptic EEGs are separable for both the datasets, but with different thresholds.

Table 6.3.5: The results for the different models on the NUH data.

Model	Accuracy Mean $\pm$ Standard deviation
Model 1	83%
Model 2	78.30% $\pm$ 3.77
Model 3	86.58% $\pm$ 2.74

6.4 DISCUSSION

We achieved a mean EEG-level classification accuracy of 81.61% (Model 1) over the 20 different combinations of folds. Most of the studies in the literature are limited to epileptic IED detection. Classification of EEGs based on the interictal epileptiform activity is a relatively novel concept. Altunay et al. has presented a similar study in [143] by applying a linear prediction filter to classify EEGs based on IEDs. The system is reported to have achieved a sensitivity of 95%, but a direct comparison cannot be made as the study was conducted on intracranial EEG data of 5 patients. We have also achieved clinical validation of our system by applying it to a ‘blind’ dataset for evaluation.

We also explored the cause for the misclassifications. We observed that the EEGs were misclassified either due to less number of IED detections (false negatives) or due to false IED detections (false positives). We reviewed a few EEGs with false positives and observed that most of the non-IED misclassifications were due to artifacts. We also compared the characteristics of the subject (age, gender) with the predictions of the EEG classification system. The complexity of EEG has been shown in the literature to vary across different age groups [144]. This could result in considerable variability in the IED morphology. Recent studies [145] have shown that the EEG is discriminable with gender. The results for the different characteristics are illustrated in Fig. 6.4.1 (age) and Fig. 6.4.2 (gender). We observed that most of the misclassification were from age groups less than 30 or greater than 70. In evaluating gender, we did not find any significant difference between male and female subjects.

Figure 6.4.1: Scatter plots for subject age vs. the output of the EEG classification system.

Figure 6.4.2: Scatter plots for subject gender vs. the output of the EEG classification system.

“The measure of greatness in a scientific idea is the extent to which it stimulates thought and opens up new lines of research.”

Paul Dirac

7

Extending the Approach for Seizure Detection and Brain-Computer Interfaces

In this chapter, we explore the applicability of our proposed system in other domains. Here we elaborate on the extended applications of our CNN-based health detection system. We have applied and verified the effectiveness of the CNN system for detection of epileptic seizures and for classifying various Brain-Computer Interfaces (BCIs) based on EEG. The application for seizures detection is detailed in Section 7.1 and the BCI work in Section 7.2. For seizure detection, a temporal segmentation of EEGs is applied similar to that of IED detection. For BCI classification, we have applied the spectrograms instead of the temporal EEG in order to make a direct comparison with the literature. In both scenarios, the CNN system was found to be more effective compared to the traditional methods in the respective domains.

7.1 SEIZURE DETECTION

7.1.1 INTRODUCTION

EEG is considered the most prominent tool for seizure detection due to adequate temporal resolution, inexpensive and painless. However, visual scanning of long-term EEG recordings becomes a time-consuming, tedious, error-prone task, and expert centered process. Besides, the interpretation of the same recording can vary from different annotators. Thus, there exists a significant need for an automated seizure detection system that would help the long-term EEG monitoring practice and treatment planning for epilepsy patients. In recent years, a variety of machine learning and bioengineering-based algorithms have been applied for seizure detection from scalp EEG [2, 4, 5]. Most of the studies employ hand-crafted features characteristic of seizure manifestations in EEG; most methods rely on spectral information [146–151], whereas some of them employ the temporal characteristics of seizures [146, 152]. However, epileptic seizures are highly non-stationary with seizure manifestations in EEG being extremely irregular, both within a patient over time and between different patients. Consequently, finding the optimal features for seizure detection is yet a challenging, important problem [146].

Here, we present a method based on the deep convolutional neural networks (CNNs) to perform unsupervised feature learning framework for automated seizure onset detection. The proposed system was evaluated on 526 hours duration of scalp EEG data, including 181 seizures of 23 pediatric patients. The different parameters of CNN were optimized through 4-fold nested cross-validation. The resulting generalized CNN seizure detection model achieved an average sensitivity of 86.29%, an average false detection rate of 0.74 h-1, and an average detection latency of 2.1 sec.

7.1.2 METHODOLOGY

7.1.2.1 SCALP EEG AND PREPROCESSING

In this study, a publically available CHB-MIT database (<https://www.physionet.org/pn-6/chbmit/>) was utilized. The data was recorded with the sampling frequency of 256 Hz at Children’s Hospital Boston (CHB) in collabora-

tion with the Massachusetts Institute of Technology (MIT). It contains scalp EEG signals recorded from 23 pediatric patients (5 males, 17 females and one missing gender information) with 198 annotated epileptic intractable seizures (not controlled with seizure medications). The different details of the patients are detailed in Table 7.1.1. There was 983 hours duration of raw EEG recordings from 23 channels which were recorded according to the International 10-20 electrodes positioning system. The recordings were organized in 24 cases, each containing EEG data from a single subject, except cases 1 and 21 which were obtained from the same subject with a 1.5-year interval. It was noticed that 17 seizures files had different montages (monopolar) to the rest of the seizure files (bipolar recordings). Thus, these 17 records were excluded from this study. We also excluded the number of session records from the analysis that contained dashes (missing data) in the original data and EEG signals with extreme artifacts. In the final dataset, 24 case's EEG recordings with length up to 526 hours' duration containing 181 seizure events were evaluated.

The block diagram of the proposed CNN based seizure detection system is illustrated in Fig. 7.1.1. Initially, the scalp EEG data were preprocessed by applying a 4th order zero-phase Butterworth bandpass filter between 1 and 30 Hz. Later we segment the data into 5-sec waveforms for CNN learning (1280 samples $\times$ 1 channel). This resulted in 2141 seizure segments, which were extracted from 181 seizure events. We also extracted 2141 non-seizure segments randomly for balanced training.

Figure 7.1.1: Workflow of the proposed CNN based seizure detection system implementation.

Table 7.1.1: Details on the CHB-MIT scalp EEG dataset. Gender:female (Fe), male (Ma). Others: not available (NA).

Case ID	Age (years)	Gender	No. of seizures	Seizure duration (sec) mean $\pm$ SD	EEG recordings used (hours)
01	11	Fe	7	63.14 $\pm$ 30.47	25
02	11	Ma	3	57.33 $\pm$ 41.86	20
03	14	Fe	7	57.43 $\pm$ 8.38	20
04	22	Ma	4	94.50 $\pm$ 30.88	40
05	7	Fe	5	111.6 $\pm$ 9.45	19
06	1.5	Fe	10	15.30 $\pm$ 2.87	28
07	14.5	Fe	3	108.30 $\pm$ 30.44	32
08	3.5	Ma	5	183.8 $\pm$ 49.20	18
09	10	Fe	4	69.00 $\pm$ 7.70	32
10	3	Ma	7	63.86 $\pm$ 17.24	24
11	12	Fe	3	268.67 $\pm$ 418.61	20
12	2	Fe	27	36.63 $\pm$ 19.00	15
13	3	Fe	10	44.00 $\pm$ 22.40	10
14	9	Fe	8	21.13 $\pm$ 8.68	21
15	16	Ma	20	99.6 $\pm$ 53.58	33
16	7	Fe	8	8.62 $\pm$ 2.50	14
17	12	Fe	3	97.67 $\pm$ 15.04	16
18	18	Fe	6	52.83 $\pm$ 14.43	24
19	19	Fe	3	78.67 $\pm$ 2.08	15
20	6	Fe	8	36.75 $\pm$ 6.25	21
21	13	Fe	4	49.75 $\pm$ 28.52	20
22	9	Fe	3	68.00 $\pm$ 8.72	18
23	6	Fe	7	60.57 $\pm$ 32.57	20
24	NA	NA	16	31.94 $\pm$ 18.41	21
18Fe / 5Ma			181		526

7.1.2.2 APPROACH

We apply four-fold cross-validation for evaluating the proposed CNN based seizure detection system. During the CNN evaluation process, the 24 epileptic cases were divided into four folds, randomly, by keeping the distribution of seizure segments similar in all the folds. The different fold-split details are given in Table 7.1.2. For each cross-validation step, three folds (18 cases) were applied for training, and the testing was performed on the remaining

fold (6 cases). The CNN model training was terminated once the validation error stopped improving. We consider the cross-entropy as the error function for the training and validation of the CNN architecture. The different hyper-parameters and the stopping criteria of the system were optimized by applying nested cross-validation on the training dataset: 70% for training and 30% for validation.

The proposed CNN structure is illustrated in Fig. 7.1.2. Initially, the convolution operation is performed by convolving each electrode signal with a 1-dimensional filter (to capture the temporal information of the EEG signal) and passed through the output function to form the feature map. Here, the output function is selected as the rectifier linear unit (ReLU) function. Following this, a pooling layer is utilized to provide a squashing function. Many pooling strategies have been proposed in the literature; however, for this study, we applied the max-pooling. The generated features from the convolutional layers are passed through a fully connected hidden layer. Finally, the CNN output layer was mapped between $[0,1]$ with a softmax function. The optimized CNN model was tested on the testing fold at 5-sec window (1280 samples $\times$ 1 channel) with a 75% overlap. For each 5-sec EEG input, the CNN generates a probability value between $[0,1]$ with the higher value representing the seizure waveform. In order to obtain the time instant detection, we combined the 21-channel CNN outputs from a single time instant together. The maximum value of 21-channel was chosen as the combined time instant output, resulting in a single value between $[0,1]$ for each 5-sec EEG window. These procedures were repeated until all the folds were employed as test data.

Table 7.1.2: Distribution of seizure segments in each folds

Fold number	Case IDs	Number of seizure events	No. of seizure segments extracted
1	CHB-01,03,04,05,10,23	37	518
2	CHB-02,08,12,20,22,24	62	594
3	CHB-07,09,11,13,17,18	29	482
4	CHB-06,14,15,16,19,21	53	547
5	Total	181	2141

Figure 7.1.2: Structure of the implemented CNN network.

Table 7.1.3 summarizes the different parameter combinations we evaluated for CNN training, while information on the dimensions of our data structures is presented in Table 7.1.4. Here, X denotes the input, F are the filters for the layer, C are the feature maps for the layer, W^0 denotes the weight matrix for the fully connected hidden layer, b^0 and b^1 are the pre-activation, and post-activation features for the fully connected hidden layer, W' and l are the weight matrix and activation of the logistic regressor.

Table 7.1.3: The different parameters evaluated for CNN implementation

S.No	Parameter Value
Number of convolutional layers	1,2,3,4,5
Number of convolutional filters	40,30,20,10,5
Dimensions of convolutional filters	$1 \times 64, 1 \times 32, 1 \times 24,$ $1 \times 16, 1 \times 8$
Number of pooling layers	5
Number of fully connected layer neurons	400
Optimizer	Adam Optimizer with learning rate of 10^{-3}
Activation function	ReLU (Rectified Linear Unit)
Dropout probability	0.5
Training termination criteria	Until the validation error stopped improving

7.1.3 RESULTS AND DISCUSSION

The proposed deep CNN models were designed in Python (v3.5.2) with Tensorflow 1.5.0 [132] with a K80 Tesla GPU. The performance of the proposed automated seizure detection system was evaluated based on three different performance measures: sensitivity, false detection rate (FDR), and latency. Sensitivity is the percentage of true seizure events detected by the system in the test dataset. FDR is calculated by the number of times algorithm misclassifies an EEG segment as a seizure event per hour. Latency is the time delay (in seconds) between the seizure detected by the algorithm and the seizure onset marked by the clinical experts. Table 7.1.5 presents the CNN seizure detection results for each fold. The proposed method achieved an average detection sensitivity of 86.29%. Furthermore, the algorithm misclassified 392 EEG segments as seizure event in total. This records an average FDR of 0.74 per hour. Finally, the algorithm achieved an average detection latency of 2.1 seconds across all the folds. Along with the detection latency we also need to account for two more latency offsets: former in the onset of the detection process, we require a delay of 5-seconds to initiate the detection process; the latter due to the sliding window of 75%, for each subsequent detection suffers from an additional latency of 1.25 seconds. The onset delay can be neglected, assuming we do not encounter a seizure dur-

Table 7.1.4: Information regarding the dimension of our data structures.

Type	Size
Input layer (X)	X: 1 channel $\times$ 1280 samples
Convolutional hidden layers ($\vec{F}, \vec{C}$)	F1: 64 coefficients $\times$ 40 filters C1: 23 channels $\times$ 640 samples $\times$ 40 feature maps F2: 32 coefficients $\times$ 30 filters C2: 1 channel $\times$ 320 samples $\times$ 30 feature maps F3: 24 coefficients $\times$ 20 filters C3: 1 channel $\times$ 160 samples $\times$ 20 feature maps F4: 16 coefficients $\times$ 10 filters C4: 1 channel $\times$ 80 samples $\times$ 10 feature maps F5: 8 coefficients $\times$ 5 filters C5: 1 channel $\times$ 40 samples $\times$ 5 feature maps
Fully connected hidden layer (W^0, b^0, b^l)	W^0 : 400 neurons $\times$ 4600 coefficients b^0, b^l : 400 features $\times$ 1 sample
Output layer (W^0, l)	W^l : 2 classes $\times$ 400 neurons l : 2 classes $\times$ 1 probability

ing the initiation of the detection system. The delay caused due to the application sliding window can be reduced by reducing the window length of EEG segments applied for training or by increasing the sliding window overlap. However, this might increase the processing time of the system. Consequently, the accumulated delay of our system is 3.75 seconds (training window length (5 seconds) $\times$ overlap [0.75]).

The sensitivity-precision curves for the different folds are displayed in Fig. 7.1.3, here the test data classes (number of seizures and non-seizures EEG segments) are highly imbalanced. The plot shows the trade-off between precision and sensitivity for different thresholds. It can be observed that the system has achieved an average precision of 0.286 for the corresponding average sensitivity of 86.29%. The obtained results illustrate that

CNN can successfully characterize the oscillatory patterns of seizure EEG by extracting higher-level features automatically from the raw EEG input.

Table 7.1.5: CNN seizure detection performance

Fold number	EEG duration	Seizure event	TP	FP	FN
1	148	37	34	95	4
2	113	62	55	79	6
3	134	29	22	108	7
4	131	53	47	110	6
Total	526	181	158	392	23

Figure 7.1.3: The sensitivity-precision curves for each fold.

The proposed algorithm is geared towards generalized seizure detection via long-term recordings of scalp EEG. Most of the previous works in this area had focused on detecting seizures using data from individual patients (i.e., patient-specific seizure onset detection). In this study, however, the focus has been to generalize seizure detection across a group of subjects from the CHB-MIT database. Therefore, we compared the performance of our algorithm with appropriate methods in the literature that analyses long-

term EEG recordings for generalized seizure detection. The performance of our CNN based seizure detection system is superior to the previous studies in the literature [4, 153–155]. In [4], Wilson et al. has achieved a sensitivity of 67% with a false detection per of 1.7 per hour based on a Reveal algorithm. Similarly, Furbass et al. has achieved a sensitivity of 67% for a false detection rate of 0.32 [153] and Thodoroff et al. has achieved a sensitivity of 81% for a false detection rate of 0.29 per hour [154] on the CHB-MIT database. The proposed CNN model achieved an average sensitivity of 86.29% for an average FDR of 0.74 per hour.

7.1.4 CONCLUSIONS

In this study, we present the preliminary results for an automated seizure detection system from long-term scalp EEG with deep CNNs. The work comprises of preprocessing, CNN training and validation, and testing with unknown EEG dataset. The proposed CNN model achieved an average sensitivity of 86.29%, an average FDR of 0.74 per hour and an average latency of 2.1 seconds. Overall, the study reveals that deep CNN provides an interesting line of enquiry when identifying seizure and non-seizure EEG records. The system may aid the neurologists in long-term EEG monitoring practice and treatment planning for epilepsy patients. In the future, we intend to utilize a larger amount of seizure data in the deep learning training phase to improve the system performance [156, 157]. It is also noted that the artifact poses a significant challenge for the seizure detection system. Therefore, we also plan to include artifact rejection techniques to reduce the misclassification rate. In addition, further investigation needs to be done to handle the dynamics of the signal sequences in the CNN output [158].

7.2 CLASSIFICATION FOR BRAIN-COMPUTER INTERFACES (BCIs)

7.2.1 INTRODUCTION

Brain-computer interface (BCI) is an emerging area of research with enormous scope in medical as well as non-medical applications. BCI is a communication channel to control external devices or applications with brain signals directly. It enables the users to interact with the environment independent of peripheral nerves and muscles, using large-scale neural signaling mechanisms in the brain. BCI finds its applications in various fields, such as medicine, neuroergonomics, smart environment, education and self-regulation, games and entertainment [34], and security and authentication [33]. In recent years, extensive work has been accomplished regarding the application of BCI in biomedical systems: detection and diagnosis of tumors and neurological disorders [159], neurorehabilitation [32], assisting activities of daily living (ADL), and in the design of neuroprosthetics.

A BCI system consists of the following parts: signal acquisition, signal preprocessing, feature extraction, and classification [160]. In signal acquisition, the electrical activity produced by the brain is recorded by performing the specified voluntary task. There are generally two methods of signal acquisition procedures: invasive and non-invasive. The invasive signal acquisition involves surgical interventions and electrodes are placed on the surface of the brain. In non-invasive acquisition, the signal is collected without any surgical interventions. The former provides better signal quality, but the latter is preferred owing to its ease of implementation. Functional magnetic resonance imaging (fMRI), functional near-infrared spectroscopy (fNIRS), magnetoencephalography (MEG), and electroencephalogram (EEG) are some of the widely applied non-invasive signal acquisition methods [160]. EEG is commonly applied for BCI problems due to its non-invasiveness, economic viability, ease of usage, and portability. EEG is a measurement of the electrical signals produced by the brain by placing electrodes on the scalp. The evoked potential (EP) is a widely studied category of BCI electrical signals. EPs are the electrical signals produced by the brain in response to stimuli. EPs can be divided into three categories based on the type of stimuli: visual (VEP), auditory (AEP), and somatosensory (SEP). In this study, we are analyzing a VEP dataset. When the stimulation frequency is low (less

than 4 Hz), the signals are referred to as transient VEPs, and when the stimulation frequency is high (greater than 6 Hz), the signals are referred to as steady-state VEPs (SSVEPs) [161]. This is because if a series of identical visual stimuli are presented at a high frequency, the brain enters into a steady-state and generates neuronal signals at the same (or multiples of) frequency as that of the stimulus, without any transient phase [162]. Here we apply the SSVEP dataset analyzed in the study designed by Oikonomou et al. in [163], available from the Physionet database [164]. Followed by signal acquisition, the data is preprocessed. Numerous features can be extracted from the preprocessed signals including wavelet coefficients, power spectral density, amplitude parameters, etc. The extracted feature vector is then applied to train a classifier. The classifier aims to identify the user's intended action into one of the predefined classes based on the feature values. Various classification algorithms have been implemented in the literature, such as k -nearest neighbors (k -NN) [25], multilayer perceptron (MLP) [165], decision trees [126], linear discriminant analysis (LDA) [166], and SVM [27]. The basic model of a BCI study is illustrated in Fig. 7.2.1.

Figure 7.2.1: The basic model of a BCI study. The training part is shown in red.

There are various challenges faced by BCI studies. The design of the experiment, data acquisition, and training requires a high degree of organization and timing. Developing a large database is quite tedious and expensive. The variability of the brain signals produced across subjects is also quite high. A system trained on a particular set of subjects may not work efficiently on a different subject. The quality of the data acquired is another concern. The non-invasive methods are prone to artifacts such as electrocardiograms (ECGs), eye blinks, muscle artifacts, etc. Since the amplitude of brain signal recordings is in the range of microvolts, artifacts pose a significant challenge in the classification process. Moreover, the desired features that provide the best classification results are unknown in most of the stud-

ies. In general, such highly predictive features are found by a trial and error method, i.e., evaluating all the features and then selecting the best set of features.

In this study, we optimize the classification component of a BCI system by applying deep learning techniques. Deep learning techniques are shown to provide higher classification accuracies in the literature. Moreover, deep networks can detect underlying structures or features from the raw data. This would help us to reduce the dependency on the feature extraction part of the BCI system. We analyze the performance of traditional classification algorithms with the newer, sophisticated deep learning methods of convolutional neural networks (CNN) [29] and recurrent neural networks (RNN) [31]. In RNN, we implement the long short-term memory (LSTM) architecture [30].

CNN has been applied for SSVEP studies: navigation in virtual environment (VE) [127], LCD flickering frequency identification [128], classification of SSVEPs [129], and other BCI studies like P300 detection [130]. RNN has been applied in BCI studies for real-time ocular artifact suppression [167] and predicting imagined mental tasks [168]. LSTM has also been applied to BCI studies such as automated visual classification [169] and smart sensor manufacturing environment application [170]. The comparison of different classification algorithms is presented in the literature [163, 171–175]. Specifically, [163] and [171] provide an extensive study of different classifiers; however, CNN and LSTM were not included in the study. We compare CNN and LSTM with various traditional classification algorithms, namely, k -NN [25], MLP [165], decision trees [126], and SVM [27]. We also test the traditional classifiers in combination with sequential forward selection (SFS) [7] of features. We analyze the accuracy of the various classification algorithms by performing a leave-one-subject-out cross-validation (LOSO-CV) on an SSVEP dataset [163]. In this study, we demonstrate the superiority of deep learning algorithms of CNN and LSTM over the traditional classification methods. SVM with Gaussian kernel employing SFS of features provided the best accuracy of 66.09% amongst the traditional classifiers, while CNN provided a superior accuracy of 69.03%.

7.2.2 METHODOLOGY

7.2.2.1 BCI DATASET AND PREPROCESSING

In this study, we analyze the SSVEP dataset introduced in [163], available from the Physionet database [164]. The experiment consists of 256 channel SSVEP recordings of 11 subjects (8 males and three females) with a sampling frequency of 250 Hz. The stimuli are non-overlapping flickering lights from five magenta boxes with frequencies 6.66 Hz, 7.5 Hz, 8.5 Hz, 10 Hz, and 12 Hz. The objective is to distinguish between the different stimuli frequencies by analyzing the SSVEP signals. The EEG signals are normalized, bandpass filtered between 5-48 Hz, and notch filtered at 50 Hz as suggested in [163].

7.2.2.2 APPROACH

We investigate various traditional classification algorithms: k -NN [25], decision trees [126], MLP [165], and SVM [27]. The different parameters of the traditional classifiers explored are detailed in Table 7.2.1. We evaluate two types of deep learning methods: CNN and RNN with LSTM architecture. Table 7.2.3 details the different optimized parameters applied for CNN. The block diagram of the CNN structure is illustrated in Fig. 7.2.2.

Figure 7.2.2: The block structure of the implemented CNN network.

In RNN, we implement a bidirectional LSTM. The optimization of LSTM parameters is performed sequentially. First, we optimized the number of units (in the range from 10 to 1500). Later we choose the number of epochs and best weight's initialization by applying LOSO-CV on the training set, for each step of the main LOSO-CV. The optimized parameter values for LSTM are detailed in Table 7.2.4.

The performance of classifiers primarily depends on feature extraction and feature selection. We performed channel selection and feature extraction,

Table 7.2.1: The different parameter values evaluated for traditional classifiers

Classifier	Parameter	Values
k -NN	Distance measure k	ED 1 to 10
C4.5	Number of leaves	Optimized by CV on training data
SVML	Kernel C	Linear 2^{-5} to 2^5
SVMG	Kernel C γ	Gaussian 2^{-5} to 2^5 2^{-5} to 2^5
AdaBoost (C4.5)	Weak learner Number of learners	C4.5 100
AdaBoost (LDA)	Weak learner Number of learners	LDA 100
MLP	Number of input layer neurons	257
	Number of hidden layer neurons	10, 20, 30, 50, 100
	Number of output layer neurons	5

Table 7.2.2: The different sets of parameter values evaluated for CNN optimization

Parameters	Values
Number of convolution layers	1, 2, 3
Number of pooling layers	1, 2, 3
Number of fully connected layers	1
Number of convolution filters	4, 8, 16, 32, 64
Dimension of convolution filters	1×3 , 1×4 , 1×5 , 1×6 , 1×8 , 1×16 each
Number of hidden layer neurons	100, 500, 1000, 1500, 2000, 3000
Activation functions	ReLu, tanh, ELU, softplus, softsign

as suggested by Oikonomou et al. in [163]. SSVEP potential manifests most prominently on the EEG channels placed over the occipital lobe. Therefore,

Table 7.2.3: Optimized parameters of the CNN network

Parameters	Value
Number of convolution layers	1
Number of pooling layers	1
Number of fully connected layers	1
Number of convolution filters	16
Dimension of convolution filters	1×4 each
Number of hidden layer neurons	2000
Drop probability	0.5
Activation function	ReLu

Figure 7.2.3: The model of the implemented bidirectional LSTM network.**Table 7.2.4:** Optimized parameters of the LSTM network

Parameters	Value
Type	Bidirectional
Number of layers	1
Learning rate	0.001
Number of units	1100
Maximum number of epochs	200
Number of initializations	100

we extract the EEG signal from the channel-126 that was located directly above the occipital lobe. We implemented a non-parametric approach of estimating the power spectral density of the SSVEP EEG signal by applying

the Welch method. The Welch method windows the signal into overlapping segments computes periodogram of each window and estimates the power spectral density by averaging the periodograms [176, 177]. It has been reported in [174] that the Welch method is an effective feature extraction method for SSVEP problems. We applied the MATLAB implementation of the Welch method by applying the ‘pwelch’ function with a Hamming window with 50% overlap between adjacent segments and 512 points discrete Fourier transform (DFT). We used the aforementioned power spectral density values computed in the frequency range from 0 to 125Hz, consisting of 257 points as the feature vector in our classification framework. We apply this 257 point feature vector as the input of classification algorithms. Feature selection is a widely applied technique to boost the performance of classifiers. Here we also compare the results for traditional classifiers with feature selection based on correlation. Very often, we are working with data consisting of many features, where only part of them are important to solve our problem. There can be redundant attributes that provide the same information as other features. We aim to select and apply only “interesting” attributes or in other words, select such a subset of features which gives the highest possible classification accuracy. There exist many types of feature selection algorithms like filters, wrapper, etc. In this study, we develop a feature ranking based on correlation coefficient values. Given a feature set $X = \{x_i | i = 1, 2, \dots, N\}$, we find a subset $Y_M = \{x_{i1}, x_{i2}, \dots, x_{iM}\}$, with M , N that optimizes the objective function, $J(Y)$, given by:

$$J(Y_M) = \frac{\sum_{i=1}^M \rho_{ic}}{\sum_{i=1}^M \sum_{j=i+1}^M \rho_{ij}}, \quad (7.1)$$

where ρ_{ic} is the correlation between feature i and the class label c , and ρ_{ij} is the correlation coefficient between features i and j . Subsequently, we apply SFS [7] for the different features. Here we start from an empty set and sequentially add up the next most important attribute. This method will provide the best subset of features for a selected classifier.

The traditional classifiers are implemented with MATLAB 2016a Statistics and Machine Learning Toolbox by applying the default parameters. CNN and LSTM are implemented using Tensorflow r0.12 [132] with Tesla K40 graphical processing unit (GPU) on Ubuntu 16.04. We perform a LOSO-

CV on the 11 patient SSVEP dataset, and the accuracies are averaged. We also perform a nested 3-fold cross-validation on the training set for optimizing the parameters of traditional classifiers as well as selecting the number of iterations for deep learning techniques.

7.2.3 RESULTS AND DISCUSSION

The different classification algorithms were tested with the SSVEP dataset, and the LOSO-CV accuracies are tabulated in Table 7.2.5. The deep learning methods of CNN and LSTM shows superior accuracies in comparison with the other traditional classification algorithms. SVM has also shown better accuracies for certain subjects when applied along with feature selection. This is mainly attributed to the fact that the performance of shallow classifiers predominantly depends on the quality of the feature selection procedure. The feature selection process improved the performance of traditional classifiers significantly, except for MLP. The deep learning methods outperformed the traditional classifiers even with the inclusion of the feature selection process. Deep learning methods can perform with minimal or no feature extraction, and the desired features are automatically learned by the network. This makes deep learning methods suitable for studies in which the desired features are unknown.

The comparison studies between various classifiers have been presented in [163, 171–175]. In [174], Carvalho et al. made a comparison between SVM, LDA, and extreme learning machine (ELM), which is a type of MLP. ELM yields the best results in this study. Oikonomou et al. proved SVM as the best classifier for BCI problems in comparison with decision trees, LDA, k -NN, and Naive Bayes classifier [163]. In [175], Singla et al. made a comparison between SVM and artificial neural networks (ANN). ANN was tested in two different configurations: feed-forward backpropagation and cascade-forward backpropagation. In this study, SVM provides better classification results compared to ANN. Similarly, an extensive study of classification algorithms was performed by Lotte et al. in [171] and presented SVM as the superior classifier. In [173] Krusienski et al. showed that LDA performs better than SVM for P300 speller classification. In this study, we have made a comparison between deep learning methods and various traditional classifiers for BCI problems. CNN is shown to outperform all the traditional

methods available in the literature. It is a proven fact that as the size of the dataset increases, deep learning techniques tend to perform better than traditional classifiers [129]. Currently, we have only proved the superiority of deep learning on a relatively small dataset. In comparison with the traditional algorithms, deep learning techniques are complex, computationally expensive, and require a large amount of data. The new development of powerful GPUs and cloud-based shared services have made deep learning systems cost-effective. Determining the different parameters of deep learning systems is still a major challenge.

7.2.4 CONCLUSIONS

In this study, we have demonstrated the superiority of deep learning methods for the classification part of BCI systems in comparison with the traditional classification algorithms explored in the literature. We have also shown that the application of deep learning methods has reduced the dependency on the feature extraction part of BCI systems. In our future work, we intend to evaluate different datasets to substantiate the performance of deep learning. Also, in order to study the various standard features of a particular BCI dataset, we plan to extract the features by the deep learning methods. This would provide a better understanding of the relation between EEG and the specified BCI tasks.

Table 7.2.5: LOO-CV classification accuracies of different classifiers on the SSVEP dataset

Classifiers	S01	S02	S03	S04	S05	S06	S07	S08	S09	S10	S11	Mean prediction accuracy
<i>k</i> -NN	52.17%	62.61%	39.13%	28.26%	23.48%	52.17%	31.30%	13.04%	83.48%	57.89%	64.35%	46.17%
<i>k</i> -NN SFS	78.26%	63.48%	36.23%	35.87%	25.22%	34.78%	61.74%	17.39%	78.26%	57.02%	73.04%	51.03%
C4.5	69.57%	53.91%	28.99%	42.39%	20.87%	30.43%	54.78%	20.29%	87.83%	58.77%	75.65%	49.41%
C4.5 SFS	66.67%	73.04%	26.09%	34.78%	18.26%	39.13%	66.09%	21.74%	85.22%	65.79%	79.13%	52.36%
AdaBoost C4.5	76.81%	44.35%	21.74%	26.09%	23.48%	22.83%	61.74%	21.74%	84.35%	45.61%	63.48%	44.75%
AdaBoost C4.5 SFS	66.67%	44.35%	21.74%	26.09%	23.48%	22.83%	61.74%	18.84%	77.39%	47.37%	63.48%	44.75%
AdaBoost LDA	88.41%	76.52%	39.13%	42.39%	17.39%	79.35%	31.30%	27.54%	93.91%	86.84%	61.74%	58.59%
AdaBoost LDA SFS	94.2%	82.61%	39.13%	55.43%	25.22%	80.43%	57.39%	24.64%	91.30%	87.72%	78.26%	65.12%
MLP	89.86%	88.70%	30.43%	47.83%	27.83%	83.70%	58.26%	27.54%	98.26%	96.49%	41.74%	62.78%
MLP SFS	88.41%	77.39%	43.48%	55.43%	23.48%	72.83%	68.70%	26.09%	88.70%	87.72%	84.35%	65.14%
SVML	75.36%	67.83%	21.74%	45.65%	23.48%	22.83%	53.91%	24.64%	96.52%	69.30%	93.91%	54.11%
SVML SFS	72.46%	79.13%	28.99%	45.65%	27.83%	41.30%	70.43%	20.29%	88.70%	88.60%	88.70%	59.28%
SVMG	89.86%	93.91%	27.54%	56.52%	23.48%	79.35%	49.57%	18.84%	97.39%	93.86%	86.09%	65.13%
SVMG SFS	89.86%	93.91%	33.33%	56.52%	22.61%	70.65%	70.43%	27.54%	97.39%	93.86%	70.90%	66.09%
CNN	90.34%	91.22%	42.24%	57.14%	27.01%	85.71%	57.32%	26.14%	98.27%	94.32%	89.65%	69.03%
LSTM	92.75%	91.30%	39.13%	53.26%	21.73%	88.04%	54.78%	17.39%	96.52%	93.91%	86.95%	66.89%

“It is strange that only extraordinary men make the discoveries, which later appear so easy and simple.”

Georg C. Lichtenberg

8

Conclusions and Future directions

This chapter presents the conclusions of our research (see Section 8.1) with the directions for future works (see Section 8.2).

8.1 CONCLUSIONS

Electroencephalography (EEG) is a widely applied diagnostic technique for the diagnosis of neural disorders. Epilepsy is one of the most prevalent neurological disorders in the world. Epilepsy diagnosis based on EEG events has been proved to be highly specific, but less sensitive. However, due to the lack of trained neurologists all over the world, especially in developing countries, a significant amount of time is spent on analyzing long EEG recordings. Still, the gold standard of EEG evaluation for epilepsy remains to be the visual inspection performed by experts, despite the technological advancement in the health care industry. This was predominantly due to the lack of a reliable and clinically validated epilepsy diagnostic system based on EEG. In this thesis, we have developed a deep learning-based reliable and clinically validated epilepsy diagnostic system based on the identification of interictal epileptiform patterns. The previous literature studies, methods, and performance indices have been detailed in Chapters 2 and 3.

We have developed an efficient, reliable, and clinically validated epilepsy diagnostic system by interictal pattern identification in EEG. The system is evaluated as an offline solution in our research. This system could be made online with an addition of a Graphical User Interface (GUI) and cloud processing platforms by configuring the EEG monitoring equipment at the hospitals. This system would aid the neurologists to identify epileptiform transients reliably and expeditiously, and thereby diagnose epilepsy effectively. Consequently, more time could be diverted towards providing proper treatment and care rather than on the diagnostic process, thereby improving patient care.

In our first step, we performed the preprocessing and understanding of the dataset (see Chapter 4). In any machine learning-based system, the quality of data determines the generalization and performance of the system. In our study, we evaluate datasets from different hospitals, namely, Massachusetts General Hospital (MGH), Boston, Medical University of South Carolina (MUSC), and National University Hospital (NUH), Singapore. The MGH dataset consisted of 561 EEGs with 18,164 annotated IEDs (by two experts). We applied the dataset from MGH for training, and the other two datasets were applied for evaluation. The MUSC data consisted of 235

annotated IED segments (by 18+ experts), and the NUH data consisted of 100 EEGs labeled as epileptic or non-epileptic. Most of the studies in the literature have been only cross-evaluated on a single dataset. Here, we have evaluated our systems on non-correlated datasets from multiple centers, which make the results more reliable.

The MGH data preprocessed by applying a 5-step preprocessing module: filtering, down-sampling, power-based artifact detection, removal of missing electrode segments, and application of an EEG montage. We extracted the IED and the non-IED segments as 500-millisecond waveforms (64 samples) from the preprocessed data. These were further applied to develop the IED detector. Currently, a standard quantitative definition of IEDs is not available in the literature. Here, we quantized the different morphological characteristics of IEDs and had put forward quantized IED characteristics definitions. It is also essential to understand the variability of IEDs across patients while developing an IED detector. We performed correlation analysis on IEDs within patients and across patients. We observed that the IEDs within a patient tends to exhibit a reasonable similarity (with a mean correlation value of 0.43) and the IEDs across patients vary significantly (with a mean correlation value of 0.21).

As a second step, we developed an efficient IED detector (see Chapter 5). From the IED correlation studies, it became evident that the IED detector development is quite challenging due to high variability across patients. Moreover, a standard feature set for IEDs has not yet been formulated. Consequently, we developed an IED detector based on raw-EEG with minimal preprocessing. This would also help us to reduce the dependency on hand-crafted features, that have been put forward in most of the studies in the literature. We explored different machine learning techniques and deep learning techniques for IED detection. We applied Template Matching (TM), traditional classifiers (k -Nearest Neighbors (k -NN), Support Vector Machines (SVM), Classification and Regression Trees (CART), Linear Discriminant Analysis (LDA), and ensemble classifier systems), and Convolutional Neural Networks (CNN) for IED detection. Since the performance of the TM system depends on the quality of the template library, we applied a clustering system comprising of Affinity Propagation (AP) and Dynamic Time Warping (DTW) to extract high-quality template library. The different traditional

classifier systems were optimized for the respective hyperparameters by applying Bayesian optimization. Similarly, the hyperparameters of the CNN system were optimized by performing 4-fold nested cross-validation on the training set. The CNN system performed superior to the other traditional system with a false detection rate of 0.2 ± 0.12 per minute and 0.92 ± 0.83 per minute for sensitivity thresholds of 80% and 90%, respectively. This result is superior to those reported in the literature (see Tables 3.3.1, 3.3.2, 3.3.3, and 3.3.4). We further evaluated the performance of our system on the MUSC dataset. Although the dataset is small (235 IEDs), the number of cross-annotations is very high (18+). This makes it useful for evaluating the majority consensus of neurologists over IEDs. We also evaluated Persyst, a commercial EEG analysis software product, on this dataset. The proposed system achieved satisfactory results on the MUSC dataset, outperformed Persyst at high annotator threshold settings (see Fig. 5.3.8), and was found to be non-inferior with the consensus of experts (see Fig. 5.3.9).

In our final step, we develop the EEG classification system based on the features extracted from the IED detector (see Chapter 6). We explored different feature sets with different system configurations. The system with an ensemble of IED detectors (TM and CNN IED detectors) at the waveform-level classification was found to be effective for EEG classification. The features employed were the number of CNN detections per second and the mean correlation distance from the template libraries (IED and background template library). The simple thresholding system performed quite reasonably. However, due to the dependency on an absolute threshold, it would be difficult to apply it across different subjects and different types of equipment. As a result, we developed a multi-feature classifier based system for EEG classification. We implemented an SVM with Gaussian kernel for EEG classification. The thresholding system achieved superior performance with a mean cross-validation balanced accuracy of 81.61% (Model 1) on the MGH dataset. The ensemble IED detector system with an SVM for EEG classification achieved a satisfactory result of 79.33% (Model 3) on the MGH dataset. While evaluating the NUH dataset, the ensemble system performed superior to the simple thresholding system with a balanced accuracy of 88%. This evaluation was performed as a clinical validation: the labels were blind to the system, and performance was evaluated in comparison with the clinician's

report at NUH.

Finally, we have extended the methodology for identifying EEG paradigms. The CNN system was applied for detecting seizures and Brain-Computer Interface (BCI) tasks. In seizure detection, we developed a generalized seizure detector with satisfactory detection performance. The system achieved an average sensitivity of 86.29%, an average FDR of 0.74 h-1 and an average latency of 2.1 seconds (with a fixed delay of 1.25 seconds). The most salient feature of the system is the non-dependency on features. Similar to IEDs, quite a lot of feature definitions have been formulated in the literature, and a standardized set is unknown. In the BCI study, we have evaluated the performance of CNN in comparison with the traditional classifier algorithms in the literature. Here we compare the conventional system with and without feature selection. The CNN system outperformed the different conventional systems with an added advantage of non-dependency on feature extraction, evaluation, and selection.

8.2 FUTURE WORK

In future work, we aim to enhance the effectiveness of the IED detector and EEG classification system. The future steps are:

- Implement artifact rejection techniques to reduce misclassifications.
- Evaluate the system on various other datasets for improving reliability.
- Make the system adaptable to different montage, equipment, and clinical settings.
- Develop a multi-observer system rather than a multi-consensus system.

8.2.1 ARTIFACT REJECTION TECHNIQUES

It is an accepted fact that artifacts in the scalp EEG pose a challenging problem, considering its low amplitude. The EEG artifacts could be subject dependent, environmental, or equipment-based. Also, it is crucial to develop an efficient, subject-specific artifact rejection mechanism in order to reduce misclassifications. In our current system, we have considered artifacts as backgrounds. However, we plan to incorporate an artifact rejection module to decrease the number of misclassifications.

8.2.2 DEVELOPMENT OF AN ADAPTIVE SYSTEM

The present system has been applied in common average referential (CAR) montage. To develop an adaptable system, we further extend our analyses to other montages such as bipolar, monopolar, etc. In addition, we plan to modify our system to receive data from different EEG equipment and diverse clinical settings such as the epilepsy monitoring unit, neonatal EEG recording, etc.

8.2.3 VALIDATION ON MULTIPLE DATASETS

Although the proposed system has been evaluated on two other non-correlated datasets from different centers, it needs to be tested on data from more centers in order to substantiate the dependability of our system. We have already acquired data and initiated the evaluation from more centers across the globe:

- National Neuroscience Institute (NNI), Singapore
- Temple University Hospital (TUH), Pennsylvania
- Lokmanya Tilak Municipal General Hospital, India

8.2.4 TOWARDS MULTI-OBSERVER SYSTEMS

We developed the IED detector system based on the IED annotations with the agreement of both the annotators. In the literature, most of the studies with multiple annotators have applied multiple consensus of annotators to develop the IED detector systems. Moreover, there are quite a lot of studies pertaining to comparing the Inter-rater Agreement (IRA) across many annotators. Almost all of them have concluded that the IRA is quite low for IED annotations. This result is unreliable due to the lack of a standard database and definition for IEDs. Moreover, most of the IRA studies does not perform any consistency evaluation on the experts. Besides, only the consensus of a single or few neurologists is applied for the diagnostic task across different centers in the world. Each expert is subjective to his/her experience and expertise, and there could be variations across different geographical areas, lifestyles, and environmental conditions.

In our prospective work, we intend to implement a multi-observer system rather than a multi-consensus system. This would provide more flexibility to the system in terms of annotations and the reliability of the neurologist. Moreover, the multi-observer system could be adaptively developed by adding consensus of different experts across the globe.

We have performed a preliminary analysis of the multi-observer system. In order to evaluate the multi-observer system, we re-annotated our MGH dataset (see Section 4.1.1) with seven annotators. Annotations are always time consuming and expensive. Hence, in order to reduce the budget, we extracted IED templates from each of the 84 subjects, independently. This was performed by applying Affinity Propagation (AP) in conjunction with Dynamic Time Warping (DTW). We extracted 2063 templates from the 84 subjects with 18,164 annotated IEDs. The IED waveforms are shown as entire EEG (19-channels) 10-second segment for the experts to annotate the data. In order to prevent skewing random 10-second segments from non-epileptic EEGs were shown intermittently. The task for the experts was to identify whether the particular segment exhibited IEDs. The annotations were 0 or 1, with 1 indicating the presence of IEDs. Additionally, in order to investigate the consistency of an annotator, few of the IED segments (5% of the IEDs) were randomly reiterated for re-annotation. This task was kept hidden from the knowledge of the experts. We obtained the annotations from 7 experts for this study. We applied the different EEG preprocessing and extraction techniques elaborated in Chapter 4.

After data extraction, we assessed the quality of the annotations by performing two different evaluations. The first evaluation was based on the ability to distinguish non-IED waveforms that were shown intermittently. The accuracy of the seven annotators for identifying the non-IED waveforms are presented in Table 8.2.1. The different annotators exhibited superior expertise in distinguishing IEDs vs. non-IED waveforms. In most of the studies, only this analysis is performed to evaluate the annotations.

Next we performed the consistency analysis. Here we utilize the duplicate annotations of the same IEDs and verify whether the annotator decision has been consistent. We derive the consistency values for each IED that were applied for re-evaluation. The consistency value c_i for an IED i is computed as $c_i = \frac{\max\{n(1), n(0)\}}{N}$, where $n(0)$ is the number of times the segment has

Table 8.2.1: The accuracy of the different annotators for distinguishing IEDs vs. non-IED waveforms.

Annotator	Accuracy
A1	99.55%
A2	98.48%
A3	99.39%
A4	99.39%
A5	99.09%
A6	98.48%
A7	95.91%

been labeled ‘0’, $n(1)$ is the number of times the segment has been labeled ‘1’, and N is the total number of times the IED has been reshown. Here $N = n(0) + n(1)$ as we apply binary labeling. The consistency score for each annotator was computed by averaging the consistency values. The results are illustrated in Table 8.2.2. Annotators A6 and A7 were more consistent amongst the seven annotators. The consistency score directly affects the IRA analysis. This makes the consistency evaluation essential for IRA studies.

Table 8.2.2: Consistency score of the different annotators.

Annotator	Consistency score
A1	89.22%
A2	89.31%
A3	92.84%
A4	79.46%
A5	89.31%
A6	94.02%
A7	95.83%

However, the subjectivity and inconsistency of experts is an inevitable challenge for IED detection. Moreover, the consistency also depends on different factors such as the physical and mental condition of the experts. In order to overcome these challenges, we propose the multi-observer system that aims to model an individual expert’s consensus rather than the crowd consensus. Here, instead of applying the multi-consensus before training the IED detector, we train independent IED detector for the different experts.

Later the decisions from the different annotator systems are combined to produce a weighted decision. The two different approaches are illustrated in Figs. 8.2.1 and 8.2.2.

Figure 8.2.1: Block diagram of the multi-consensus IED detector system.

Figure 8.2.2: Block diagram of the proposed multi-observer IED detector system.

We analyze the effectiveness of the proposed approaches based on Support Vector Machine (SVM)-based IED detector. We have chosen SVM as the number of IEDs with seven annotations is low. Here our objective is to verify the ability of the two different approaches to mimic the expert annotations. The average of the seven expert annotations is applied as the original test labels. Consequently, we employ Root Mean Square Error (RMSE) as the performance measure for comparison. We repeat the evaluation procedure for 100 trials and average the RMSE values.

At each trial, we randomly divide the 84 patient MGH data (2063 annotated IEDs) into two sets: 80% percentage for training and 20% for testing. We employ similar training conditions as elaborated in Chapter 4. The non-IED waveforms were selected randomly from the non-IED database and are kept constant throughout the analysis. We also compute the consistency scores for the different annotators based on the training data. For each trial, we evaluate the eight systems: 6 systems with different annotator consensus (varying from 2 to 7), multi-observer system, and the modified

multi-observer system. In the basic multi-observer system, we apply equal weights for the different annotator systems, whereas the consistency scores are employed as the weights in the modified multi-observer system. The RMSE boxplots across the trials for the different systems are presented in Fig. 8.2.3. The multi-observer systems have shown superior RMSE values in comparison with the multi-consensus systems. We will employ the CNN-based detector once we procure sufficient data points. Moreover, this system would be integrated with the EEG classification system in the future.

Figure 8.2.3: Boxplots for the RMSE values for the different approaches.

List of Publications

1. **Thomas J.**, Jin J., Dauwels J., Cash S.S., and Westover M.B. “Clustering of interictal spikes by dynamic time warping and affinity propagation” in Acoustics, Speech, and Signal Processing (ICASSP), 2016 IEEE International Conference on, pp. 749-753. IEEE, 2016.
2. **Thomas J.**, Jin J., Dauwels J., Cash S.S., and Westover M.B. “Automated epileptiform spike detection via affinity propagation-based template matching” in Engineering in Medicine and Biology Society (EMBC), 2017 39th Annual International Conference of the IEEE, pp. 3057-3060. IEEE, 2017.
3. **Thomas J.**, Maszczyk T., Sinha N., Kluge T., and Dauwels J. “Deep learning-based classification for brain-computer interfaces” in Systems, Man, and Cybernetics (SMC), 2017 IEEE International Conference on, pp. 234-239. IEEE, 2017.
4. D. Tejas, P. Das, T. Pedapati, G. Bramble, V. Muthusamy, H. Samulowitz, Kush R., Yuvaraj R., **J. Thomas**, and J. Dauwels. “Neurology-as-a-Service for the Developing World” in NIPS 2017 Workshop on Machine Learning for the Developing World. arXiv preprint arXiv:1711.06195 (2017).
5. **Thomas J.**, Jin J., Shaju N., Maszczyk T., Sinha N., Dauwels J., Cash S.S., and Westover M.B. P290 “Convolutional Neural Network-based Interictal Epileptiform Discharge Detection” in 26th Annual Computational Neuroscience Meeting (CNS*2017): Part 3. BMC Neuroscience. 2017 Aug;18(1):60. Available from: <https://doi.org/10.1186/s12868-017-0372-1>.
6. **Thomas J.**, Comoretto L., Jin J., Dauwels J., Cash S.S., and Westover M.B. “EEG Classification via Convolutional Neural Network-Based Interictal Epileptiform Event Detection” in Engineering in Medicine and Biology Society (EMBC), 2018 40th Annual International Conference of the IEEE, pp. 3148-3151, 2018.
7. Yuvaraj R., **J. Thomas**, and J. Dauwels. “Hybrid deep convolutional neural network and hidden markov model for automatic seizure detection from long-term scalp EEG” in ICNF Neuroinformatics 2018, (doi: 10.7490/f1000research.1115871.1).

8. Rathakrishnan R., **Thomas J.**, Jin J., Dauwels J., Cash S.S., and Westover M.B. “Machine learning-based epileptic EEG identification system: a multi-center study” in AES 2018 Annual Meeting Abstract Database. AESnet.org.
9. Yuvaraj R., **J. Thomas**, T. Kluge, and J. Dauwels. “A Deep Learning Scheme for Automatic Seizure Detection from Long-Term Scalp EEG” in 2018 52th Asilomar Conference on Signals, Systems, and Computers, IEEE, pp. 368-372, 2018.

PUBLICATIONS UNDER PREPARATION

1. **Thomas et al.** “ Interictal Epileptiform Discharge Detection with Convolutional Neural Networks”, International Journal of Neural Systems, 2018.
2. **Thomas et al.** “Automated Multi-stage Cascade System for Epileptic EEG Identification”, IEEE Journal of Biomedical and Health Informatics, 2018.
3. **Thomas et al.** “Data-driven Quantitative Definition of Epileptiform Spikes”, Clinical Neurophysiology, 2018.
4. **Thomas et al.** “Universal Template library for Epileptic Spikes via Dynamic Time Warping-based Clustering”, IEEE Transactions in Neural Systems and Rehabilitation Engineering, 2018.
5. Yuvaraj R., **J. Thomas**, T. Kluge, and J. Dauwels. “Epileptic seizure detection using convolutional neural network and statistical models from scalp EEG recordings”, Clinical Neurophysiology, 2018.

References

- [1] Epilepsy Foundation of America. *About Epilepsy: The Basics*. URL: <http://www.epilepsy.com/learn/about-epilepsy-basics>. Accessed: 2019-08-01.
- [2] D. J. Thurman, E. Beghi, C. E. Begley, A. T. Berg, J. R. Buchhalter, D. Ding, D. C. Hesdorffer, W. A. Hauser, L. Kazis, R. Kobau, et al. Standards for epidemiologic studies and surveillance of epilepsy. *Epilepsia*, 52(s7):2–26, 2011.
- [3] M. J. England, C. T. Liverman, A. M. Schultz, and L. M. Strawbridge. Epilepsy across the spectrum: Promoting health and understanding.: A summary of the institute of medicine report. *Epilepsy & Behavior*, 25(2):266–276, 2012.
- [4] S. B. Wilson, M. L. Scheuer, R. G. Emerson, and A. J. Gabor. Seizure detection: evaluation of the reveal algorithm. *Clinical Neurophysiology*, 115(10):2280–2291, 2014.
- [5] J. J. Halford, R. J. Schalkoff, J. Zhou, S. R. Benbadis, W. O. Tatum, R. P. Turner, S. R. Sinha, N. B. Fountain, A. Arain, P. B. Pritchard, et al. Standardized database development for eeg epileptiform transient detection: Eegnet scoring system and machine learning analysis. *Journal of neuroscience methods*, 212(2):308–316, 2013.
- [6] O. Khouma, M. L. Ndiaye, S. M. Farsi, J.-j. Montois, I. Diop, and B. Diouf. Comparative methods of spike detection in epilepsy. In *Science and Information Conference (SAI), 2015*, pages 749–755. IEEE, 2015.
- [7] A. W. Whitney. A direct method of nonparametric measurement selection. *IEEE Transactions on Computers*, 100(9):1100–1103, 1971.
- [8] A. L. Samuel. Some studies in machine learning using the game of checkers. *IBM Journal of research and development*, 3(3):210–229, 1959.
- [9] P. Sedgwick. Pearson’s correlation coefficient. *Bmj*, 345:e4483, 2012.
- [10] M. Müller. Dynamic time warping. *Information retrieval for music and motion*, pages 69–84, 2007.

- [11] R. De Maesschalck, D. Jouan-Rimbaud, and D. L. Massart. The mahalanobis distance. *Chemometrics and intelligent laboratory systems*, 50(1):1–18, 2000.
- [12] H. Sakoe and S. Chiba. Dynamic programming algorithm optimization for spoken word recognition. *Acoustics, Speech and Signal Processing, IEEE Transactions on*, 26(1):43–49, 1978.
- [13] T. Rakthanmanon, B. Campana, A. Mueen, G. Batista, B. Westover, Q. Zhu, J. Zakaria, and E. Keogh. Searching and mining trillions of time series subsequences under dynamic time warping. In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 262–270. ACM, 2012.
- [14] J. MacQueen et al. Some methods for classification and analysis of multivariate observations. In *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, volume 1, pages 281–297. Oakland, CA, USA., 1967.
- [15] S. C. Johnson. Hierarchical clustering schemes. *Psychometrika*, 32(3):241–254, 1967.
- [16] J. C. Bezdek, R. Ehrlich, and W. Full. Fcm: The fuzzy c-means clustering algorithm. *Computers & Geosciences*, 10(2):191–203, 1984.
- [17] B. J. Frey and D. Dueck. Clustering by passing messages between data points. *science*, 315(5814):972–976, 2007.
- [18] A. K. Jain and R. C. Dubes. *Algorithms for clustering data*. Prentice-Hall, Inc., 1988.
- [19] M. Halkidi and M. Vazirgiannis. Clustering validity assessment: Finding the optimal partitioning of a data set. In *Data Mining, 2001. ICDM 2001, Proceedings IEEE International Conference on*, pages 187–194. IEEE, 2001.
- [20] D. L. Davies and D. W. Bouldin. A cluster separation measure. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, (2):224–227, 1979.
- [21] P. J. Rousseeuw. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of computational and applied mathematics*, 20:53–65, 1987.
- [22] Y. Liu, Z. Li, H. Xiong, X. Gao, and J. Wu. Understanding of internal clustering validation measures. In *Data Mining (ICDM), 2010 IEEE 10th International Conference on*, pages 911–916. IEEE, 2010.

- [23] K. Fu. *Sequential methods in pattern recognition and machine learning*, volume 52. Academic press, 1968.
- [24] L. Breiman. *Classification and regression trees*. Routledge, 2017.
- [25] L. E. Peterson. K-nearest neighbor. *Scholarpedia*, 4(2):1883, 2009.
- [26] S. Mika, G. Ratsch, J. Weston, B. Scholkopf, and K.-R. Mullers. Fisher discriminant analysis with kernels. In *Neural networks for signal processing IX, 1999. Proceedings of the 1999 IEEE signal processing society workshop.*, pages 41–48. Ieee, 1999.
- [27] M. A. Hearst, S. T. Dumais, E. Osuna, J. Platt, and B. Scholkopf. Support vector machines. *IEEE Intelligent Systems and their Applications*, 13(4):18–28, 1998.
- [28] J. Zhu, H. Zou, S. Rosset, and T. Hastie. Multi-class adaboost. *Statistics and its Interface*, 2(3):349–360, 2009.
- [29] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.
- [30] S. Hochreiter and J. Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.
- [31] R. J. Williams and D. Zipser. A learning algorithm for continually running fully recurrent neural networks. *Neural computation*, 1(2):270–280, 1989.
- [32] A. Budhota, A. Hussain, C. Hughes, C. Hansen, S. Kager, D. A. Vishwanath, C. W. K. Kuah, K. Chua, and D. Campolo. Role of emg as a complementary tool for assessment of motor impairment. In *Biomedical Robotics and Biomechatronics (BioRob), 2016 6th IEEE International Conference on*, pages 692–697. IEEE, 2016.
- [33] K. P. Thomas, A. Vinod, and N. Robinson. Online biometric authentication using subject-specific band power features of eeg. In *Proceedings of the 2017 International Conference on Cryptography, Security and Privacy*, pages 136–141. ACM, 2017.
- [34] K. P. Thomas and A. Vinod. A study on the impact of neurofeedback in eeg based attention-driven game. In *Systems, Man, and Cybernetics (SMC), 2016 IEEE International Conference on*, pages 000320–000325. IEEE, 2016.
- [35] W. O. Tatum. Ellen r. grass lecture: Extraordinary eeg. *The Neurodiagnostic Journal*, 54(1):3–21, 2014.

- [36] Mayo clinic. *Epilepsy*. URL: <https://www.mayoclinic.org/diseases-conditions/epilepsy/symptoms-causes/syc-20350093>. Accessed: 2019-08-01.
- [37] R. S. Fisher, C. Acevedo, A. Arzimanoglou, A. Bogacz, J. H. Cross, C. E. Elger, J. Engel, L. Forsgren, J. A. French, M. Glynn, et al. Ilae official report: a practical clinical definition of epilepsy. *Epilepsia*, 55(4):475–482, 2014.
- [38] W. H. Organization et al. Epilepsy management at primary health level in rural china: Who/ilae/ibe. *A Global Campaign Against Epilepsy Demonstration Project*, 2009.
- [39] M. E. C. Andraus, C. F. Andraus, and S. V. Alves-Leon. Periodic eeg patterns: importance of their recognition and clinical significance. *Arquivos de neuro-psiquiatria*, 70(2):145–151, 2012.
- [40] L. Hirsch and R. Brenner. Periodic discharges and other controversial eeg patterns. *Atlas of EEG in critical care*. Chichester: Wiley-Blackwell, pages 129–160, 2010.
- [41] L. J. Hirsch, R. P. Brenner, F. W. Drislane, E. So, P. W. Kaplan, K. G. Jordan, S. T. Herman, S. M. LaRoche, B. Young, T. P. Bleck, et al. The acns subcommittee on research terminology for continuous eeg monitoring: proposed standardized terminology for rhythmic and periodic eeg patterns encountered in critically ill patients. *Journal of clinical neurophysiology*, 22(2):128–135, 2005.
- [42] M. Zijlmans, P. Jiruska, R. Zelmann, F. S. Leijten, J. G. Jefferys, and J. Gotman. High-frequency oscillations as a new biomarker in epilepsy. *Annals of neurology*, 71(2):169–178, 2012.
- [43] G. Worrell and J. Gotman. High-frequency oscillations and other electrophysiological biomarkers of epilepsy: clinical studies. *Biomarkers in medicine*, 5(5):557–566, 2011.
- [44] J. Jirsch, E. Urrestarazu, P. LeVan, A. Olivier, F. Dubeau, and J. Gotman. High-frequency oscillations during human focal seizures. *Brain*, 129(6):1593–1608, 2006.
- [45] R. J. Staba, C. L. Wilson, A. Bragin, D. Jhung, I. Fried, and J. Engel Jr. High-frequency oscillations recorded in human medial temporal lobe during sleep. *Annals of neurology*, 56(1):108–115, 2004.
- [46] J. W. Britton, L. C. Frey, J. Hopp, P. Korb, M. Koubeissi, W. Lievens, E. Pestana-Knight, and E. L. St. *Electroencephalography (EEG): An introductory text and atlas of normal and abnormal findings in adults, children, and infants*. American Epilepsy Society, Chicago, 2016.

- [47] G. G. Celesia and R.-C. Chen. Parameters of spikes in human epilepsy. *Diseases of the nervous system*, 1976.
- [48] G. Deuschl, A. Eisen, et al. Recommendations for the practice of clinical neurophysiology: guidelines of the international federation of clinical neurophysiology. 1999.
- [49] K. A. Kooi. Voltage-time characteristics of spikes and other rapid electroencephalographic transients semantic and morphological considerations. *Neurology*, 16(1):59–59, 1966.
- [50] P. Gloor. Contributions of electroencephalography and electrocorticography to the neurosurgical treatment of the epilepsies. *Neurosurgical management of the epilepsies*, pages 59–105, 1975.
- [51] S. B. Wilson and R. Emerson. Spike detection: a review and comparison of algorithms. *Clinical Neurophysiology*, 113(12):1873–1881, 2002.
- [52] P. Y. Ktonas and J. R. Smith. Quantification of abnormal eeg spike characteristics. *Computers in biology and medicine*, 4(2):157–163, 1974.
- [53] J. Gotman. Quantitative measurements of epileptic spike morphology in the human eeg. *Electroencephalography and clinical neurophysiology*, 48(5):551–557, 1980.
- [54] C. J. James, R. D. Jones, P. J. Bones, and G. J. Carroll. Detection of epileptiform discharges in the eeg by a hybrid system comprising mimetic, self-organized artificial neural network, and fuzzy logic stages. *Clinical Neurophysiology*, 110(12):2049–2063, 1999.
- [55] C. F. Boos, F. M. de Azevedo Geovani R. Scolaro, and V. P. Maria do Carmo. *Automatic Detection of Paroxysms in EEG Signals using Morphological Descriptors and Artificial Neural Networks*. INTECH Open Access Publisher, 2011.
- [56] Y.-C. Liu, C.-C. K. Lin, J.-J. Tsai, and Y.-N. Sun. Model-based spike detection of epileptic eeg data. *Sensors*, 13(9):12536–12547, 2013.
- [57] Z. H. İnan and M. Kuntalp. A study on fuzzy c-means clustering-based systems in automatic spike detection. *Computers in Biology and Medicine*, 37(8):1160–1166, 2007.
- [58] A. Tzallas, P. Karvelis, C. Katsis, D. Fotiadis, S. Giannopoulos, and S. Konitsiotis. A method for classification of transient events in eeg recordings: application to epilepsy diagnosis. *Methods of Information in Medicine*, 45(6):610–621, 2006.

- [59] T. P. Exarchos, A. T. Tzallas, D. I. Fotiadis, S. Konitsiotis, and S. Giannopoulos. Eeg transient event detection and classification using association rules. *IEEE Transactions on Information Technology in Biomedicine*, 10(3):451–457, 2006.
- [60] N. Acir, I. Oztura, M. Kuntalp, B. Baklan, and C. Guzelis. Automatic detection of epileptiform events in eeg by a three-stage procedure based on artificial neural networks. *IEEE Transactions on Biomedical Engineering*, 52(1):30–40, 2005.
- [61] M. Adjouadi, D. Sanchez, M. Cabrerizo, M. Ayala, P. Jayakar, I. Yaylali, and A. Barreto. Interictal spike detection using the walsh transform. *IEEE Transactions on Biomedical Engineering*, 51(5):868–872, 2004.
- [62] M. Adjouadi, M. Cabrerizo, M. Ayala, D. Sanchez, I. Yaylali, P. Jayakar, and A. Barreto. A new mathematical approach based on orthogonal operators for the detection of interictal spikes in epileptogenic data. *Biomedical sciences instrumentation*, 40:175–180, 2003.
- [63] D. Flanagan, R. Agarwal, Y. Wang, and J. Gotman. Improvement in the performance of automated spike detection using dipole source features for artefact rejection. *Clinical Neurophysiology*, 114(1):38–49, 2003.
- [64] T. Sugi, M. Nakamura, A. Ikeda, and H. Shibasaki. Adaptive eeg spike detection: determination of threshold values based on conditional probability. *Frontiers of Medical and Biological Engineering*, 11(4):261–277, 2001.
- [65] M. A. Black, R. D. Jones, G. J. Carroll, A. A. Dingle, I. M. Donaldson, and P. J. Parkin. Real-time detection of epileptiform activity in the eeg: a blinded clinical trial. *Clinical EEG and Neuroscience*, 31(3):122–130, 2000.
- [66] S. B. Wilson, C. A. Turner, R. G. Emerson, and M. L. Scheuer. Spike detection ii: automatic, perception-based detection and clustering. *Clinical neurophysiology*, 110(3):404–411, 1999.
- [67] B. Ramabhadran, J. D. Frost Jr, J. R. Glover, and P. Y. Ktonas. An automated system for epileptogenic focus localization in the electroencephalogram. *Journal of clinical neurophysiology*, 16(1):59–68, 1999.
- [68] R. Benlamri, M. Batouche, S. Rami, and C. Bouanaka. An automated system for analysis and interpretation of epileptiform activity in the eeg. *Computers in biology and medicine*, 27(2):129–139, 1997.

- [69] W. Webber, B. Litt, K. Wilson, and R. Lesser. Practical detection of epileptiform discharges (eds) in the eeg using an artificial neural network: a comparison of raw and parameterized eeg data. *Electroencephalography and clinical Neurophysiology*, 91(3):194–204, 1994.
- [70] T. Pietilä, S. Vapaakoski, U. Nousiainen, A. Värri, H. Frey, V. Häkkinen, and Y. Neuvo. Evaluation of a computerized system for recognition of epileptic activity during long-term eeg recording. *Electroencephalography and clinical neurophysiology*, 90(6):438–443, 1994.
- [71] A. A. Dingle, R. D. Jones, G. J. Carroll, and W. R. Fright. A multi-stage system to detect epileptiform activity in the eeg. *IEEE Transactions on Biomedical Engineering*, 40(12):1260–1268, 1993.
- [72] W. E. Hostetler, H. J. Doller, and R. W. Homan. Assessment of a computer program to detect epileptiform spikes. *Electroencephalography and clinical neurophysiology*, 83(1):1–11, 1992.
- [73] J. Gotman and L.-Y. Wang. State dependent spike detection: validation. *Electroencephalography and clinical neurophysiology*, 83(1):12–18, 1992.
- [74] J. Gotman and L. Wang. State-dependent spike detection: concepts and preliminary results. *Electroencephalography and clinical Neurophysiology*, 79(1):11–19, 1991.
- [75] P. G. De Oliveira, C. Queiroz, and F. L. Da Silva. Spike detection based on a pattern recognition approach using a microcomputer. *Electroencephalography and clinical neurophysiology*, 56(1):97–103, 1983.
- [76] J. Gotman, J. Ives, and P. Gloor. Automatic recognition of inter-ictal epileptic activity in prolonged eeg recordings. *Electroencephalography and clinical neurophysiology*, 46(5):510–520, 1979.
- [77] J. Gotman and P. Gloor. Automatic recognition and quantification of interictal epileptic activity in the human scalp eeg. *Electroencephalography and clinical neurophysiology*, 41(5):513–529, 1976.
- [78] R. Sankar and J. Natour. Automatic computer analysis of transients in eeg. *Computers in biology and medicine*, 22(6):407–422, 1992.
- [79] S. S. Lodder and M. J. van Putten. A self-adapting system for the automated detection of inter-ictal epileptiform discharges. *PloS one*, 9(1):e85180, 2014.

- [80] A. Nonclercq, M. Foulon, D. Verheulpen, C. De Cock, M. Buzatu, P. Mathys, and P. Van Bogaert. Cluster-based spike detection algorithm adapts to interpatient and inpatient variation in spike morphology. *Journal of neuroscience methods*, 210(2):259–265, 2012.
- [81] E. I. Zacharaki, I. Mporas, K. Garganis, and V. Megalooikonomou. Spike pattern recognition by supervised classification in low dimensional embedding space. *Brain informatics*, 3(2):73–83, 2016.
- [82] V. Chavakula, I. S. Fernández, J. M. Peters, G. Popli, W. Bosl, S. Rakhade, A. Rotenberg, and T. Loddenkemper. Automated quantification of spikes. *Epilepsy & Behavior*, 26(2):143–152, 2013.
- [83] K. Indiradevi, E. Elias, P. Sathidevi, S. D. Nayak, and K. Radhakrishnan. A multi-level wavelet approach for automatic detection of epileptic spikes in the electroencephalogram. *Computers in Biology and Medicine*, 38(7):805–816, 2008.
- [84] F. I. Argoud, F. M. De Azevedo, J. M. Neto, and E. Grillo. Sade3: an effective system for automated detection of epileptiform events in long-term eeg based on context information. *Medical and Biological Engineering and Computing*, 44(6):459–470, 2006.
- [85] H. S. Liu, T. Zhang, and F. S. Yang. A multistage, multimethod approach for automatic detection and classification of epileptiform eeg. *IEEE Transactions on Biomedical Engineering*, 49(12):1557–1566, 2002.
- [86] H. J. Carey, M. Manic, and P. Arsenovic. Epileptic spike detection with eeg using artificial neural networks. In *Human System Interactions (HSI), 2016 9th International Conference on*, pages 89–95. IEEE, 2016.
- [87] Y. Song and J. Zhang. Automatic recognition of epileptic eeg patterns via extreme learning machine and multiresolution feature extraction. *Expert systems with applications*, 40(14):5477–5489, 2013.
- [88] C. Castellaro, G. Favaro, A. Castellaro, A. Casagrande, S. Castellaro, D. Puthenparampil, and C. F. Salimbeni. An artificial intelligence approach to classify and analyse eeg traces. *Neurophysiologie Clinique/Clinical Neurophysiology*, 32(3):193–214, 2002.
- [89] M. Feucht, K. Hoffmann, K. Steinberger, H. Witte, F. Benninger, M. Arnold, and A. Doering. Simultaneous spike detection and topographic classification in pediatric surface eegs. *NeuroReport*, 8(9):2193–2197, 1997.

- [90] M. C. Tjepkema-Cloostermans, R. C. de Carvalho, and M. J. van Putten. Deep learning for detection of focal epileptiform discharges from scalp eeg recordings. *Clinical Neurophysiology*, 2018.
- [91] T. X. Le, T. T. Le, V. V. Dinh, Q. L. Tran, L. T. Nguyen, and D. T. Nguyen. Deep learning for epileptic spike detection. *VNU Journal of Science: Computer Science and Communication Engineering*, 33(2): 1–13, 2018.
- [92] T. Fukami, T. Shimada, and B. Ishikawa. Fast eeg spike detection via eigenvalue analysis and clustering of spatial amplitude distribution. *Journal of neural engineering*, 15(3):036030, 2018.
- [93] W. Ganglberger, G. Gritsch, M. M. Hartmann, F. Fürbass, H. Perko, A. Skupch, and T. Kluge. A comparison of rule-based and machine learning methods for classification of spikes in eeg. j. *Journal of Communications*, 12(10), 2017.
- [94] J. Thomas, J. Jin, J. Dauwels, S. S. Cash, and M. B. Westover. Automated epileptiform spike detection via affinity propagation-based template matching. In *Engineering in Medicine and Biology Society (EMBC), 2017 39th Annual International Conference of the IEEE*, pages 3057–3060. IEEE, 2017.
- [95] J. Le Douget, A. Fouad, M. M. Filali, J. Pyrzowski, and M. Le Van Quyen. Surface and intracranial eeg spike detection based on discrete wavelet decomposition and random forest classification. In *Engineering in Medicine and Biology Society (EMBC), 2017 39th Annual International Conference of the IEEE*, pages 475–478. IEEE, 2017.
- [96] M. L. Scheuer, A. Bagic, and S. B. Wilson. Spike detection: Inter-reader agreement and a statistical turing test on a large data set. *Clinical Neurophysiology*, 128(1):243–250, 2017.
- [97] A. R. Johansen, J. Jin, T. Maszczyk, J. Dauwels, S. S. Cash, and M. B. Westover. Epileptiform spike detection via convolutional neural networks. In *Acoustics, Speech and Signal Processing (ICASSP), 2016 IEEE International Conference on*, pages 754–758. IEEE, 2016.
- [98] A. Rosado and A. C. Rosa. Automatic detection of epileptiform discharges in the eeg. *arXiv preprint arXiv:1605.06708*, 2016.
- [99] S. S. Lodder, J. Askamp, and M. J. van Putten. Inter-ictal spike detection using a database of smart templates. *Clinical neurophysiology*, 124(12):2328–2335, 2013.

- [100] M. De Lucia, J. Fritschy, P. Dayan, and D. S. Holder. A novel method for automated classification of epileptiform activity in the human electroencephalogram-based on independent component analysis. *Medical & biological engineering & computing*, 46(3):263–272, 2008.
- [101] P. Van Hese, B. Vanrumste, H. Hallez, G. J. Carroll, K. Vonck, R. D. Jones, P. J. Bones, Y. D’Asseler, and I. Lemahieu. Detection of focal epileptiform events in the eeg by spatio-temporal dipole clustering. *Clinical Neurophysiology*, 119(8):1756–1770, 2008.
- [102] M. Latka, Z. Was, A. Kozik, and B. J. West. Wavelet analysis of epileptic spikes. *Physical Review E*, 67(5):052902, 2003.
- [103] C. Kurth, F. Gilliam, and B. Steinhoff. Eeg spike detection with a kohonen feature map. *Annals of Biomedical Engineering*, 28(11):1362–1369, 2000.
- [104] C.-W. Ko and H.-W. Chung. Automatic spike detection via an artificial neural network using raw eeg data: effects of data preparation and implications in the limitations of online recognition. *Clinical neurophysiology*, 111(3):477–481, 2000.
- [105] H. Goelz, R. D. Jones, and P. J. Bones. Wavelet analysis of transient biomedical signals and its application to detection of epileptiform activity in the eeg. *CLINICAL ELECTROENCEPHALOGRAPHY-CHICAGO-*, 31(4):181–191, 2000.
- [106] F. Sartoretto and M. Ermani. Automatic detection of epileptiform activity by single-level wavelet analysis. *Clinical Neurophysiology*, 110(2):239–249, 1999.
- [107] Ö. Özdamar and T. Kalayci. Detection of spikes with artificial neural networks using raw eeg. *Computers and Biomedical Research*, 31(2):122–142, 1998.
- [108] H. S. Park, Y. H. Lee, N. G. Kim, D. S. Lee, and S. I. Kim. Detection of epileptiform activities in the eeg using neural network and expert system. *Studies in health technology and informatics*, 52:1255–1259, 1997.
- [109] L. Senhadji, J.-L. Dillenseger, F. Wendling, C. Rocha, and A. Kinie. Wavelet analysis of eeg for three-dimensional mapping of epileptic events. *Annals of biomedical engineering*, 23(5):543–552, 1995.
- [110] T. Kalayci and O. Ozdamar. Wavelet preprocessing for automated neural network detection of eeg spikes. *IEEE engineering in medicine and biology magazine*, 14(2):160–166, 1995.

- [111] A. J. Gabor and M. Seyal. Automated interictal eeg spike detection using artificial neural networks. *Electroencephalography and Clinical neurophysiology*, 83(5):271–280, 1992.
- [112] G. Fischer, N. Mars, and F. L. da Silva. Pattern recognition of epileptiform transients in the electroencephalogram. *Progr. Rep. Inst. Med. Phys.*, 7:22–31, 1980.
- [113] J. J. Halford, M. B. Westover, S. M. LaRoche, M. P. Macken, E. Kutluay, J. C. Edwards, L. Bonilha, G. P. Kalamangalam, K. Ding, J. L. Hopp, et al. Interictal epileptiform discharge detection in eeg in different practice settings. *Journal of Clinical Neurophysiology*, 35(5):375–380, 2018.
- [114] S. Smith. Eeg in the diagnosis, classification, and management of patients with epilepsy. *Journal of Neurology, Neurosurgery & Psychiatry*, 76(suppl 2):ii2–ii7, 2005.
- [115] R. Gregory, T. Oates, and R. Merry. Electroencephalogram epileptiform abnormalities in candidates for aircrew training. *Electroencephalography and clinical Neurophysiology*, 86(1):75–77, 1993.
- [116] M. Sundaram, T. Hogan, M. Hiscock, and N. Pillay. Factors affecting interictal spike discharges in adults with epilepsy. *Electroencephalography and clinical Neurophysiology*, 75(4):358–360, 1990.
- [117] M. A. King, M. R. Newton, G. D. Jackson, G. J. Fitt, L. A. Mitchell, M. J. Silvapulle, and S. F. Berkovic. Epileptology of the first-seizure presentation: a clinical, electroencephalographic, and magnetic resonance imaging study of 300 consecutive patients. *The Lancet*, 352(9133):1007–1011, 1998.
- [118] C. Binnie. Epilepsy in adults: diagnostic eeg investigation. *Recent advances in clinical neurophysiology.*, pages 217–222, 1996.
- [119] J. Thomas, L. Comoretto, J. Jin, J. Dauwels, S. S. Cash, and M. B. Westover. Eeg classification via convolutional neural network-based interictal epileptiform event detection. In *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 3148–3151. IEEE, 2018.
- [120] R. Rahul, J. Thomas, J. Jin, J. Dauwels, S. S. Cash, and M. B. Westover. Machine learning-based epileptic eeg identification system: a multi-center study. In *AES 2018 Annual Meeting Abstract Database*. AESnet.org.

- [121] J. J. Halford, A. Arain, G. P. Kalamangalam, S. M. LaRoche, B. Leonardo, M. Basha, N. J. Azar, E. Kutluay, G. U. Martz, W. J. Bethany, et al. Characteristics of eeg interpreters associated with higher interrater agreement. *Journal of clinical neurophysiology: official publication of the American Electroencephalographic Society*, 34(2):168, 2017.
- [122] M. U. Guide. The mathworks. *Inc., Natick, MA*, 5:333, 1998.
- [123] A. Halevy, P. Norvig, and F. Pereira. The unreasonable effectiveness of data. *IEEE Intelligent Systems*, 24(2):8–12, 2009.
- [124] J. J. Halford. Computerized epileptiform transient detection in the scalp electroencephalogram: Obstacles to progress and the example of computerized eeg interpretation. *Clinical Neurophysiology*, 120(11):1909–1915, 2009.
- [125] S. Kim and J. McNames. Automatic spike detection based on adaptive template matching for extracellular neural recordings. *Journal of neuroscience methods*, 165(2):165–174, 2007.
- [126] J. R. Quinlan. Induction of decision trees. *Machine learning*, 1(1):81–106, 1986.
- [127] V. Bevilacqua, G. Tattoli, D. Buongiorno, C. Loconsole, D. Leonardi, M. Barsotti, A. Frisoli, and M. Bergamasco. A novel bci-ssvep based approach for control of walking in virtual environment using a convolutional neural network. In *Neural Networks (IJCNN), 2014 International Joint Conference on*, pages 4121–4128. IEEE, 2014.
- [128] H. Cecotti and A. Graeser. Convolutional neural network with embedded fourier transform for eeg classification. In *Pattern Recognition, 2008. ICPR 2008. 19th International Conference on*, pages 1–4. IEEE, 2008.
- [129] N.-S. Kwak, K.-R. Müller, and S.-W. Lee. A convolutional neural network for steady state visual evoked potential classification under ambulatory environment. *PloS one*, 12(2):e0172578, 2017.
- [130] H. Cecotti and A. Graser. Convolutional neural networks for p300 detection with application to brain-computer interfaces. *IEEE transactions on pattern analysis and machine intelligence*, 33(3):433–445, 2011.
- [131] E. Bagheri, J. Jin, J. Dauwels, S. Cash, and M. B. Westover. Fast and efficient rejection of background waveforms in interictal eeg. In *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 744–748. IEEE, 2016.

- [132] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, et al. Tensorflow: Large-scale machine learning on heterogeneous distributed systems. *arXiv preprint arXiv:1603.04467*, 2016.
- [133] T.-W. Shen, X. Kuo, and Y.-L. Hsin. Ant k-means clustering method on epileptic spike detection. In *Natural Computation, 2009. ICNC'09. Fifth International Conference on*, volume 6, pages 334–338. IEEE, 2009.
- [134] H. Sahbi and N. Boujemaa. Robust face recognition using dynamic space warping. In *Biometric Authentication*, pages 121–132. Springer, 2002.
- [135] W.-D. Chang and C.-H. Im. Enhanced template matching using dynamic positional warping for identification of specific patterns in electroencephalogram. *Journal of Applied Mathematics*, 2014, 2014.
- [136] V. Niennattrakul and C. A. Ratanamahatana. On clustering multimedia time series data using k-means and dynamic time warping. In *Multimedia and Ubiquitous Engineering, 2007. MUE'07. International Conference on*, pages 733–738. IEEE, 2007.
- [137] J. Hu, B. K. Ray, and L. Han. An interweaved hmm/dtw approach to robust time series clustering. *ICPR (3)*, 6:145–148, 2006.
- [138] G. Chen, Q. Wei, and H. Zhang. Discovering similar time-series patterns with fuzzy clustering and dtw methods. In *IFSA World Congress and 20th NAFIPS International Conference, 2001. Joint 9th*, volume 4, pages 2160–2164. IEEE, 2001.
- [139] P. J. Karoly, D. R. Freestone, R. Boston, D. B. Grayden, D. Himes, K. Leyde, U. Seneviratne, S. Berkovic, T. O'Brien, and M. J. Cook. Interictal spikes and epileptic seizures: their relationship and underlying rhythmicity. *Brain*, page aww019, 2016.
- [140] C. C. Aggarwal. *Data classification: algorithms and applications*. CRC Press, 2014.
- [141] R. Martinez-Cantin. Bayesopt: A bayesian optimization library for nonlinear optimization, experimental design and bandits. *The Journal of Machine Learning Research*, 15(1):3735–3739, 2014.
- [142] F. Iglesias and W. Kastner. Analysis of similarity measures in times series clustering for the discovery of building energy patterns. *Energies*, 6(2):579–597, 2013.

- [143] S. Altunay, Z. Telatar, and O. Erogul. Epileptic eeg detection using the linear prediction error energy. *Expert Systems with Applications*, 37(8):5661–5665, 2010.
- [144] F. Zappasodi, L. Marzetti, E. Olejarczyk, F. Tecchio, and V. Pizzella. Age-related changes in electroencephalographic signal complexity. *PloS one*, 10(11):e0141995, 2015.
- [145] M. J. Putten, S. Olbrich, and M. Arns. Predicting sex from brain rhythms with deep learning. *Scientific reports*, 8(1):3069, 2018.
- [146] A. T. Tzallas, M. G. Tsipouras, D. G. Tsalikakis, E. C. Karvounis, L. Astrakas, S. Konitsiotis, and M. Tzaphlidou. Automated epileptic seizure detection methods: A review study. *Epilepsy-Histological, Electroencephalographic and Physiological Aspects*, pages 75–98, 2012.
- [147] S. Ramgopal, S. Thome-Souza, M. Jackson, N. EsterKadish, I. S. Fernandez, J. Klehm, W. Bosl, C. Reinsberger, S. Schachter, and T. Loddenkemper. Seizure detection, seizure prediction, and closed-loop warning systems in epilepsy. *Epilepsy & Behavior*, 37:291–307, 2014.
- [148] P. Fergus, D. Hignett, A. Hussain, D. Al-Jumeily, and K. Abdel-Aziz. Automatic epileptic seizure detection using scalp eeg and advanced artificial intelligence techniques. *BioMed Research International*, 2015: 17, 2015.
- [149] K. C. Chua, V. Chandran, U. R. Acharya, and C. M. Lim. Analysis of epileptic eeg signals using higher order spectra. *Journal of Medical Engineering and Technology*, 33(1):42–50, 2009.
- [150] S. Mousavi, M. Niknazar, and B. V. Vahdat. Epileptic seizure detection using ar model on eeg signals. In *Biomedical engineering conference, 2008. CIBEC 2008. Cairo International*, pages 1–4. IEEE, 2008.
- [151] Z. Iscan, Z. Dokur, and D. Tamer. Classification of electroencephalogram signals with combined time and frequency features. *Expert Systems with Applications*, 38:10499–10505, 2011.
- [152] A. Shoeb. Application of machine learning to epileptic seizure onset detection and treatment. *Ph.D thesis, Massachusetts Institute of Technology*, 2009.
- [153] F. Furbass, P. O. P, M. Hartmann, H. Perko, A. M. Skupch, G. Lindinger, L. Elezi, E. Pataraiia, A. J. Colon, C. Baumgartner, and T. Kluge. Prospective multi-center study of an automatic online seizure detection system for epilepsy monitoring units. *Clinical Neurophysiology*, 126(6):1124–31, 2015.

- [154] P. Thodoroff, J. Pineau, and A. Lim. Learning robust features using deep learning for automatic seizure detection. In *Machine learning for healthcare conference*, pages 178–190, 2016.
- [155] J. Saini and M. Dutta. An extensive review on development of eeg-based computer-aided diagnosis systems for epilepsy detection. *Network: Computation in Neural systems*, 28(1):1–27, 2017.
- [156] O. Faust, Y. Hagiwara, T. J. Hong, O. S. Lih, and U. R. Acharya. Deep learning for healthcare applications based on physiological signals: A review. *Computer Methods and Programs in Biomedicine*, 161:1–13, 2018.
- [157] Y. Guo, Y. Liu, A. Oerlemans, S. Lao, S. Wu, and M. S. Lew. Deep learning for visual understanding: A review. *Neurocomputing*, 187:27–48, 2016.
- [158] Q. Guo, F. Wang, J. Lei, D. Tu, and G. Li. Convolutional feature learning and hybrid cnn-hmm for scene number recognition. *Neurocomputing*, 184:78–90, 2016.
- [159] J. Thomas, L. Comoretto, J. Jin, J. Dauwels, S. S. Cash, and M. B. Westover. Eeg classification via convolutional neural network-based interictal epileptiform event detection. In *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 3148–3151. IEEE, 2018.
- [160] S. N. Abdulkader, A. Atia, and M.-S. M. Mostafa. Brain computer interfacing: Applications and challenges. *Egyptian Informatics Journal*, 16(2):213–230, 2015.
- [161] A. C. N. Society. *Guideline 9B: Recommended Standards for Visual Evoked Potentials*. Surflin.com, <http://www.acns.org/pdf/guidelines/Guideline-9B.pdf>.
- [162] S. J. Luck. *An introduction to the event-related potential technique*. MIT press, 2014.
- [163] V. P. Oikonomou, G. Liaros, K. Georgiadis, E. Chatzilari, K. Adam, S. Nikolopoulos, and I. Kompatsiaris. Comparative evaluation of state-of-the-art algorithms for ssvep-based bcis. *arXiv preprint arXiv:1602.00904*, 2016.
- [164] A. L. Goldberger, L. A. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley. Physiobank, physiotoolkit, and physionet. *Circulation*, 101(23):e215–e220, 2000.

- [165] H. Ramchoun, M. Amine, J. Idrissi, Y. Ghanou, and M. Ettaouil. Multilayer perceptron: Architecture optimization and training. *IJIMAI*, 4(1):26–30, 2016.
- [166] D. F. Morrison. *Multivariate analysis, overview*. Wiley Online Library, 1998.
- [167] A. Erfanian and B. Mahmoudi. Real-time ocular artifact suppression using recurrent neural network for electro-encephalogram based brain-computer interface. *Medical and Biological Engineering and Computing*, 43(2):296–305, 2005.
- [168] E. M. Forney and C. W. Anderson. Classification of eeg during imagined mental tasks by forecasting with elman recurrent neural networks. In *Neural Networks (IJCNN), The 2011 International Joint Conference on*, pages 2749–2755. IEEE, 2011.
- [169] C. Spampinato, S. Palazzo, I. Kavasidis, D. Giordano, M. Shah, and N. Souly. Deep learning human mind for automated visual classification. *arXiv preprint arXiv:1609.00344*, 2016.
- [170] D. Balderas, A. Molina, and P. Ponce. Alternative classification techniques for brain-computer interfaces for smart sensor manufacturing environments. *IFAC-PapersOnLine*, 48(3):680–685, 2015.
- [171] F. Lotte, M. Congedo, A. Lécuyer, F. Lamarche, and B. Arnaldi. A review of classification algorithms for eeg-based brain-computer interfaces. *Journal of neural engineering*, 4(2):R1, 2007.
- [172] T. Hinterberger, A. Kübler, J. Kaiser, N. Neumann, and N. Birbaumer. A brain-computer interface (bci) for the locked-in: comparison of different eeg classifications for the thought translation device. *Clinical Neurophysiology*, 114(3):416–425, 2003.
- [173] D. J. Krusienski, E. W. Sellers, F. Cabestaing, S. Bayouth, D. J. McFarland, T. M. Vaughan, and J. R. Wolpaw. A comparison of classification techniques for the p300 speller. *Journal of neural engineering*, 3(4):299, 2006.
- [174] S. N. Carvalho, T. B. Costa, L. F. Uribe, D. C. Soriano, G. F. Yared, L. C. Coradine, and R. Attux. Comparative analysis of strategies for feature extraction and classification in ssvep bcis. *Biomedical Signal Processing and Control*, 21:34–42, 2015.
- [175] R. Singla and B. Haseena. Comparison of ssvep signal classification techniques using svm and ann models for bci applications. *International Journal of Information and Electronics Engineering*, 4(1):6, 2014.

- [176] A. V. Oppenheim. *Discrete-time signal processing*. Pearson Education India, 1999.
- [177] J. G. Proakis. Digital signal processing: principles, algorithms, and application-3/e. 1996.