DETECTING THE ONSET OF AN EPILEPTIC SEIZURE USING A NOVEL TIME-SERIES APPROACH

by

Suhani Shrivastava, B.E.

THESIS

Presented to the Faculty of

The University of Houston-Clear Lake

In Partial Fulfillment

Of the Requirements

For the Degree

MASTER OF SCIENCE

in Computer Science

THE UNIVERSITY OF HOUSTON-CLEAR LAKE

MAY, 2018

DETECTING THE ONSET OF AN EPILEPTIC SEIZURE USING A NOVEL TIME-SERIES APPROACH

by

Suhani Shrivastava

APPROVED BY

_____

Gary D. Boetticher, Ph.D., Chair

_____

Terry Feagin, Ph.D., Committee Member

_____

Richard L. Puzdrowski, Ph.D., Committee Member

APPROVED/RECEIVED BY THE COLLEGE OF SCIENCE AND ENGINEERING:

_____

Said Bettayeb, Ph.D., Associate Dean

_____

Ju H. Kim, Ph.D., Dean

Dedication

I dedicate this thesis to my loving husband for his dedicated partnership through thick and thin in every aspect of life. I also dedicate this work to my late grandmother, all my family members and friends.

Acknowledgements

I am extremely thankful to my thesis advisor Dr. Gary D. Boetticher for constantly steering me in the right direction. Without his constant support and motivation, this work would not have been possible. I would also like to thank Dr. Terry Feagin and Dr. Richard L. Puzdrowski for their suggestions and feedback in writing this thesis.

ABSTRACT

DETECTING THE ONSET OF AN EPILEPTIC SEIZURE USING A NOVEL TIME-SERIES APPROACH

Suhani Shrivastava
The University of Houston-Clear Lake, 2018

Thesis Chair: Gary D. Boetticher, Ph.D

Electroencephalography (EEG) is one of the most popular non-invasive techniques for acquiring electrical signals from the brain. Data mining EEG signals finds numerous applications in the field of neuroscience for obtaining crucial information about the neural activities. EEG data is very complex in that it is non-stationary and multidimensional. Therefore, the task is how to convert voluminous raw EEG data into a succinct representation. This research provides a methodology for representing EEG data in a concise and comprehensible format using a minimum number of data points without the loss of useful information. Among many applications of studying EEG data, detecting epileptic seizures concerns neurologists the most. Epilepsy is a serious disorder characterized by the occurrence of epileptic seizures. These seizures occur as a result of abnormal neuronal activities of the brain. Today, more than 65 million people in the world suffer from epileptic seizures which can be life-threatening. It is not just the physical effects

of seizure that impacts patients adversely but also the social isolation that the patient and their families face. If EEG signals are analyzed properly, seizures can be predicted at their onset. This thesis proposes a seizure prediction method which uses a novel time-series approach to provide a useful method for the diagnosis of epileptic seizures. The key to the method identifies transitions from non-epileptic(pre-ictal) to epileptic(ictal) segments of the EEG signal using offset statistical moving averages. This research examines EEG data of multiple epileptic patients from CHB MIT database. The method analyzes EEG signals for common transitional patterns using multiple inter-patient and intra-patient seizure files. The experiments provide substantial results and predict seizures early in some situations and with a minimal latency in a few other situations.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1 INTRODUCTION

With tremendous developments in neuroimaging technology, there has been an explosion of neuroscience data. Electroencephalography (EEG) is one of the most widely used techniques to obtain electrical signals from the brain. Hans Berger, a German psychiatrist developed the first human EEG in the year 1924 [Teplan02]. Many researchers and neuroscientists study EEG data to acquire an in-depth understanding of the human brain network.

EEG exemplifies one of the non-invasive methods for acquiring electrical signals from the brain [Morshed14]. Data Mining EEG signals can be extremely useful for interpreting neural activities of the brain by analyzing frequency patterns associated with different neurological tasks.

One of the major challenges when data mining EEG signals is the lack of any symmetry and consistency of the signals. EEG patterns border on a chaotic state make representation and interpretation difficult.

Ideally, it would desirable to be able to map EEG signals to specific thoughts. However, due to the embryonic nature of the discipline, such a goal is unrealistic at this point. Moving in the direction of this ideal requires the distillation of EEG signals into a succinct form and to identifying replicated EEG patterns.

It would not be surprising to find very few replicated EEG patterns because of the chaotic nature of the data. One way to find more patterns would be to relax the constraints of acceptable matches. Thus increasing the number of matches at the expense of accuracy.

2 BACKGROUND

2.1 EEG Recording Setup

These electrical impulses are generated by neural activities of the brain and can be recorded by placing electrodes on different regions of the scalp. However, electrode placement can be a challenging task as it requires multiple permutations according to different cortical regions of the scalp [Simkin14].

An EEG recording setup uses the following equipment (refer Figure 1):

1) Electrodes (gel-less, pre-gelled, saline based electrodes, and reusable metal electrodes)
2) Conductive gel
3) Amplifiers and filters
4) Analog to Digital Converter
5) Recording Hardware & Software

Scalp electrodes vary from 1 to 3 mm in diameter with leads that are attached to the amplifier. For long term and invasive EEG recordings, needle electrodes which penetrated into the skin are used. Needles can cause infection and bleeding therefore proper hygiene must be maintained. For electrodes, an abrasive paste is applied on the scalp to create a minute abrasion for obtaining EEG recordings. For recording multiple channels electrode caps are usually preferred [Teplan02].

Electrodes generate signals in microvolts. These signals are amplified for required digitization. The A/D converter changes analog signals into digital format. Once the signals are in digital format, they can be stored and displayed on a computer.

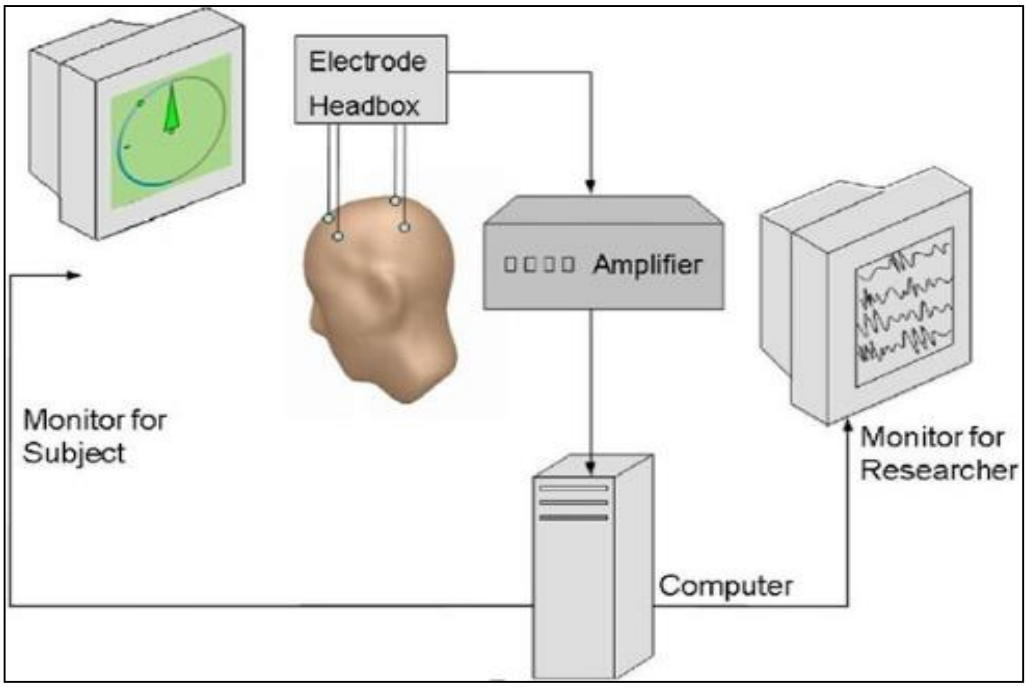

Figure 1: EEG Recording Setup [Teplan02]

2.2 Electrode Placement System

EEG uses the International 10-20 System as a standard for electrode placement on the scalp. It is called 10-20 system because the electrodes are placed at distances in increments of 10%-20% over the anatomical regions of the brain(as shown in Figure 2). The electrodes are placed at these particular sites. EEG data varies with specific regions of the brain. On the basis of these regions, letters are used to represent these sites. Central Sulcus (C), Frontal Lobe (F), Frontopolar area (FP), Occipital Lobe (O), Parietal Lobe (P), and Temporal Lobe (T). Label 'A' is used for reference sites such as ear positions and ground positions. Label 'Z' represents the central line along the hemisphere of the brain. Odd numbers are used to reference electrode sites to the left of the central line (Z) and even numbers for electrode positions to the right of the central line (Z).

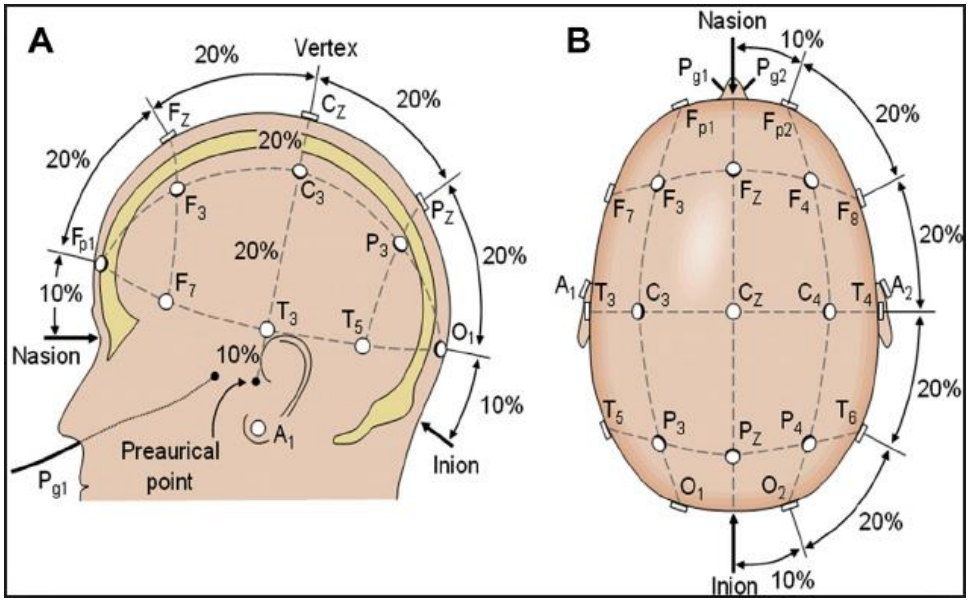

Figure 2: 10 – 20 International System of electrode placement [Cahn 06]

EEG channels represent spatial information via various electrode sites. It becomes important to categorize data according to cortical regions of the brain which requires data mining EEG signals.

2.3 EEG Data

Data obtained using electroencephalography (EEG) in the form of electrical rhythms or frequency bands (refer to Fig. 3) falls under the following categories [Teplan02]:

- **Delta** (<4 Hz) – While in the Delta state, the mind experiences deep healing. These waves are observed when a person is sleeping. Delta waves have the highest amplitude. Sleep walking and sleep talking is observed in the Delta state.
- **Theta** (4 - 8 Hz) - Theta waves indicate a deep relaxation state while the mind is still conscious. Theta waves indicate exceptional insight. These waves also occur when a person enters into sleep from alpha.
- **Alpha** (8 - 14 Hz) - Alpha waves are observed when thoughts are smoothly flowing in the mind. It is a comparatively relaxed state. Increased amount of alpha waves is characterized by positive health benefits.
- **Beta** (14 - 30 Hz) - These frequencies exist when the subject is in highly attentive state. For instance, if a person is having conversation with someone, engaged in some decision making or problem solving activity, then the brain emits Beta waves.
- **Gamma** (>30 Hz) - Gamma frequencies occur when the subject or patient reaches an extremely deep meditative states like those in monks and nuns. It is a difficult

task to detect gamma frequencies using EEG recording since most recorders are capable of detecting signals only up to 25 Hz.

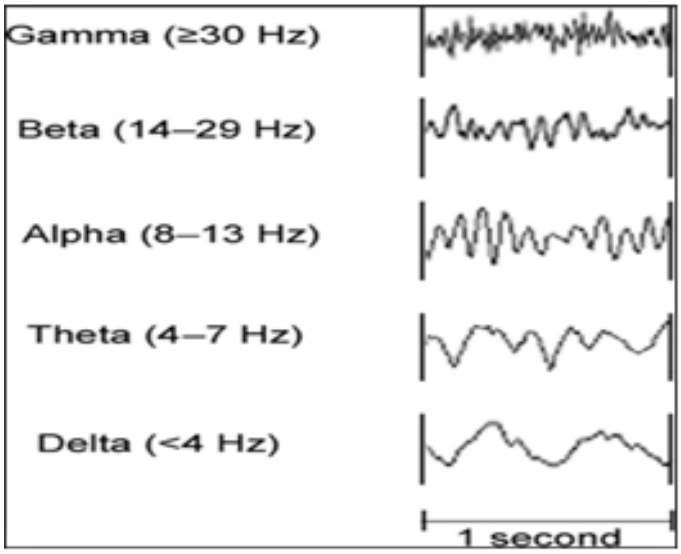

Figure 3: EEG Frequency bands [Teplan02]

2.4 Epileptic Data

Epilepsy is a brain disorder which occurs as a result of transient changes in the activity of neurons. Recurrent electrical discharges in the EEG signals represent seizure activity. It is prevalent worldwide among people of different age groups and therefore remains an issue of utmost concern. Around 65 million people suffer from epilepsy across the globe [Ngugi, Anthony K, 2010]. Researchers and scientists have been experimenting with different strategies to address and prevent epilepsy.

Common causes of epilepsy include:

- Brain Stroke
- Head injury
- Trauma
- Infections in certain part of the brain.

The nature of epilepsy varies in terms of severity, type of epilepsy and the brain region from where it originates. The epilepsy pattern also changes from person to person. Collecting EEG data can be a time consuming task which takes several days [Guttag J., Shoeb A, 2010]. Continuous and long-term evaluation of EEG patterns helps monitor and diagnose seizure patterns in patients.

2.5 Epileptic Seizures

Epilepsy is accompanied by seizures that are unpredictable in nature. Epileptic seizures can be categorized into the following types:

- Focal seizures. Focal seizures originate from specific region of the brain. These are also called partial seizures. During the time of seizure, it is possible that the patient is aware or may have impaired awareness.
- Generalized seizures. A generalized seizure involves electrical disturbances in the entire brain. It affects both hemispheres of the brain. The patient usually has

impaired awareness during a generalized seizure [Tzallas, Alexandros T, Tsipouras, Markos G, 2012].

2.6 Dimensions of EEG Data

EEG data is multidimensional [Jrad11] and has following dimensions:

- **Spatial** - EEG signals from various areas of the brain. This is done by selecting specific EEG channels (electrode sites).
- **Spectral** - EEG signal representation based on specific frequency bands.
- **Temporal** - representation of EEG signals based on time domain.

Initially, EEG data is obtained in time-domain [Polikar96, Suleiman07]. However, the information obtained from time-domain analysis is not sufficient for obtaining useful information. Therefore, there is a need for frequency analysis or spectral analysis of raw EEG signals for obtaining relevant information from the signal. Various signal transformation techniques are available for converting a signal in time-domain to frequency-domain such as a Fourier Transform. However, if Fourier transform is used, only spectral information about the signal is available but not the time-domain information at the same time. Furthermore, Fourier Transform requires EEG signals to be stationary. However, since the EEG signals are non-stationary, it is not an efficient method to use Fourier Transform for analysis of EEG signals. Moreover, a signal processing technique is required which can provide both spectral and time domain analysis of the signal. Wavelet Transform (WT) can be used for time-domain as well as spectral analysis of EEG signals [Kong14], [Polikar96].

2.7 Data Mining Approaches

Data mining extracts useful information in the form of patterns or models from a given data set [Fayyad96]. Data obtained from various databases could be structured, semi-structured, or unstructured. For efficient utilization of data, it must be analyzed using various data mining techniques. Data Mining deploys numerous approaches such as regression, classification, clustering, and association [Bharati10].

- **Regression -** Regression techniques model the relationship between already known variables or data inputs and the variables or data that we want to predict. Linear Regression (LR), Non- Linear Regression and Multivariate Linear/Non-Linear Regression are some of the regression techniques in data mining.
- **Classification -** Classification is a common data mining technique to distinguish and classify data records based on a pre-determined set of rules. Neural Networks (NN), Support Vector Machines (SVM), Decision Trees, etc. are examples of data mining tools capable of classification.
- **Clustering –** Clustering partitions data based on similarities and dissimilarities. This algorithm seeks to create clusters having strong associations among its members and weak association across different clusters. There are various clustering methods available such as partitioning methods, hierarchical methods, grid-based methods, etc. [Bharati10]
- **Association Rule Learner -** Association rules examine multiple attributes looking for correlation patterns. Different types of association rules include Quantitative

Association Rule, Multilevel Association Rule, Multidimensional Association rule, etc.[Bharati10]

2.8 Data Mining Strategies of EEG Data

Based on the source of EEG data, signal processing and classification is done. A typical EEG signal may require one or more processing techniques such as artifact removal, feature extraction, classification algorithms to categorize EEG data, and for post-processing of signals etc.

- **Artifact Removal:** EEG signals are contaminated with artifacts from external as well as internal sources. Any external activity in the environment such as electrode displacement, cable movements, applying too much electrode gel on the scalp surface, broken wire contacts, etc. add noise to the EEG signals [Nolan10]. Internal sources include artifacts arising from muscle movements, eye blinks, sweating etc.
- **Feature Extraction:** A feature is a distinguishing property or a recognizable component that is derived from a pattern. Feature extraction seeks to obtain important information from huge and multidimensional data sets [Jrad11, Al-Fahoum14]. There are a variety of feature extraction methods that provide dimensionality reduction.
- **Classification of Features:** After extracting relevant features, various classification techniques or classifiers can be applied to categorize the extracted features into classes. For example, Linear Discriminant Analysis (LDA), Support Vector Machine (SVM), clustering algorithms for modeling data, etc. are simple classifiers for feature

classification. Once classification is done, EEG data can be used for required applications.

2.9 Challenges in Data Mining EEG Signals

Mining EEG signals can lead to significant scientific breakthroughs. However, there are some major challenges related to data mining EEG signals.

- Electrode placement is a critical task and it requires various adjustments to extract the desired EEG signals.
- Electrodes used for collecting EEG data are highly sensitive and can easily pick up noises from other electrical activities within the surrounding environment. Also, other factors such as excessive application of gel on the scalp or any loose cables make it difficult to obtain clean EEG signals.
- Any movement in subject's muscle activity, eye movements, blinks, etc. can highly distort the original EEG signals. Also, external changes to the environment (e.g. noises, odors, light) can distort the EEG recordings.
- EEG data is complex. A recording from a single electrode probe is composed of many other electrical signals from thousands of neurons, each of which has different amplitudes and frequencies. Therefore, separating and classifying individual wave frequencies is a difficult task [AlZoubi08].
- EEG data collection is spatial. Integrating EEG readings from different brain locations poses a challenge.

- EEG data is temporal. The signals are susceptible to constant change within microseconds. The signals also tend to jitter from trial to trial.
- Each person's brainwave pattern is unique. Comparing two different person's brainwaves requires calibration.
- Despite the knowledge of the source of EEG signals, the neurons do not just depict the signals originating from that particular source. Rather it is a mixture of signals from other neurons in the surrounding space. Thus poor spatial resolution requires spatial filtering and preprocessing of raw EEG signals [Evans 08].
- EEG Signals can be chaotic.

3 LITERATURE REVIEW

3.1 General Overview

In the past two decades, a considerable amount of work has been published regarding data mining EEG signals. EEG data mining involves pre-processing, artifact removal, feature extraction, and classification techniques. These approaches vary with their area of application.

- Flexer[2000] reviews of Data Mining techniques in EEG that have been used in the past. These include Neural Networks, Hidden Markov Models, Fuzzy Logic.
- Bialas[2014] provides spatial filtering mechanism for filtering out irrelevant information from the raw EEG signal. An automatic EEG spike detection method is proposed by Ko[1998] for classifying EEG data from multiple channels using ANN. Ayhan[2011], Pullaiah[2017] et al. use discrete wavelet transform (DWT) for analyzing non-stationary EEG signals. Most of the artifact removal methods depend on the type of artifacts and their area of origin, for example artifacts arising from eye movements will be treated differently from artifacts originating from the head movements. Scholgel[2007] et al. provide such methods. Assi[2014] uses Independent Component Analysis (ICA) for ocular artifact removal.

  However, some artifact removal methods have a broader scope. For example, Junghofer[2000] et al. provides a SCADS method which can detect artifacts from various sources. Also, Nolan[2010] et al. describes some advanced artifact removal techniques.

- For feature extraction, various techniques exist. Azlan[2014] conducts feature extraction using ICA and PCA. Some researchers combine these techniques. Adeli[2003] uses Discrete-Wavelet Transform (DWT) for feature extraction.
- In Classification algorithms, Subasi[2010] uses Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA), ICA, and Support Vector Machines (SVM) for classifying EEG signals. [Orhan11] deploys cluster-based approach for classification. Acharya[2012] apply classifiers such as K-Nearest Neighbor (KNN), SVM, ANN, Decision Trees (DT) and Fuzzy Sugeno Classifier in their work.

Table 1 provides a general research perspective of machine learners and other signal processing techniques applied to EEG data. These methodologies are used for feature extraction and feature selection, classification of EEG signals, and also for processing EEG signals. Table 2 gives a distribution of the research papers mentioned in Table 1 spanning the past two decades.

Although some significant amount work has been published in data mining EEG signals, many questions remain unanswered considering the complexity brain signals.

Table 1. Related Work - Machine Learners and Signal Processing for EEG Signals

Machine Learner/Methodology	Paper
Clustering	[Exarchos05,Lee10, KG06, Georgiev07,Panuccio02, Moerchen03]
Fast Fourier	[Geng14,Shirazi14, Moerchen03, Rivero15, Gevins74, Thieu15, Poulos98, Ramaraju11]
Filtering	[Ko98, Exarchos05,Bialas14]
GA, GP	[Schroder03,Yaacoub17,Petrantonakis09, Rejer13, Rivero15, Firpi05, Aguiar2000]
ICA	[Vigario2000,Muller2000,Subasi10,Hosni07,Shamlo13, Jirayucharoensak13, Assi14,Joyce03, Kroupi14, Azlan14, Lee09]
K-NN or KNN	[Wang11,Islam11,Chan15,Hu12, Swetapadma16,Acharya12, Lotte07,Schuster10, Chaovalitwongse07, Alzoubi08, Firpi05]
Neural Network	[Hazarika97, Ko98, Wang05,Basssani08, Yaacoub17, Islam11,Guan16,Swetapadma16, Wiechert16, Lhotska09, Acharya12, KG06, Garett03,Sen14, Lotte07, Jahankhani07, Jahankhani06, Flexer2000, Purnamasari16, Nurse16]
K- Means Clustering	[Assi14]
Prin. Comp. Analysis (PCA)	[Subasi10,Elsawy13,Labib16, Lhotska09, Azlan14, Jahankhani07]
Spectral Analysis, Common Frequency Pattern	[Pregenzer99,Dalponte07,Lin10,Suk11,Pan13, Geng14, Kroupi14, Alzoubi08, Gevins74, Poulos98]
Support Vector Machine	[Schroder03, Wang05, Subasi10, Hosni07,Lin10, Chaovalitwongse11, Wang_ Shouyi11, Islam11, Chen12,Hussein13, Pan13,Su13, Chan15, Labib16, Bugeja16, Swetapadma16, Alzoubi08,Chatchinarat16, Wiechert16, Deedwaniya16, Acharya12, KG06, Garett03, Sen14, Xu14, Lotte07, Kroupi14, Markopoulos16,Horlings08, Chaovalitwongse11,Chandrashekar14, Rejer13, Santana12, Pinto15]
Wavelet Transform	[Hazarika97,Xue03,Wang05, Subasi10,Takajyo06, Basssani08, Lee10,Ramaraju11,Ayhan11, Chen12, Jirayucharoensak13, Labib16,Robinson16,Pullaiah16, Chatchinarat16, Salah11, Jahankhani07, Jahankhani06, Moerchen03, Wang11,Wang_ Shouyi11, Geng14,Ting07, Gao10, Salah11, Purnamasari16]
Quadratic Discriminant Analysis	[Petrantonakis09, Swetapadma16]
Self Organizing Maps	[Yamagutchi07, Lee10]
Linear Discriminant Analysis (LDA)	[Subasi10,Wang10, Wang_ Shouyi11, Bialas14, Assi14, Swetapadma16, Garett03,Lotte07, Santana12]
Common Spatial Patterns	[Chen12,Xu14, Markopoulos16, Su12]

Table 2. Distribution of Related Research over the Past Two Decades

Methodology	Year																				
	97	98	99	00	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17
Clustering						1	1		1	1	1							1			
Fourier Transform		1					1								1			2	2		
Filtering		1							1									1			
GA, GP				1			1	1	1				1				1		1		1
ICA				2			1				1		1	1			2	3			
KNN									1		2	1		1	2	2			1	1	
NN	1	1		1			1		1	2	2	1	1		1	1		1		4	1
PCA											1		1	1			1	1		1	
Spectral Analysis		1	1								1	1		1	1		1	2			
SVM							2		1	1	2	1		2	4	3	4	4	2	7	
Wavelet Transform	1						2		1	1	3	1		3	6	1	1	1		5	
LDA							1				1			2	1	1		2		1	
QDA													1							1	
SOM											1			1							
CSP																1		2		1	

Figure 4 below shows a strong upward trend in publications on Data Mining EEG signals.

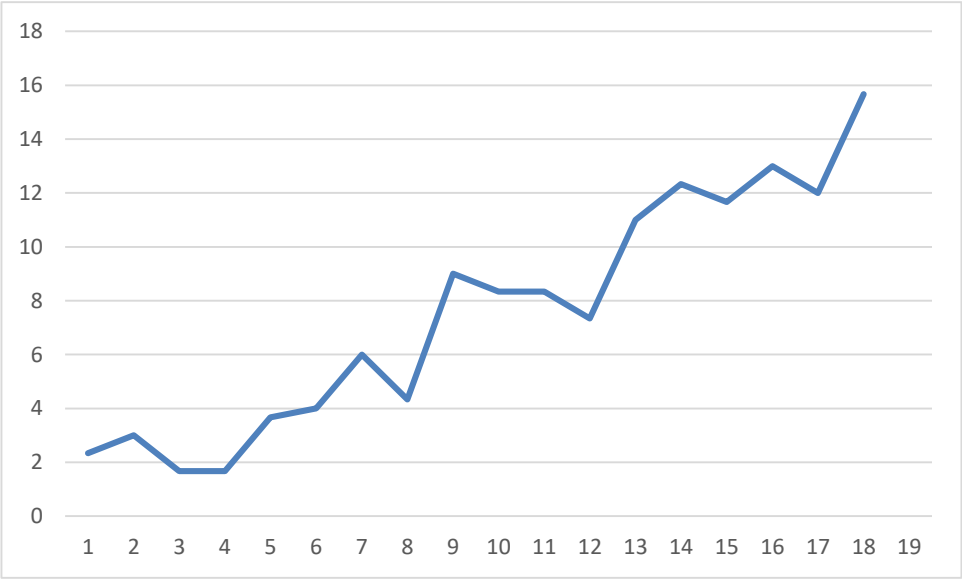

Figure 4: Publication Rate (3-Year Average) Since 1997

3.2 Related Research

There is a trending interest in the development of algorithms that not only detect the onset of seizures but also predict the seizures before they occur. Many linear and non-linear methods have been proposed. Linear methods include approaches such as examining variance, kurtosis, skewness etc. of the EEG signals. Non-linear methods include correlation density, similarity/dissimilarity index etc.

Tarassenko et al. [1998] use a neural network classifier that predicts seizure for patient specific data as well as across patients. The authors detect spike patterns (inter-ictal activity) in EEG signals using time-domain parameters. These parameters are fed as input to a multi-layer perceptron. For patient specific classifiers, the authors achieved sensitivity between 83.1% to 97.3%. However, for other patients, the results gave an increased number of false-positives. Hence, clinical application across multiple subjects cannot be considered at this moment.

Tzallas et al. [2009] also use artificial neural networks that classifies seizure and non-seizure activity. The authors apply time-frequency analysis to extract features from EEG segments. These features depict energy distribution of the signal segments. These features are applied as input to neural network for classifying polyspikes (seizure activity). The authors tested this on publicly available datasets and obtained an accuracy varying from 89% to 100%.

Sakkalis et al. [2013] use three approaches to identify absence seizures using Order Index(non-linear), Multiscale Variance Index(linear) and approximate entropy measures. All these methods achieved reasonable sensitivity in terms of detecting seizure. However, Multiscale variance gave biased results for detecting long duration seizures. This lead to lower specificity rates. Better performance is expected when these measures combine with other classifiers.

Shoeb et.al [2010] propose a method that predicts the onset of seizure using Support Vector Machine(SVM) classifier. Due to the classification boundary issues between seizure and non-seizure, a radial basis function (RBF) kernel is also applied. The prediction of seizure is patient specific. The detector using this algorithm predicted seizure onset with 96% sensitivity. The authors further combine EEG and ECG recordings to predict the onset of seizure. Seizure prediction performance increases with this approach.

Chaovalitwongse et. al [2006] apply optimization-based techniques to inter-cranial EEG(iEEG) data. The classification of normal and abnormal signals is prepared using support vector machines and statistical cross-validation. Rasekhi et al [2013] use linear univariate features and predict seizure using machine learning methods. Several univariate features are combined from 6 channels to create a feature space. This is repeated for each of the 10 patients.48 combinations of methods are applied to achieve best setting.

Hasan et al. [2017] use the k-nearest neighbors (k-NN) algorithm for classification of seizure and non-seizure. Statistical parameters such as approximate entropy, standard

deviation, mean absolute value, standard error etc. are used followed by regression analysis for predicting seizure. The algorithm is applied to patients of different age groups.

McSharry et al. [2003] discuss how linear methods like variance statistics perform equivalently well when compared with other complex non-linear techniques like correlation dimension between multiple channels. They also discuss how any seizure prediction scheme should test out-of-sample datasets that are not known to contain any seizures. Using this approach, seizure prediction algorithms are compared using false-positives/outliers in a non-seizure dataset.

Tzallas et al. [2012] lists common methods that have been used for detecting abrupt changes in EEG signals. These include knowledge-based rules, artificial neural networks, template-based methods, clustering techniques and other data mining techniques. They also mention the use of mimetic techniques which requires base rules set by an expert neurophysiologist to identify a spike as seizure. Spike attributes such as height, frequency, slope etc. are compared with the values provided by the neurophysiologist. All of these methods use either one of single-channel data or multiple-channel data. The authors provide a great review on existing seizure prediction and seizure identification techniques.

Unlike the above approaches, this research uses a simple approach to identify and predict a seizure. Initially, the proposed method uses a classification approach using adjacent moving average groups. Later on, using additional offset values (in a loop), the method predicts seizure onset when the patient transitions from a non-seizure to a seizure mode.

Using the concept of different offsets, provides an adjusted threshold for better prediction of the seizure onset time.

4 STATEMENT OF PROBLEM

EEG data represents different neural activities of the brain which is of great value to the field of neuroscience. Using EEG, the neurological state of the subject or patient can be determined, which can further be used for clinical, therapeutic, medicinal and physiological applications.

From a clinical perspective, EEG diagnoses brain disorders such as attention deficit hyperactivity disorder (ADHD), Alzheimer's disease, Schizophrenia, Epilepsy, etc. It is also capable of monitoring the state of coma, alertness and brain death. EEG analysis also helps in testing the effect of drugs and for detecting the origin of seizures, brain injuries and lesions.

By observing EEG patterns, trauma and stress levels of a subject can be determined which significantly helps in cognitive behavioral therapies and for providing the required remedy or treatment to the subject or patient. Besides its clinical and medicinal applications, EEG data also finds application in Brain Computer Interface devices, which aim to provide assistance to the disabled.

For the above mentioned applications of EEG data, representing EEG signals is a crucial issue. However, one of the biggest problems is the non-stationary and multidimensional nature of the EEG data subject to frequent and non-predictable changes. These problems hinder the ability to represent EEG signals in a condensed form.

Furthermore, while recording EEG signals, the presence of errors and artifacts of various types requires extensive pre-processing and filtering mechanisms.

What is needed is a method of representing EEG waves in a succinct format with minimal loss of useful information that supports extensive pattern recognition by relaxing pattern constraints.

The succinct representation is necessary for many EEG applications. One of these include monitoring EEG signals for the diagnosis and prediction of epileptic seizures in patients. The analysis of seizure patterns requires feature extraction. A feature essentially represents an important segment of a pattern. While predicting a seizure, features would be the spots where abrupt changes in the signal patterns occur. Typically, an epileptic seizure is characterized by irregular patterns with spikes and waves. Part of the problem is to identify such spots.

EEG signals are prone to artifacts and noises which are also identified by abrupt spikes, in such cases it is difficult to differentiate whether the change occurs due to onset of seizure or due to some random artifact/noise signal. It can increase the number of false-positives (outliers counted as seizures) and hence it becomes important to distinguish an epileptic seizure from an outlier. This requires using some form of classification mechanism.

The frequency of seizures varies across different patients. For patients with infrequent seizures, it becomes difficult to discover seizures patterns. For obvious reasons, the lack of sufficient historical data hinders the process of pattern discovery in such cases. The results in such cases can be misleading.

Even if any patient has frequent seizures, the types of seizures (focal, generalized, etc.) can vary.

Also, since the seizure patterns may differ from patient to patient, it becomes intuitively difficult to generalize parameters that give expected results across different patients.

In order to locate the brain region from where the seizure actually originates, it becomes important to identify the onset of a seizure. Patients usually take anti-seizure drugs. For cases where patients suffer from recurrent seizures, these drugs don't suffice alone, hence brain surgery becomes a requirement. If the onset of seizure is known, the surgeons can trace and locate the origin spot inside the brain. [Khan et al. [2012]]

5 CONTRIBUTIONS/TECHNICAL INNOVATIONS

The following list shows the technical innovations of this research.

- Distill EEG signals to a very simple representation using as few as three data points.

  Examining EEG signals from visual inspection can be a tedious task even for medical experts. This becomes all the more difficult in case of long-term EEG signals. Therefore, a simple representation of the EEG signal is needed that does not require manual analysis and can be embedded in clinical equipment that keeps track of the patient's EEG signal.

  The proposed method uses data points from the original EEG signal and represents the important data in the form of peaks and valleys. The points of interest combine together to form triangular patterns.

- The ability to relax constraints for what constitutes a match in EEG signals. This would help in recognizing similar patterns in the EEG waves which further depicts somewhat similar neural activities.

  This implies finding matches using threshold parameters. For instance, distinguishing between epileptic seizure data and non-seizure data, threshold boundaries can be utilized.

- The ability to differentiate changes in state information. This implies detecting changes in frequency or amplitude of the EEG signal. This is helpful for detecting a change from non-epileptic zone to epileptic seizure.

The apparent chaotic nature of EEG signals with the added presence of artifacts makes it difficult to characterize a state transition. Other transitions such as seizure to non-seizure, or within seizures can also be detected using the proposed method. However, this research focuses on the prediction of non-seizure to seizure, or the onset of a seizure.

- The ability to describe more complex patterns by extending beyond three data points.

  The EEG datasets used for study are long-term recordings ranging from minutes to hours, and thus representing the signals using few data points alone cannot suffice. The triangular patterns formed using data points are further coalesced into groups. The groups are analyzed for transition.

- This approach may be applied to other domains that use time-series data.

  Time-series data involves observations with constant time interval between the readings. Time-series analysis focuses on predicting trends, forecasting events, identifying patterns, detecting abnormalities in the data to identify fluctuations and outliers. Several estimation and interpolation techniques can be used to analyze time-series data depending on the complexity, application and nature of the data. Examples include weather forecasting, stock market prediction, or earthquake prediction.

- Detect the onset of a seizure.

  The proposed method predicts the onset of seizure using supervised learning using both intra-patient and inter-patient data files.

- The use of offset Moving Averages

  This research uses moving averages to describe seizure and non-seizure groups, or signal segments. Different moving average approaches such as Simple Moving Average, Weighted Moving Average and Exponential Moving Averages can be used to determine changes in signal segments. At this point, the proposed method uses simple moving average as a parameter to detect change. The Methods section describe moving average offsets in greater detail.

In addition to the above points, an efficient EEG signal processing algorithm could be developed or existing signal processing algorithms can be enhanced for extracting and representing relevant information from the EEG signals without a significant loss of useful information. This implies **retaining parts of the EEG signal that are needed and discarding the ones not needed.** So far, there is not a single signal processing algorithm that offers "best" results.

6 STATEMENT OF WORK

Overview

This research seeks to predict the onset of a seizure for a patient as early as possible. Seizures are accompanied by change in the EEG signals from normal to abnormal activity that usually occurs for few seconds.

Since the seizure activity occurs for a comparatively smaller duration of time than non-seizure activity, it is difficult to separate seizure and non-seizure parts of the signal. The presence of outliers creates problems as they also occur in spike and wave patterns.

Details

The initial step distills EEG signals into its simplest representation while minimizing the loss of useful information that might as well have seizure content.

To understand the details of this research, consider the following sine wave below.

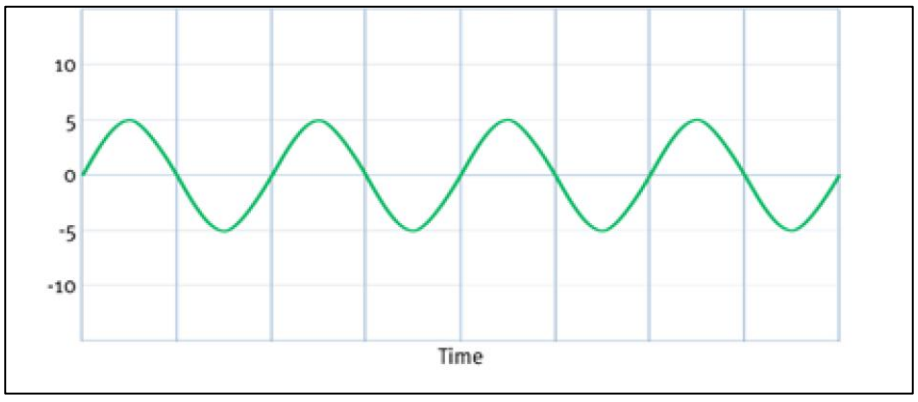

Figure 5: A stationary wave pattern

The initial step identifies all peaks and valleys as depicted below.

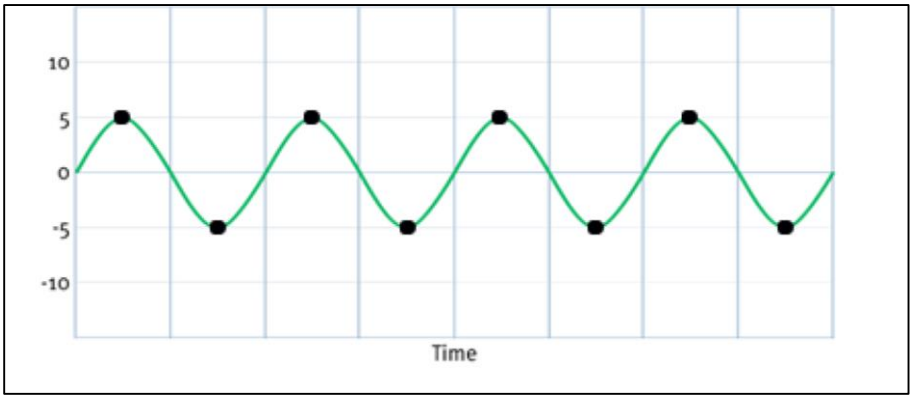

Figure 6: A stationary wave pattern with Peaks and Valleys Marked

Next, distances are calculated as illustrated with the lines below.

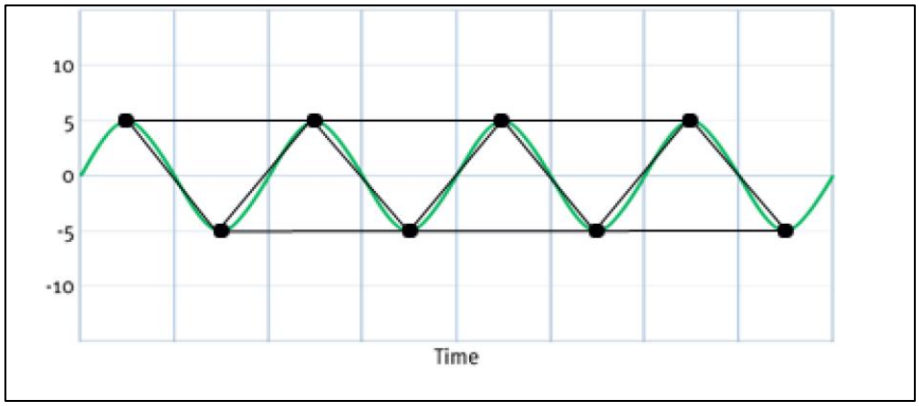

Figure 7: A stationary wave pattern with Peaks/Valleys and Distances

Similar distances are eliminated.

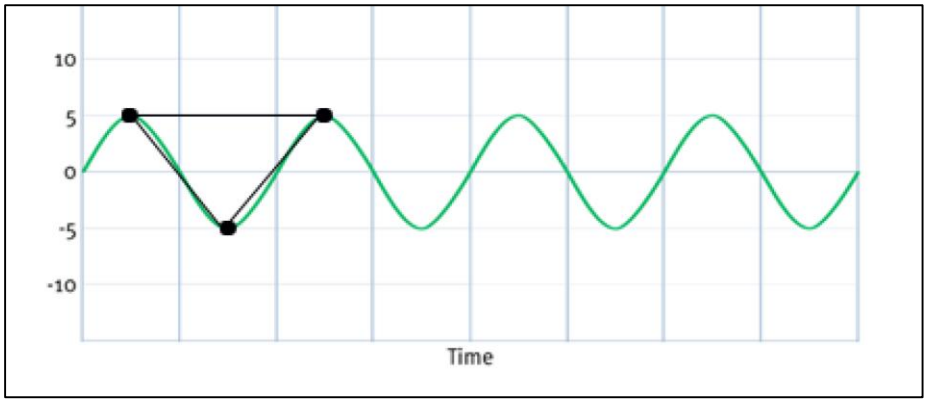

Figure 8: A stationary wave pattern with Reduced Peaks/Valleys and Distances

The final representation would consist of reduced number of peaks and valleys by removing the similar distances that keep repeating throughout the waveform and hence reducing the representation.

Anticipated issues modeling EEG waves with this approach include:

- Tracking significant amplitude and frequency changes by determining the *peak* and *valley* positions.
- Reducing the number of repeating symmetrical points, this would further be useful in removing unwanted segments of the EEG waves.

Challenges representing EEG Waves

The contrived example above assumes an extensive amount of symmetry in the wave pattern. Actual EEG waves contain an extensive amount of chaos as seen in Figure 9 below. Thus, making the representation much more challenging.

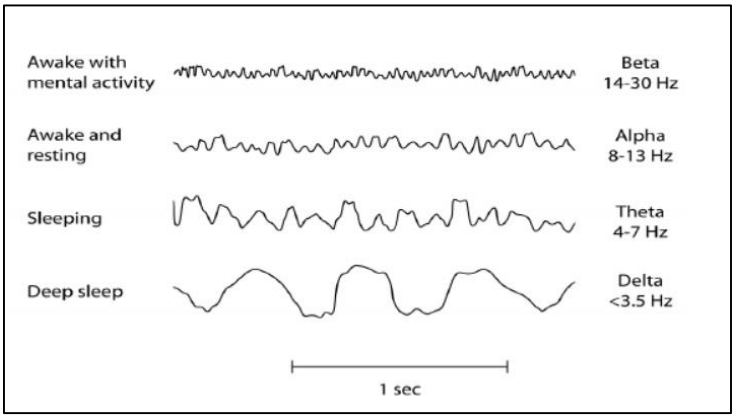

Figure 9: EEG frequency bands

In contrast to the stationary sine wave represented in Figure 5, the EEG waves in Figure 9 are constantly changing in terms of successive *peak* and successive *valley* distances due to changes in frequencies. Furthermore, even the amplitudes are continuously changing. In

such a case, where the waves are non-stationary and no definite pattern is observed, it is a challenging task to succinctly represent the EEG waves. Publically available EEG data sets of normal subjects can be studied for exploring various EEG patterns. Further, data mining approaches can be applied to categorize EEG data based on required applications.

Considering the chaotic nature of EEG data, this proposed method uses an algorithm that efficiently represents non-stationary signals with minimal loss of information. Only the trivial data points of the signal are discarded. A more detailed description of relevant points in the signal can be understood from the details described in the methods section.

7 EXPERIMENTS

7.1 Description of Data

The experiments use a publically available database collected from the Boston Children's Hospital [Venable et al. 2000]. The database consists of EEG recordings from 22 pediatric patients. Each patient folder contains EDF files with seizure/non-seizure data. Most of the files consist of one hour of data, while a few others have two to four-hours of data. The files adopt the follow naming convention chb*xx_nn*, where *xx* represents patient case and *nn* represents the file number. For instance, chb01_03 refers to third file for first patient.

The European Data Format (EDF) is a standard format for representation and exchange of medical time series data between different systems [Kemp et al. 2003]. The programming language R provides a package called *edfReader* for reading EDF files. Once imported, EDF files can be converted to CSV format for processing the EDF data.

The recordings for CHB MIT database use the standard International 10-20 system. Some files also record Electrocardiogram(ECG) data. The respective folders for all the patients in CHB MIT database contain a summary file with a description of the actual seizure start and end times. For each patient, multiple EDF files exist with multiple seizures events.

The initial experiments use the chb01 (patient 1) EEG recordings from CHB-MIT database. The EDF files are imported into R Studio for analysis. In chb01, the "01" indicates the patient ID. This folder consists of EEG recordings of an 11-year-old female subject. These

EDF files are a mixture of seizure files and normal EEG data. Likewise, other patient folders also contain a mixture of seizure and non-seizure files.

7.2 Methods

All experiments use the programming language R, version 3.4.2, and RStudio Version 1.1.383. R provides statistical computing and data mining functionalities with graphics support [1]. It also produces high quality plots for time-series analysis. R offers over 10,000 packages.

The original voltage values sampled at 256Hz. The proposed R algorithm discards any trivial data and considers only relevant signal values of interest in terms of peaks and valleys.

The basic requirement requires finding peaks and valleys. After calculating the maximum number data points from the EDF file, local peaks and local valleys are determined. Peaks and valleys are defined as follows.

- *Peak*: A data point ‘d[i]’is defined as a peak if the voltage values are higher than its neighbors [i-n] and [i+n] respectively. Hence, d[i] is a peak if it satisfies the following condition:

  d[i] > d[i-n] and d[i] > d[i+n], where *n* may range from 1 to k.

- *Valley*: A data point ‘d[i]’is defined as a valley if the voltage values are lower than its neighbors [i-n] and [i+n] respectively. Hence, d[i] is a valley if it satisfies the following condition:

d[i]<d[i-n] and d[i]<d[i+n], where n may range from 1 to k.

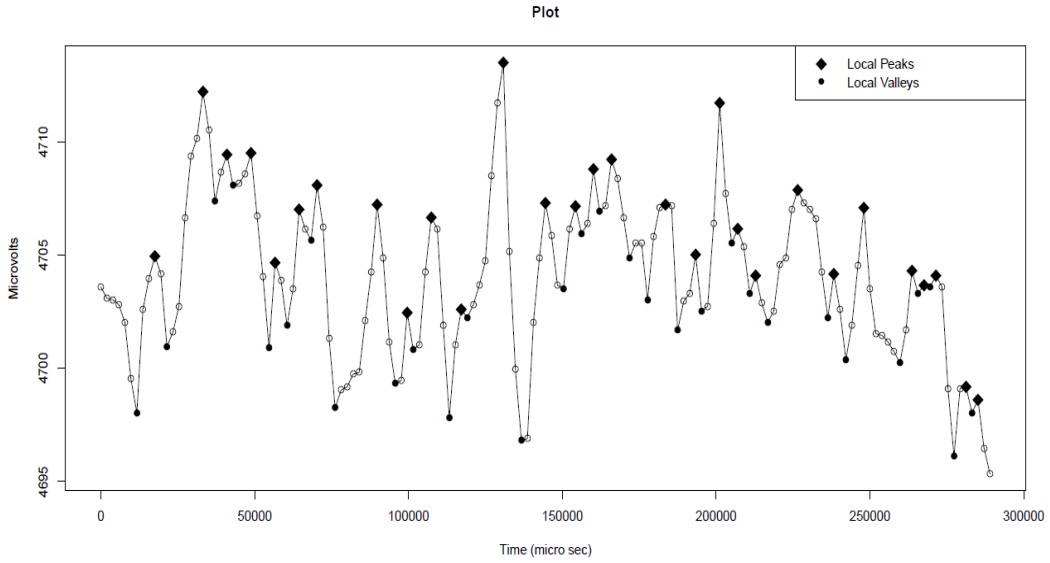

Figure 10: Local Peaks and Local Valleys (Open Circle Points are Ignored)

After the local peaks and valleys are determined, the distance between successive peaks and valleys is calculated. The median of all the distances is the reference value. The set of peak and valley points with distances greater than the reference distance are the candidates for prominent peaks and valleys. They represent abrupt peak to valley and valley to peak changes in the signal. This approach tracks changes in signal amplitudes. Despite the ability to locate peaks and valleys, the following problems arose:

- There are many consecutive peaks (and valleys) that are very close to each other.
- Signal patterns with trivial size represent information that does not add any value. The main problem here is to ignore such peaks and valleys.

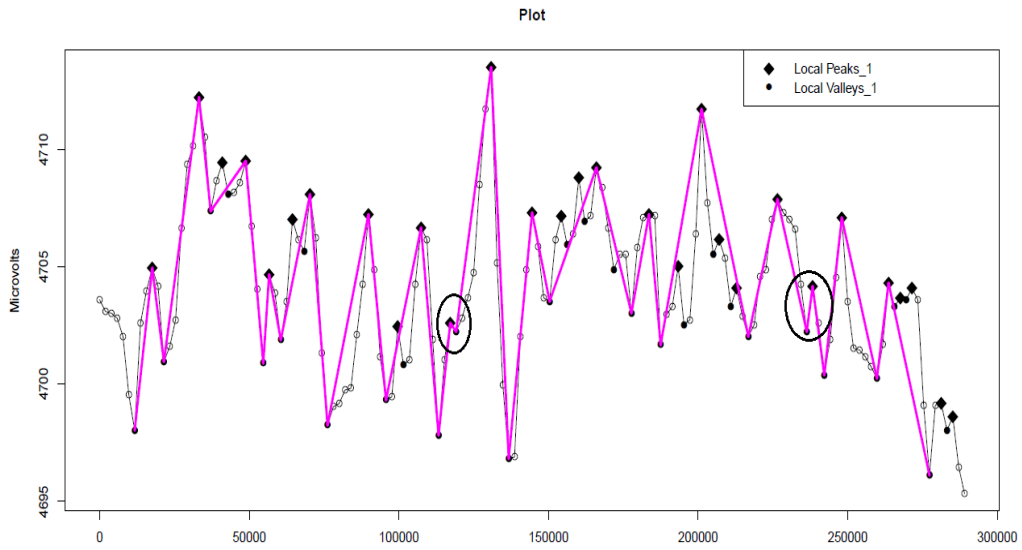

Figure 11: Refined peaks and valleys (Local Peaks_1 and Local Valleys_1)

For resolving the above-mentioned problems, the Euclidian distance between every pair of consecutive Local Peaks_1 (and Local Valleys_1) is calculated. The calculated distances are sorted and one-third of the median distance is taken as the reference distance. If the distance between any two successive Local Peaks_1 or Local Valleys_1 is greater than the new reference distance, then only those peak and valley points are considered and the rest are ignored. Furthermore, if the height of any of the Local Peaks_1 is higher than the previous and the next peak, then again a reference measure is taken to eliminate previous and the next peak. Only the highest peak is retained. Same process applies to Local valleys_1. Only this time the lowest Local valleys_1 is considered ignoring the comparatively higher previous and the next Local valleys_1. These peaks (Local Peaks_2) and valleys (Local Valleys_2) capture dominant information about the signal. Finally, the

identified Local Peaks_2 and Local Valleys_2 are connected through lines to obtain triangular patterns.

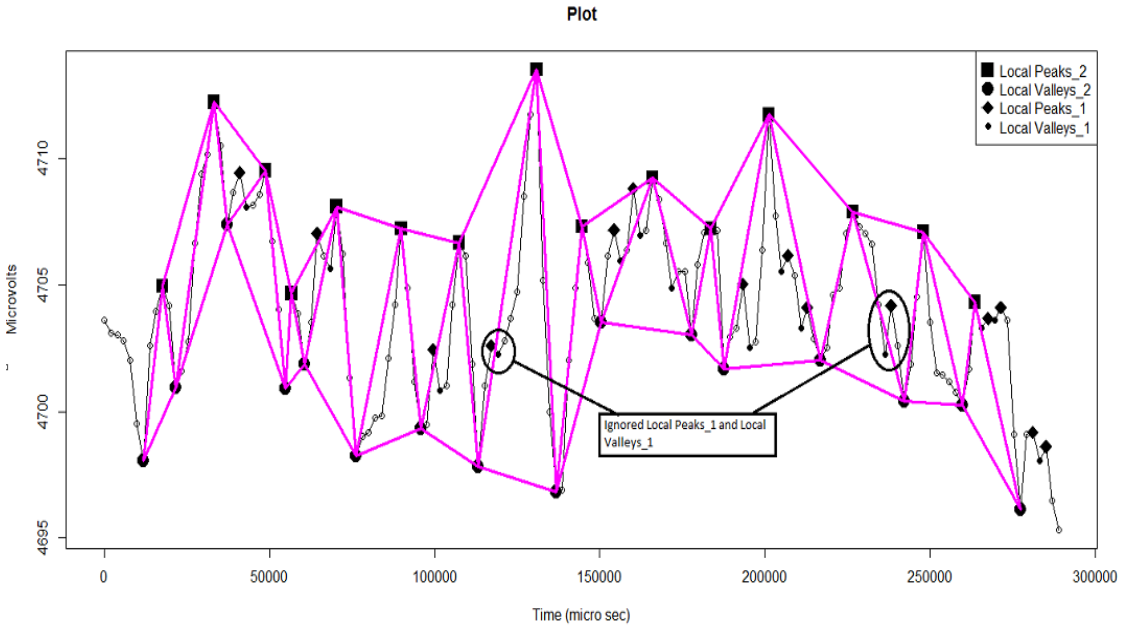

Figure 12: Refined peaks and valleys (Local Peaks_2 and Local Valleys_2)

Volatility of each triangle is the Euclidean distance between highest peak/valley and the lowest valley/peak of the three points forming the triangle. These are stored in a list. Once the volatilities are derived for all the triangles formed using extracted data from the EDF file, the concept of moving averages is introduced. This approach applies to all chb01 EDF seizure files.

The obtained volatilities for the entire EEG recording in each of the EDF file are further divided into groups. In this context, a *group*, see Figure 13, refers to collection of volatilities that are indexed from the beginning till the end within a group. Since, the

average for each group is calculated in a sliding window, these are referred to as *Moving Average Groups* throughout the text.

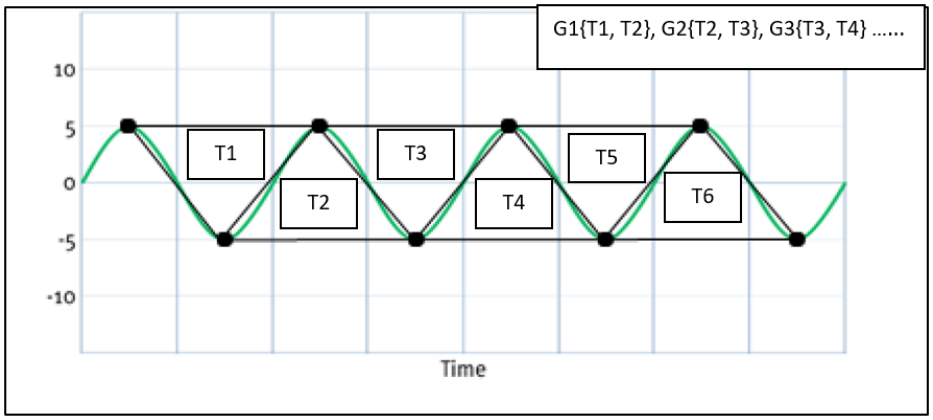

Figure 13: Triangle Groups without Offset

Assuming there are *M* moving average groups (G_1, G_2, G_3, G_4, …. G_M) generated from any of the EDF seizure files, the R Script calculates the average of all the volatilities in each group. *N* period indicates number of volatilities in each group. In another way, the length of each group is referred to as *N period*. In this case group G_1 consists of *N* (V_1, V_2, V_3, V_4, ….V_N) volatility values.

Moving Average Offset

Once the volatilities are obtained, the idea of an *offset* with respect to moving average groups is introduced. Without an *offset*, a 10 period simple moving average would use volatilities(triangles) inside groups in the order {V_1, V_2…. V_{10}}, {V_2, V_2...V_{11}}, {V_3, V_2…...V_{12}} … and so on. However, using an *offset* value of 4 changes the ordering of

volatilities to {V_1, V_2…. V_{10}}, {V_5, V_6.....V_{14}}, {V_9, V_{10}…...V_{18}} … and so on. Using an offset generates multiple permutations of settings. This allows for an exhaustive approach that finds optimized settings. Figure 14 illustrates the idea of groups with an offset.

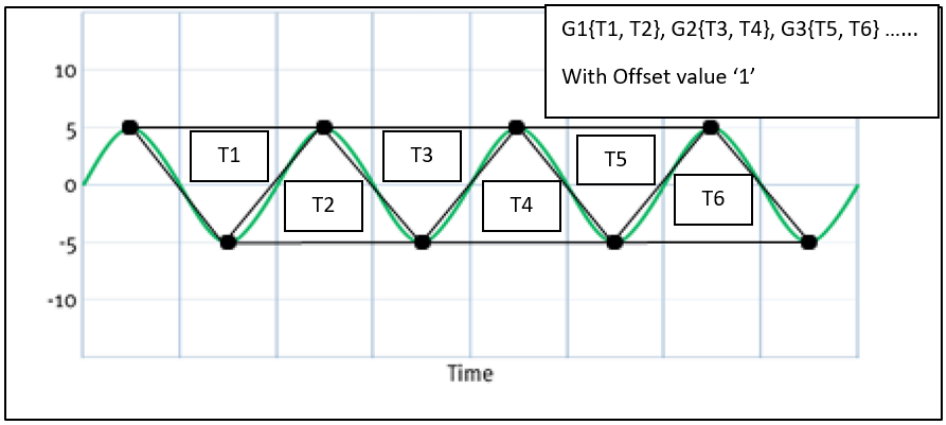

Figure 14: Groups with offset value as '1'

For any two adjacent groups G_1 and G_2 (assuming G_1 occurs before G_2 on the time line), following permutations are possible:

a) G_1 is non-epileptic G_2 is epileptic. (transition from pre-ictal to ictal stage)

b) G_1 is epileptic G_2 is non-epileptic (transition from ictal to post-ictal stage)

c) Both G_1 and G_2 are epileptic (inter-ictal stage)

d) Both G_1 and G_2 are non-epileptic (transition within normal stage)

The transition from non- epileptic to epileptic group (pre-ictal to ictal stage) is of interest.

Two approaches that determine the change between adjacent groups:

- **Absolute**

  This includes absolute changes and classifies a group as epileptic or non-epileptic.

- **Relative**

  This approach calculates relative percentage change in the average values of two successive moving average groups in pairs {{G_1, G_2}, {G_2, G_3}, {G_3, G_4} ……{$G_{(M-1)}$, G_M}}. Let the percentage change be denoted by C_1, C_2, C_3, C_4…$C_{(M-1)}$. For any change C_K, the percentage change between previous groups is observed. If {C_1, C_2, C_3, C_4 ……$C_{(K-1)}$} all are less than C_K, then seizure begins at the last volatility value for group $G_{(K+1)}$, i.e. V_N for $G_{(K+1)}$.

7.3 Experiment 1

Goal: To optimize the settings for the patient chb01_03 seizure file.

Experiment 1 uses the seizure data extracted from the R GUI application. The first application (as shown in figure 15) takes input with the following options: *EDF File path*, *Select Channel*, *Select Time Interval.*

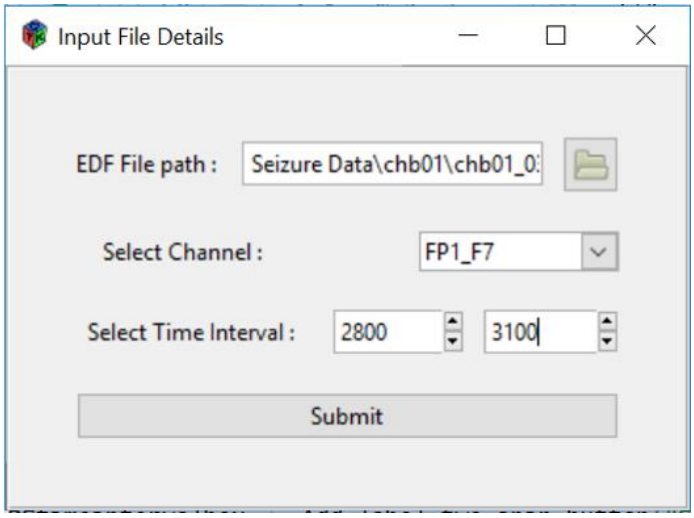

Figure 15: First Input Window

Details of the screen layout follow.

- **EDF seizure file path**: For this experiment, patient 1 seizure file chb01_03.edf is selected.
- **Channel Information**: Out of the 23 channel electrodes, any one of the channels can be selected. However, for this experiment, channel FP1_F7 is selected on a random basis.

- **Time Interval**: Since the actual seizure onset and ending time are already known, an approximate interval of 100 seconds before the onset and 100 seconds after the seizure ends is provided in the interval field. This selection is manual. For dataset chb01_03, the actual seizure begins at 2996 sec and ends at 3036 seconds according to the description in the CHB MIT database. Out of the 1-hour data available in the chb01_03 EDF file, only the data within the provided interval is extracted and experimented with. Clicking on *Submit* button generates two more windows: the plot window with the extracted data (figure 16) and the parameter setting window (figure 18) that takes further inputs.

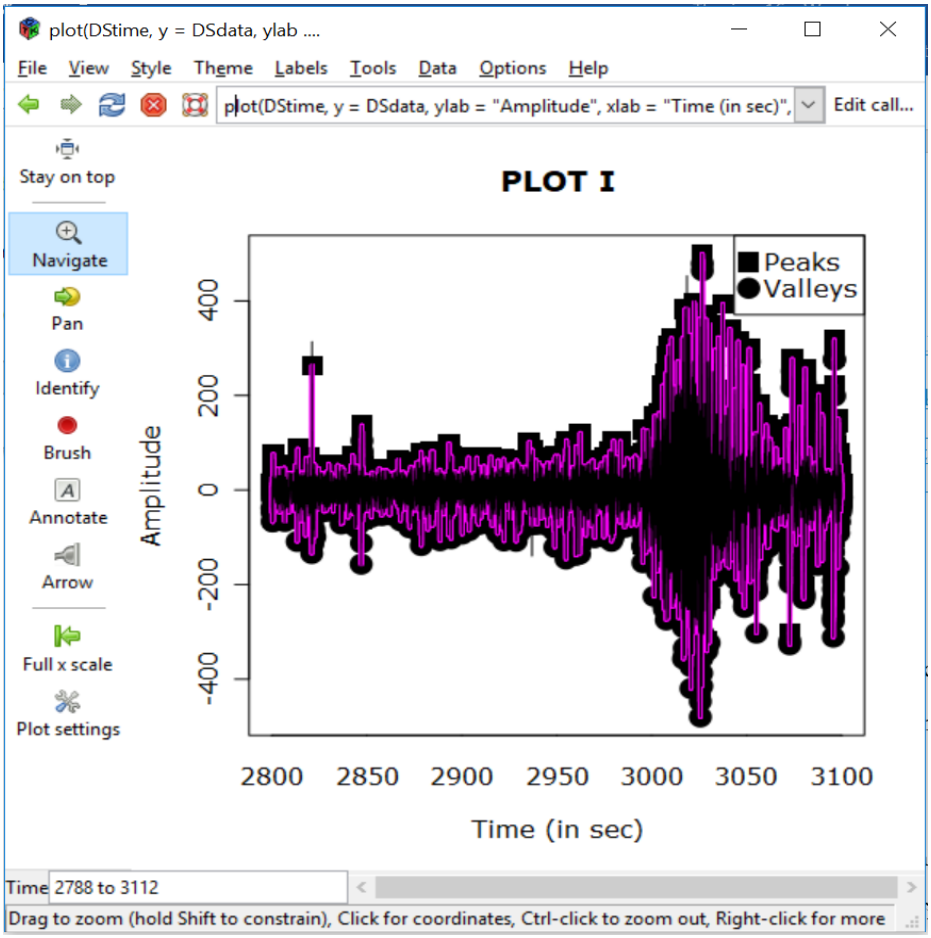

Figure 16: Plot for chb01_03 data (2800 to 3100 sec)

The above plot shows the extracted data between 2800 seconds to 3100 seconds for chb01_03 EDF file. These plots have zooming capability for better viewing of data. Figure 17 below zooms in using a range of 2879 to 2880, which corresponds to 1 second's worth of data and gives a finer granularity view.

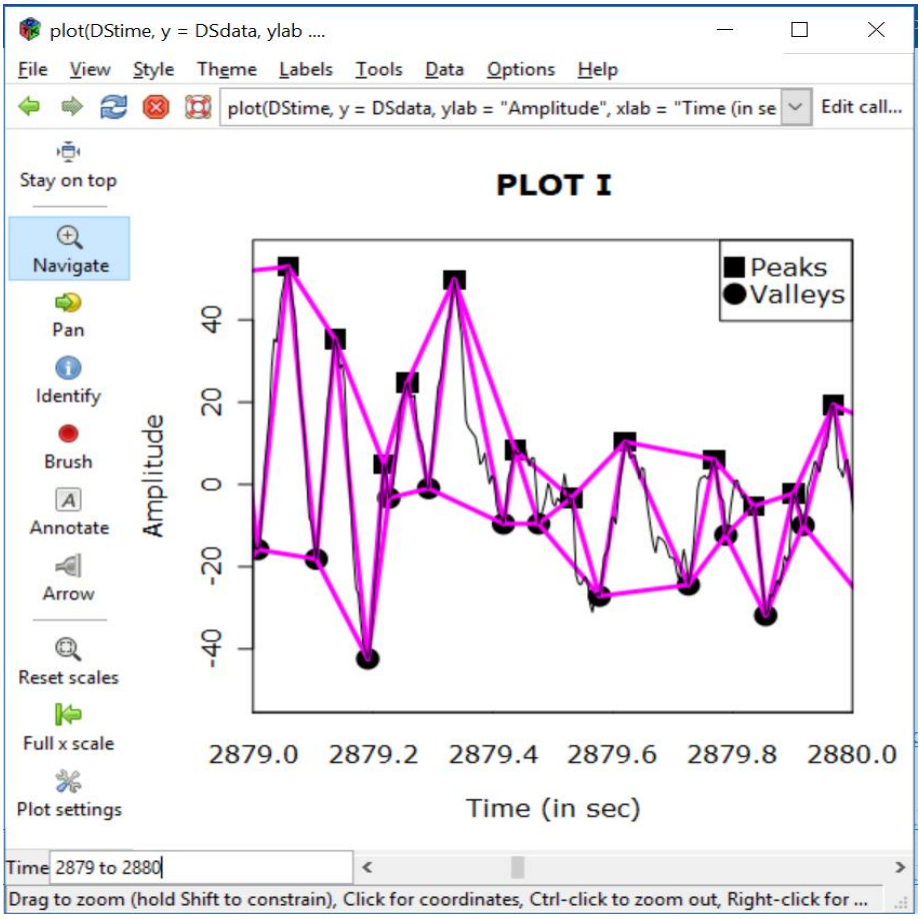

Figure 17: Detailed view of 1 second data (2879 to 2880 seconds) for chb01_03.edf

The fine-grained view in figure 17 shows the data encompassed (in the form of peaks and valleys) into triangular patterns. It also confirms the fact that maxima (peaks) and minima (valleys) cover all the data with insignificant loss of data (if at all).

Figure 18 below shows the second window that appears with the plot window.

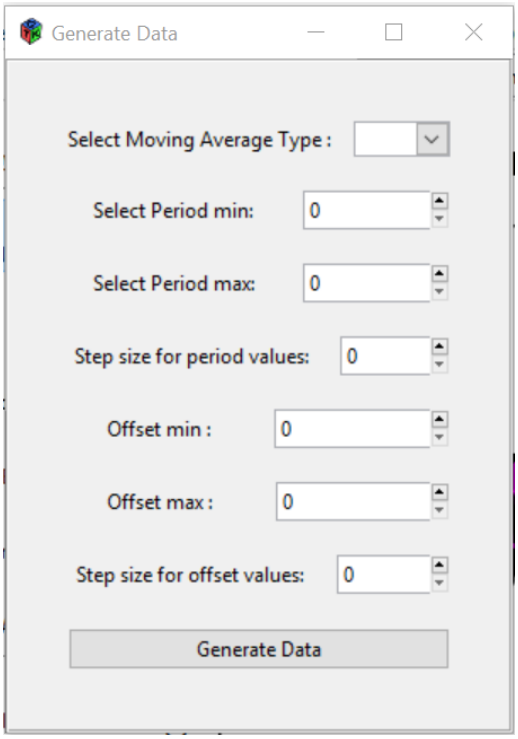

Figure 18: Offset Moving Average Window

The GUI allows for the selection of moving average type (SMA, in this case), period range, offset range and step sizes. Providing an *N* period range (both min and max), an offset range (both min and max) and their respective step sizes, generates a CSV file (Table 3). Each row in the file consists of a period value, offset value, group start time and group end time. Group end times refer to the time where the last volatility for any particular group occurs. Group end times indicate the onset of seizure.

Table 3. Sample from generated data

Setting#	Nth_Period	Step size for N Period	Offset	Step size for Offset	Start_Time	End_Time
1	20	10	150	10	3016.67	3016.91
2	30	10	150	10	3016.67	3017.05
3	40	10	150	10	3018.26	3018.68

Multiple permutations and combinations of range values in the R GUI window for patient chb01 file# chb01_03 generates multiple CSV files. The EDF data is converted into CSV format because R offers limited functionality for using and manipulating EDF content. Table 4 shows the settings which provided the best results.

However, **setting 1** (as shown in figure 19) generates the best results for chb01_03. Setting 1 populates an output CSV file with 774 sub-settings. These sub-settings are derived from the values provided in the **setting 1** GUI window and one of these sub-settings generates closest possible seizure start time to the actual seizure time. For this data file, the actual seizure starts at 2996 seconds as provided in the summary details of CHB MIT database for file# chb01_03. The experiment predicts seizure onset at 3001.70 seconds with a latency of 5.7 seconds.

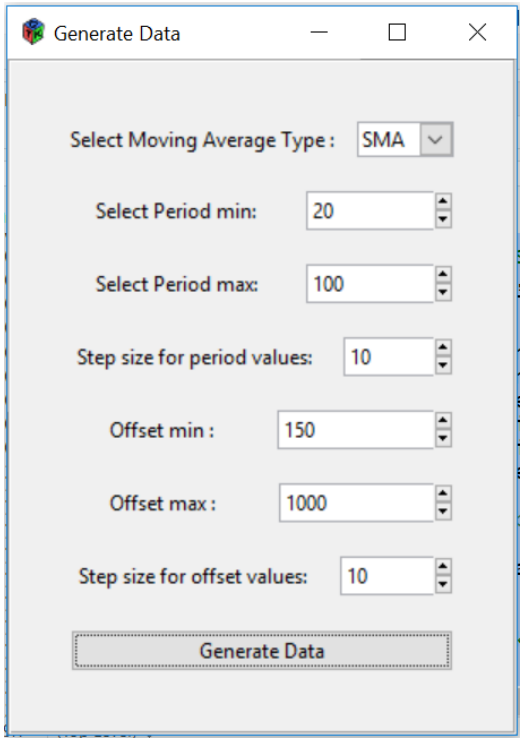

Figure 19: GUI Setting 1 for Generating data

Since setting 1 generates optimum results for chb01_03, it is selected as a training parameter for other seizure files in patient chb01 folder. The same setting is applied to testing data sets chb01_04, chb01_15, chb01_16, chb01_18, chb01_21 seizure files for chb01. These are other seizure files in patient folder chb01.

Table 4. Settings with Best Result chb01_03

Setting#	Nth_Period	Step size for N Period	Offset	Step size for Offset	Start_Time	End_Time
631	20	10	850	10	2954.74	2955.57
478	20	10	680	10	2954.77	2955.62
632	30	10	850	10	2954.74	2956.07
568	20	10	780	10	2978.62	2979.57
532	**20**	**10**	**740**	**10**	**3000.54**	**3001.70**

7.4 Experiment 2

Goal: Identify common sub-settings that predict seizure onset for all chb01 files.

This experiment uses the optimized setting from experiment 1, which generates CSV files for testing patient chb01's data sets chb01_04, chb01_15, chb01_16, chb01_18 and chb01_21. Each generated file consists of 774 sub-settings. Within each file, the sub-setting that gives the closest results in terms of the seizure start time is selected. For setting 1, some of the testing datasets detect seizure too early, while a few others predict seizure onset with minimal latency.

7.4.1 Experiment 2A (Using the 999 Penalty)

In this experiment, a penalty of '999' is applied to sub-settings that predict seizure start time way too early than the actual seizure start time. The results are sorted in ascending order on group end times, then again sorted by setting#. The remaining sub-settings are given rankings from 1 onwards till the last settings number. The rankings are reproduced for all chb01 files.

The new file with combined rankings of all files is again sorted in increasing order of the sum of ranks (as shown in Table 5). This affects the cumulative rankings of the sub-settings for all seizure files within the same patient chb01_03. As a result, the sorted rankings degrade for sub-settings with early seizure predicting times.

The sub-settings shown in figure d do not exhibit any particular pattern, they are scattered throughout the file. Since there is no consistent pattern of sub-settings, it is unclear which set of settings compete for predicting onset of seizure. Although, the cumulative rankings are available for all 774 sub-settings for all chb01 files, the rankings do not indicate a prevalent group based on settings.

Table 5: Sub-settings in Ascending order using sum of the ranks with '999' penalty

Setting#	chb01_03	chb01_04	chb01_15	chb01_16	chb01_18	chb01_21	Sum of Ranks	Average Rank
671	95	51	72	270	80	125	693	115.50
389	29	85	174	50	184	207	729	121.50
433	7	37	357	35	170	162	768	128.00
434	10	40	369	37	173	170	799	133.17
121	76	252	335	124	10	46	843	140.50
47	25	61	426	106	181	77	876	146.00
239	93	15	176	82	144	385	895	149.17
122	89	267	342	140	14	54	906	151.00
240	101	17	184	94	152	396	944	157.33
340	40	180	26	348	205	165	964	160.67
241	108	19	192	110	160	408	997	166.17
341	42	193	36	352	209	172	1004	167.33
342	43	208	48	356	213	175	1043	173.83
127	28	486	170	51	116	208	1059	176.50
665	60	49	71	217	603	69	1069	178.17
164	208	72	20	38	352	400	1090	181.67
125	115	317	364	181	27	95	1099	183.17
666	71	50	79	231	608	84	1123	187.17
390	35	92	182	59	194	572	1134	189.00
165	215	81	29	42	366	412	1145	190.83
46	5	54	413	92	180	406	1150	191.67
126	122	337	374	196	34	109	1172	195.33
739	332	334	171	20	310	10	1177	196.17
166	226	88	39	48	376	421	1198	199.67
711	236	163	407	6	20	373	1205	200.83

7.4.2 Experiment 2B (No penalty for early detection)

Unlike the previous experiment 2A, this experiment does not impose any penalty on sub-setting rankings. Instead of '999', a '0' value is added to the sub-setting rankings, hence there is no effect on the cumulative rankings for all chb01 files.

Within each file, the data is sorted by group end time in an increasing order. All the settings that detect seizure too early, i.e. before the predicted seizure start time, are ranked '0', this time with no penalty. The remaining ones are ranked 1,2,3…till the end. The entire csv file is again sorted in the original order of setting numbers (1-774). This is repeated for each csv file. A sample for chb01_04 can be seen in Table 6.

Table 6. Ranked sub-settings for chb01_04

Setting#	N_period	Step size for N Period	Offset	Step size for Offset	Start Time	End Time	Rank
1	20	10	150	10	1494.46	1495.46	616
2	30	10	150	10	1502.78	1503.69	672
3	40	10	150	10	1494.46	1496.05	629
4	50	10	150	10	1483.46	1483.84	347
5	60	10	150	10	1483.46	1483.91	368
6	70	10	150	10	1483.46	1483.97	389
7	80	10	150	10	1483.46	1484.04	409
8	90	10	150	10	1483.46	1484.11	425
9	100	10	150	10	1483.46	1484.20	441
10	20	10	160	10	1458.96	1460.59	0
11	30	10	160	10	1458.96	1461.14	0
12	40	10	160	10	1483.44	1483.73	324
13	50	10	160	10	1483.44	1483.81	342
14	60	10	160	10	1483.44	1483.89	363
15	70	10	160	10	1483.44	1483.96	384
16	80	10	160	10	1483.44	1484.02	404
17	90	10	160	10	1483.44	1484.10	421
18	100	10	160	10	1483.44	1484.19	438
19	20	10	170	10	1468.98	1469.67	1
20	30	10	170	10	1468.98	1470.48	2

A new observation sheet consists of the following column headers: sub-setting#, names of testing files in separate column headers followed by a column for sum of ranks. The file

has 774 rows which represent sub-settings. The sum of ranks for each testing file is determined. The sheet data is again sorted in the increasing order of sum of ranks for sub-settings (1-774).

An interesting pattern can be observed from this experiment. A cluster pattern of settings ranging from sub-setting# (670-700) is visible, each of which is under top 15 among all 774 ranked sub-settings. However, for the remaining ranked setting numbers no such cluster appears. A fragment of the results can be seen in Table 7.

Table 7. Sub-settings in ascending order of sum of the ranks for all chb01 files no penalty

Setting#	chb01_03	chb01_04	chb01_15	chb01_16	chb01_18	chb01_21	Sum of Ranks
685	137	0	94	0	81	190	502
694	3	0	113	0	245	232	593
695	167	0	0	0	258	237	662
676	111	0	398	0	5	159	673
696	171	0	0	0	268	240	679
697	173	0	0	0	273	242	688
671	95	51	72	270	80	125	693
689	155	0	118	0	217	218	708
389	29	85	174	50	184	207	729
690	160	0	122	0	229	223	734
691	163	0	125	0	244	227	759
433	7	37	357	35	170	162	768
698	174	73	0	0	276	247	770
692	166	0	127	0	257	231	781
699	175	82	0	0	281	255	793
712	214	125	0	435	6	14	794
434	10	40	369	37	173	170	799
693	170	0	130	0	267	236	803
121	76	252	335	124	10	46	843
47	25	61	426	106	181	77	876
605	301	3	0	507	69	6	886
239	93	15	176	82	144	385	895
631	0	541	0	23	334	0	898
122	89	267	342	140	14	54	906
307	407	38	198	0	95	193	931

Tables 9 to 13 show results for setting# 685, 694,695, 676 and 696 respectively. The experiment calculates the seizure start time for each setting for all chb01 files. Latency can be defined as follows:

Latency = (Predicted Seizure Start Time - Actual Seizure Start Time) (Eq. 1)

Latency values can be positive or negative. Detecting a seizure after it occurs indicates a positive latency. Negative latency refers to seizures predicted before they actually occurred. The average latencies vary across all five settings. For patient chb01_03, the settings with average latencies is shown in the below table 8.

Table 8. Average latency for all chb01 files using all 5 settings

Setting #	Average Start Time(Predicted)	Average Start Time(Actual)	Average Latency (sec) for all chb01 seizures
685	1550.93	1542.83	8.1
694	1552.27	1542.83	9.44
695	1552.93	1542.83	10.1
676	1546.65	1542.83	3.82
696	1553.22	1542.83	10.39

Table 9. Results with setting# 685 when applied to all chb01 files

Setting# 685 File#	Predicted Seizure Start Time(sec)	Actual Seizure Start Time(sec)	Latency(sec)
chb01_03	3012.93	2996	16.93
chb01_04	1417.59	1467	-49.41
chb01_15	1740.76	1732	8.76
chb01_16	1008.17	1015	-6.83
chb01_18	1745.06	1720	25.06
chb01_21	381.06	327	54.06

Table 10. Results with setting# 694 when applied to all chb01 files

Setting# 694 File#	Predicted Seizure Start Time(sec)	Actual Seizure Start Time(sec)	Latency(sec)
chb01_03	3003.56	2996	7.56
chb01_04	1417.58	1467	-49.42
chb01_15	1742.57	1732	10.57
chb01_16	1008.66	1015	-6.34
chb01_18	1758.27	1720	38.27
chb01_21	383.01	327	56.01

Table 11. Results with setting# 695 when applied to all chb01 files

Setting# 695 File#	Predicted Seizure Start Time(sec)	Actual Seizure Start Time(sec)	Latency(sec)
chb01_03	3013.64	2996	17.64
chb01_04	1459.85	1467	-7.15
chb01_15	1693.64	1732	-38.36
chb01_16	1008.75	1015	-6.25
chb01_18	1758.51	1720	38.51
chb01_21	383.2	327	56.2

Table 12. Results with setting# 676 when applied to all chb01 files

Setting# 676 File#	Predicted Seizure Start Time(sec)	Actual Seizure Start Time(sec)	Latency(sec)
chb01_03	3012.41	2996	16.41
chb01_04	1377.23	1467	-89.77
chb01_15	1774.71	1732	42.71
chb01_16	1007.75	1015	-7.25
chb01_18	1728.64	1720	8.64
chb01_21	379.16	327	52.16

Table 13. Results with setting# 696 when applied to all chb01 files

Setting# 696 File#	Predicted Seizure Start Time(sec)	Actual Seizure Start Time(sec)	Latency(sec)
chb01_03	3013.74	2996	17.74
chb01_04	1460.64	1467	-6.36
chb01_15	1694.14	1732	-37.86
chb01_16	1008.84	1015	-6.16
chb01_18	1758.64	1720	38.64
chb01_21	383.31	327	56.31

7.5 Experiment 3

Goal: To test optimized settings for chb01 on chb02, chb03 and chb04 patient files.

The previous experiments predict seizures for the same patient(chb01). This experiment seeks to examine generalizability of the algorithm and predicts the onset of seizure for other patients in CHB MIT dataset. Training data uses chb01 settings which applies the settings to the chb02, chb03 and chb04 seizure files.

The optimized sub-settings (ranked top 5) derived from experiment 2B are tested for all seizure files in patient chb02, chb03 and chb04 folders. Folders chb02, chb03 and chb04 consist of 3, 7 and 3 EDF seizure files respectively. File chb04_28, which is one of the seizure files for patient chb04, consists of two seizures within the same file. Tables 14(a-e), 15(a-e), 16(a-e) show the experimental results.

Table 14 a: Results with chb01 cluster setting# 685 applied to patient chb02 files

Setting# 685 File#	Predicted Seizure Start Time(sec)	Actual Seizure Start Time(sec)	Latency (sec)
chb02_16	194.59	130	64.59
chb02_16_plus	2847.25	2972	-124.75
chb02_19	3487.83	3369	118.83

Table 14 b: Results with chb01 cluster setting# 694 applied to patient chb02 files

Setting# 694 File#	Predicted Seizure Start Time(sec)	Actual Seizure Start Time(sec)	Latency (sec)
chb02_16	221.05	130	91.05
chb02_16_plus	2847.39	2972	-124.61
chb02_19	3316.47	3369	-52.53

Table 14 c: Results with chb01 cluster setting# 695 applied to patient chb02 files

Setting# 695 File#	Predicted Seizure Start Time(sec)	Actual Seizure Start Time(sec)	Latency (sec)
chb02_16	221.11	130	91.11
chb02_16_plus	2910.2	2972	-61.8
chb02_19	3260.07	3369	-108.93

Table 14 d: Results with chb01 cluster setting# 676 applied to patient chb02 files

Setting# 676 File#	Predicted Seizure Start Time(sec)	Actual Seizure Start Time(sec)	Latency (sec)
chb02_16	234.14	130	104.14
chb02_16_plus	3069.25	2972	97.25
chb02_19	3485.53	3369	116.53

Table 14 e: Results with chb01 cluster setting# 696 applied to patient chb02 files

Setting# 696 File#	Predicted Seizure Start Time(sec)	Actual Seizure Start Time(sec)	Latency (sec)
chb02_16	221.16	130	91.16
chb02_16_plus	2910.3	2972	-61.7
chb02_19	3260.14	3369	-108.86

For patient chb02, optimized settings predict seizure with an average delay of 8.76 seconds. For patient files chb03 and chb04, the experiments detect seizure with an average of 13.9 and 24.82 seconds respectively.

Table 15 a: Results with chb01 cluster setting# 685 applied to patient chb03 files

Setting# 685 File#	Predicted Seizure Start Time(sec)	Actual Seizure Start Time(sec)	Latency (sec)
chb03_01	375.27	362	13.27
chb03_02	747.89	731	16.89
chb03_03	445.91	432	13.91
chb03_04	2112.83	2162	-49.17
chb03_34	1953.24	1982	-28.76
chb03_35	2516.76	2592	-75.24
chb03_36	1631.02	1725	-93.98

Table 15 b: Results with chb01 cluster settings# 694 applied to patient chb03 files

Setting# 694 File#	Predicted Seizure Start Time(sec)	Actual Seizure Start Time(sec)	Latency (sec)
chb03_01	375.85	362	13.85
chb03_02	822.02	731	91.02
chb03_03	446.72	432	14.72
chb03_04	2231.27	2162	69.27
chb03_34	2019.71	1982	37.71
chb03_35	2480.3	2592	-111.7
chb03_36	1737.79	1725	12.79

Table 15 c: Results with chb01 cluster setting# 695 applied to patient chb03 files

Setting# 695 File#	Predicted Seizure Start Time(sec)	Actual Seizure Start Time(sec)	Latency (sec)
chb03_01	375.92	362	13.92
chb03_02	822.13	731	91.13
chb03_03	446.8	432	14.8
chb03_04	2231.38	2162	69.38
chb03_34	1997.41	1982	15.41
chb03_35	2480.36	2592	-111.64
chb03_36	1737.9	1725	12.9

Table 15 d: Results with chb01 cluster setting# 676 applied to patient chb03 files

Setting# 676 File#	Predicted Seizure Start Time(sec)	Actual Seizure Start Time(sec)	Latency (sec)
chb03_01	374.7	362	12.7
chb03_02	755.33	731	24.33
chb03_03	463.83	432	31.83
chb03_04	2244.32	2162	82.32
chb03_34	1953.16	1982	-28.84
chb03_35	2685.18	2592	93.18
chb03_36	1745.76	1725	20.76

Table 15 e: Results with chb01 cluster setting# 696 applied to patient chb03 files

Setting# 696 File#	Predicted Seizure Start Time(sec)	Actual Seizure Start Time(sec)	Latency (sec)
chb03_01	375.98	362	13.98
chb03_02	822.18	731	91.18
chb03_03	446.91	432	14.91
chb03_04	2231.49	2162	69.49
chb03_34	1997.49	1982	15.49
chb03_35	2480.41	2592	-111.59
chb03_36	1851.45	1725	126.45

Table 16 a: Results with chb01 cluster setting# 685 applied to patient chb04 files

Setting# 685 File#	Predicted Seizure Start Time(sec)	Actual Seizure Start Time(sec)	Latency (sec)
chb04_05	7859.52	7804	55.52
chb04_08	6477.48	6446	31.48
chb04_28_a	1710.41	1679	31.41
chb04_28_a	3860.13	3782	78.13

Table 16 b: Results with chb01 cluster setting# 694 applied to patient chb04 files

Setting# 694 File#	Predicted Seizure Start Time(sec)	Actual Seizure Start Time(sec)	Latency (sec)
chb04_05	7707.61	7804	-96.39
chb04_08	6517.65	6446	71.65
chb04_28_a	1742.07	1679	63.07
chb04_28_a	3861.53	3782	79.53

Table 16 c: Results with chb01 cluster setting# 695 applied to patient chb04 files

Setting# 695 File#	Predicted Seizure Start Time(sec)	Actual Seizure Start Time(sec)	Latency (sec)
chb04_05	7707.71	7804	-96.29
chb04_08	6517.8	6446	71.8
chb04_28_a	1742.13	1679	63.13
chb04_28_a	3861.66	3782	79.66

Table 16 d: Results with chb01 cluster setting# 676 applied to patient chb04 files

Setting# 676 File#	Predicted Seizure Start Time(sec)	Actual Seizure Start Time(sec)	Latency (sec)
chb04_05	7874.75	7804	70.75
chb04_08	6565.07	6446	119.07
chb04_28_a	1747.92	1679	68.92
chb04_28_a	3850.37	3782	68.37

Table 16 e: Results with chb01 cluster setting# 696 applied to patient chb04 files

Setting# 696 File#	Predicted Seizure Start Time(sec)	Actual Seizure Start Time(sec)	Latency (sec)
chb04_05	7707.78	7804	-96.22
chb04_08	6508.16	6446	62.16
chb04_28_a	1742.22	1679	63.22
chb04_28_a	3861.78	3782	79.78

Table 17. Best Settings for all chb01, chb02, chb03 and chb04 files

Patient Seizure File#	Setting#	Nth_Period	Step size for N Period	Offset	Step size for Offset	Group Start_Time(sec)	Group End_Time(sec)
chb01_04	676	20	10	900	10	1376.35	1377.23
chb01_15	695	30	10	920	10	1691.75	1693.64
chb01_16	676	20	10	900	10	1007.60	1007.75
chb01_18	676	20	10	900	10	1728.51	1728.64
chb01_21	676	20	10	900	10	378.55	379.16
chb02_16	685	20	10	910	10	194.23	194.59
chb02_16plus	685	20	10	910	10	2847.13	2847.26
chb02_19	695	30	10	920	10	3259.86	3260.07
chb03_01	676	20	10	900	10	374.56	374.70
chb03_02	685	20	10	910	10	747.75	747.89
chb03_03	685	20	10	910	10	445.73	445.91
chb03_04	685	20	10	910	10	2111.85	2112.84
chb03_34	676	20	10	900	10	1953.00	1953.16
chb03_35	694	20	10	920	10	2480.17	2480.30
chb03_36	685	20	10	910	10	1630.92	1631.02
chb04_05	694	20	10	920	10	7707.47	7707.61
chb04_08	685	20	10	910	10	6476.83	6477.49
chb04_28a	685	20	10	910	10	1710.24	1710.41
chb04_28b	676	20	10	900	10	3850.14	3850.38

Patient demographic may contribute to the experimental results, where similar age patients show similar results due to the commonalities of their brain morphologies. There is a high chance that patients of similar age groups exhibit similar seizure patterns. Patient chb01 is a 11-year-old female, chb02 is a 11-year-old male, chb03 is 14-year-old female and chb04

is a 22-year-old male. The average latency for all test settings using chb01 increases with the age for other patients [Table 17].

Table 18. Average Prediction Delay for all patient files with all 5 settings

Patient#	Age	Gender	Average Latency for all files & settings(sec)
chb01	11	Female	8.36
chb02	11	Male	8.76
chb03	14	Female	13.9
chb04	22	Male	24.82

The above observation shows comparable average delays for patient chb01_01 and chb01_02 (with same age). However, the experiment requires more samples in order to statistically validate conclusions regarding patient demographics and the average prediction delays.

8 DISCUSSION

The CHB MIT data set contains about 1 hours' worth of readings. Considering the slow execution of plots in R, each data set is truncated to approximately 90 seconds before and after the seizure occurs. This allows for sufficient data points to perform the experiments. These points include pre-ictal, ictal and post-ictal data.

The experiments assess the results based on degree of deviation from the actual seizure time. All the experiments seek to minimize latency between the onset of a seizure and when it is predicted (after the onset). If a neurologist wants to detect the origin of seizure, then exactness of predicted onset timing is vital.

The preprocessing step distills key information from the original EDF data files by selecting critical points referred to earlier as peaks and valleys. The approach performs substantially well for all the files in the CHB MIT dataset.

EEG recordings use signals from 23 channels. To reduce the complexity, all the experiments focused on EEG signals of channel FP1_F7. Additional analysis of other channels may improve the results. Certainly fodder for future research.

Experiment 1 examines the best settings for chb01_03 that serves as a training parameter for other files. The negative latency values indicate false positives. It includes all results which predict the onset of seizure before it actually occurs. The optimized settings used for this experiment generated positive average latency for all chb01 files. The experiment also provided foundation settings for other experiments. The pre-processing step was part of the

first experiment. The best results for all 6 data sets were within 20 percent of the best rankings.

Considering the lack of correlation between results, the high ranking of the best settings for all 6 data sets showed a high level of consistency. Figures 20 and 21 show the correlation between chb01 files with and without penalty respectively.

Table 19. Correlation Matrix (For '999' penalty)

	chb01_03	**chb01_04**	**chb01_15**	**chb01_16**	**chb01_18**	**chb01_21**
chb01_03	1	0.16	-0.2	-0.35	0.14	0.13
chb01_04	0.16	1	0.14	0.11	-0.38	-0.43
chb01_15	-0.2	0.14	1	-0.38	-0.6	-0.18
chb01_16	-0.35	0.11	-0.38	1	0.08	-0.41
chb01_18	0.14	-0.38	-0.6	0.08	1	0.03
chb01_21	0.13	-0.43	-0.18	-0.41	0.03	1

Table 20: Correlation Matrix (No penalty)

	chb01_03	**chb01_04**	**chb01_15**	**chb01_16**	**chb01_18**	**chb01_21**
chb01_03	1	-0.34	-0.43	0.22	-0.13	0.08
chb01_04	-0.34	1	0.02	0.1	-0.06	-0.61
chb01_15	-0.43	0.02	1	-0.14	-0.48	-0.17
chb01_16	0.22	0.1	-0.14	1	-0.62	-0.64
chb01_18	-0.13	-0.06	-0.48	-0.62	1	0.5
chb01_21	0.08	-0.61	-0.17	-0.64	0.5	1

Experiment 2 uses optimized settings derived from experiment 1 and studies the impact of optimized settings on other seizure files for the same patient. It predicts the seizure onset for other intra-patient seizure files. The experiment tests chb01_04, chb01_15, chb01_16, chb01_18 and chb01_21 data with training settings from chb01_03. Penalizing the rankings did not show any cluster pattern among settings. However, rankings without penalty show a consistent pattern of settings.

Experiment 3 tests the optimized settings from experiment 1 on inter-patient data to determine robustness of these settings. EEG patterns usually differ from patient to patient. While this assumption did not require testing inter-patient data, it may potentially provide an analysis for patients belonging to same age group having similar seizure patterns. However, we did not test this for any other dataset apart from CHB MIT datasets. Thus, due to small number of samples the effectiveness of experiment 3 cannot be generalized for datasets outside CHB MIT dataset.

It is interesting to observe how intra-patient settings had lower latency values when compared with inter-patient data. For same patient files 8.36 sec was the average delay but for other patients it ranged between 8 to 25 seconds approximately.

The previous related research mostly deals with predicting seizure before it occurs. This research focused on predicting the onset of seizure even if it was detected after the actual onset but not before.

The data set includes the time of the actual seizure. One approach would be to claim that predicting a seizure before it occurs is a good thing. The opposite approach is assumed for this research. A data set might not contain any seizures, so predicting the occurrence needs to be viewed as a False Positive.

9 CONCLUSION

The experimental results predict seizure with minimal deviation from the actual seizure. For experiment 1, after executing multiple runs of settings, resulted in an optimized setting. The results clearly indicate that the approach is very responsive to the onset of a seizure. This setting gives the best result for patient chb01_03 and predicts seizure onset at 3001.70 seconds with a latency of 5.7 seconds. The same settings when applied to other chb01 files provided both positive and negative latency. This may be a matter of chance but for all the files the average latency was positive.

The rankings with no penalty showed a cluster of settings that percolated to the top. However, the individual rankings for all chb01 files did not show any significant correlation between them, both for no penalty and for penalized rankings. The common settings when applied to other intra-patient seizure files in experiment 2, resulted in variable latencies and performed well in most cases.

For experiment 3, patient chb01 and chb02 had similar average latency. This may be attributed to the fact that patients were of same age. The average latency for all patients showed a linear relation with the age of the patient. However, due to the limited number of samples, the results cannot be generalized until tested on multiple other patients with varying demographics. The latency for all four patient files (chb01, chb02, chb03, chb04) ranged from a few seconds to approximately 2 minutes and not more than that. The proposed approach provided viable results for all datasets used in the experiments.

10 FUTURE DIRECTIONS

Current experiments may be enhanced as follows: More patients can be compared with each other on the basis of their seizure data. Patients belonging to diverse age groups can be studied. Instead of building a model from single patient files, all patient files can be used to build models. This can make the model robust and the model itself can be applied to patients of different age groups.

The current method uses simple moving average. Other types of moving averages, such as Weighted Moving Average(WMA), Exponential Moving Average (EMA) may be explored to predict the onset of seizure. These methods can be combined with machine learners to further increase the prediction efficiency. Neural networks can also be used in combination with the existing proposed method.

For the experiments, multiple channels may be explored instead of single channel. Another possible approach would be the use of surrogate channel that combines all important aspects of multiple channel signals into a single channel.

This research can be applied to different time-series domains. It may be used to predict stocks in the financial world. Other possible use includes predicting trends in other time-series analysis such as weather forecasting and earthquake prediction. The approach may produce different results according to the complexity of data and their applications.

REFERENCES

[1] U. R. Acharya, S. V. Sree, P. C. A. Ang, R. Yanti, and J. S. Suri, "Application of non-linear and wavelet based features for the automated identification of epileptic EEG signals," *International journal of neural systems,* vol. 22, p. 1250002, 2012.

[2] P. Aguiar, A. David, S. Paulo, and A. Rosa, "EEGSOLVER—brain activity and genetic algorithms," in *Proceedings of the 2000 ACM symposium on Applied computing-Volume 1*, 2000, pp. 80-84.

[3] H. Akil, M. E. Martone, and D. C. Van Essen, "Challenges and opportunities in mining neuroscience data," *science,* vol. 331, pp. 708-712, 2011.

[4] A. S. Al-Fahoum and A. A. Al-Fraihat, "Methods of EEG signal features extraction using linear analysis in frequency and time-frequency domains," *ISRN neuroscience,* vol. 2014, 2014.

[5] T. Allison and D. V. Cicchetti, "Sleep in mammals: ecological and constitutional correlates," *Science,* vol. 194, pp. 732-734, 1976.

[6] O. AlZoubi, I. Koprinska, and R. A. Calvo, "Classification of brain-computer interface data," in *Proceedings of the 7th Australasian Data Mining Conference-Volume 87*, 2008, pp. 123-131.

[7] E. B. Assi, S. Rihana, and M. Sawan, "Kmeans-ICA based automatic method for ocular artifacts removal in a motorimagery classification," in *Engineering in Medicine and Biology Society (EMBC), 2014 36th Annual International Conference of the IEEE*, 2014, pp. 6655-6658.

[8] T. Ayhan and S. Şeker, "Determination of decision moments in EEG signals using DWT," in *Innovations in Intelligent Systems and Applications (INISTA), 2011 International Symposium on*, 2011, pp. 268-271.

[9] W. A. W. Azlan and Y. F. Low, "Feature extraction of electroencephalogram (eeg) signal-a review," in *Biomedical Engineering and Sciences (IECBES), 2014 IEEE Conference on*, 2014, pp. 801-806.

[10] T. Bassani and J. C. Nievola, "Pattern recognition for brain-computer interface on disabled subjects using a wavelet transformation," in *Computational Intelligence in Bioinformatics and Computational Biology, 2008. CIBCB'08. IEEE Symposium on*, 2008, pp. 180-186.

[11] M. Bharati and M. Ramageri, "Data mining techniques and applications," 2010.

[12] P. Białas and P. Milanowski, "A high frequency steady-state visually evoked potential based brain computer interface using consumer-grade EEG headset," in *Engineering in Medicine and Biology Society (EMBC), 2014 36th Annual International Conference of the IEEE*, 2014, pp. 5442-5445.

[13] N. Bigdely-Shamlo, K. Kreutz-Delgado, C. Kothe, and S. Makeig, "EyeCatch: Data-mining over half a million EEG independent components to construct a fully-automated eye-component detector," in *Engineering in Medicine and Biology Society (EMBC), 2013 35th Annual International Conference of the IEEE*, 2013, pp. 5845-5848.

[14] A. Bousbia-Salah, M. Ait-Ameur, and M. Talha-Kedir, "Biorthogonal wavelet for EEG signal compression," in *Proceedings of the 4th International Symposium on Applied Sciences in Biomedical and Communication Technologies*, 2011, p. 9.

[15] B. R. Cahn and J. Polich, "Meditation states and traits: EEG, ERP, and neuroimaging studies," *Psychological bulletin,* vol. 132, p. 180, 2006.

[16] A. Chan, C. E. Early, S. Subedi, Y. Li, and H. Lin, "Systematic analysis of machine learning algorithms on EEG data for brain state intelligence," in *Bioinformatics and Biomedicine (BIBM), 2015 IEEE International Conference on*, 2015, pp. 793-799.

[17] G. Chandrashekar and F. Sahin, "A survey on feature selection methods," *Computers & Electrical Engineering,* vol. 40, pp. 16-28, 2014.

[18] W. A. Chaovalitwongse, Y.-J. Fan, and R. C. Sachdeo, "On the time series $ k $-nearest neighbor classification of abnormal brain activity," *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans,* vol. 37, pp. 1005-1016, 2007.

[19] W. A. Chaovalitwongse, O. A. Prokopyev, and P. M. Pardalos, "Electroencephalogram (EEG) time series classification: Applications in epilepsy," *Annals of Operations Research,* vol. 148, no. 1, pp. 227-250, 2006.

[20] A. Chatchinarat, K. W. Wong, and C. C. Fung, "A comparison study on the relationship between the selection of EEG electrode channels and frequency bands used in classification for emotion recognition," in *Machine Learning and Cybernetics (ICMLC), 2016 International Conference on,* 2016, pp. 251-256.

[21] K. Chen, Q. Liu, and Q. Ai, "Research of classification methods of EEG signal based on wavelet packet transform and CSP," 2012.

[22] M. Dalponte, F. Bovolo, and L. Bruzzone, "Automatic selection of frequency and time intervals for classification of EEG signals," *Electronics Letters,* vol. 43, pp. 1406-1408, 2007.

[23] S. Deedwaniya and T. K. Gandhi, "An ensemble approach for brain computer interface applications," in *Humanitarian Technology Conference (R10-HTC), 2016 IEEE Region 10*, 2016, pp. 1-4.

[24] R. Ebrahimpour, K. Babakhan, S. A. A. A. Arani, and S. Masoudnia, "Epileptic seizure detection using a neural network ensemble method and wavelet transform," *Neural Network World,* vol. 22, no. 3, p. 291, 2012.

[25] A. S. Elsawy, S. Eldawlatly, M. Taher, and G. M. Aly, "A principal component analysis ensemble classifier for P300 speller applications," in *Image and Signal Processing and Analysis (ISPA), 2013 8th International Symposium on*, 2013, pp. 444-449.

[26] P. Esling and C. Agon, "Time-series data mining," *ACM Computing Surveys (CSUR),* vol. 45, p. 12, 2012.

[27] N. Evans, "Single trial analyses of novel cognitive tasks in an EEG-based Brain-Computer Interface framework," 2008.

[28] T. P. Exarchos, A. T. Tzallas, D. I. Fotiadis, S. Konitsiotis, and S. Giannopoulos, "A data mining based approach for the EEG transient event detection and classification," in *Computer-Based Medical Systems, 2005. Proceedings. 18th IEEE Symposium on*, 2005, pp. 35-40.

[29] U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth, "From data mining to knowledge discovery in databases," *AI magazine,* vol. 17, p. 37, 1996.

[30] H. Firpi, E. Goodman, and J. Echauz, "Epileptic seizure detection by means of genetically programmed artificial features," in *Proceedings of the 7th annual conference on Genetic and evolutionary computation*, 2005, pp. 461-466.

[31] A. Flexer, "Data mining and electroencephalography," *Statistical Methods in Medical Research,* vol. 9, pp. 395-413, 2000.

[32] J. Gao, H. Sultan, J. Hu, and W.-W. Tung, "Denoising nonlinear time series by adaptive filtering and wavelet shrinkage: a comparison," *IEEE signal processing letters,* vol. 17, pp. 237-240, 2010.

[33] D. Garrett, D. A. Peterson, C. W. Anderson, and M. H. Thaut, "Comparison of linear, nonlinear, and feature selection methods for EEG signal classification,"

IEEE Transactions on neural systems and rehabilitation engineering, vol. 11, pp. 141-144, 2003.

[34] L. Geng and P. Cui, "Research of feature extraction algorithm based on steady-state visual evoked potential," in *Control and Decision Conference (2014 CCDC), The 26th Chinese*, 2014, pp. 2979-2983.

[35] P. Georgiev, F. Theis, A. Cichocki, and H. Bakardjian, "Sparse component analysis: a new tool for data mining," *Data mining in biomedicine,* pp. 91-116, 2007.

[36] A. S. Gevins, C. L. Yeager, and S. L. Diamond, "Heuristic real time feature extraction of the electroencephalogram (EEG)," in *Proceedings of the 1974 annual ACM conference-Volume 2*, 1974, pp. 491-497.

[37] T. Guan, X. Zeng, L. Huang, and M. Seok, "Neural network based seizure detection system using raw EEG data," in *SoC Design Conference (ISOCC), 2016 International*, 2016, pp. 211-212.

[38] M. Hasan, M. Ahamed, M. Ahmad, and M. Rashid, "Prediction of Epileptic Seizure by Analysing Time Series EEG Signal Using-NN Classifier," *Applied bionics and biomechanics,* vol. 2017, 2017.

[39] R. Horlings, D. Datcu, and L. J. Rothkrantz, "Emotion recognition using brain activity," in *Proceedings of the 9th international conference on computer systems and technologies and workshop for PhD students in computing*, 2008, p. 6.

[40] S. M. Hosni, M. E. Gadallah, S. F. Bahgat, and M. S. AbdelWahab, "Classification of EEG signals using different feature extraction techniques for mental-task BCI," in *Computer Engineering & Systems, 2007. ICCES'07. International Conference on*, 2007, pp. 220-226.

[41] B. Hu, X. Li, S. Sun, and M. Ratcliffe, "Attention Recognition in EEG-Based Affective Learning Research Using CFS+ KNN Algorithm," *IEEE/ACM transactions on computational biology and bioinformatics,* 2016.

[42] R. Hussein, A. Mohamed, K. Shahan, and A. A. Mohamed, "EEG feature extraction and selection techniques for epileptic detection: A comparative study," in *Computers & Informatics (ISCI), 2013 IEEE Symposium on*, 2013, pp. 170-175.

[43] M. Islam, M. R. Fraz, Z. Zahid, and M. Arif, "Optimizing common spatial pattern and feature extraction algorithm for brain computer interface," in *Emerging Technologies (ICET), 2011 7th International Conference on*, 2011, pp. 1-6.

[44] P. Jahankhani, V. Kodogiannis, and K. Revett, "EEG signal classification using wavelet feature extraction and neural networks," in *Modern Computing, 2006. JVA'06. IEEE John Vincent Atanasoff 2006 International Symposium on*, 2006, pp. 120-124.

[45] P. Jahankhani, K. Revett, and V. Kodogiannis, "Data mining an EEG dataset with an emphasis on dimensionality reduction," in *Computational Intelligence and Data Mining, 2007. CIDM 2007. IEEE Symposium on*, 2007, pp. 405-412.

[46] S. Jirayucharoensak, P. Israsena, S. Pan-ngum, and S. Hemrungrojn, "Online EEG artifact suppression for neurofeedback training systems," in *Biomedical Engineering International Conference (BMEiCON), 2013 6th*, 2013, pp. 1-5.

[47] C. A. Joyce, I. F. Gorodnitsky, and M. Kutas, "Automatic removal of eye movement and blink artifacts from EEG data using blind component separation," *Psychophysiology*, vol. 41, pp. 313-325, 2004.

[48] N. Jrad, R. Phlypo, and M. Congedo, "SVM feature selection for multidimensional EEG data," in *Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on*, 2011, pp. 781-784.

[49] M. Junghöfer, T. Elbert, D. M. Tucker, and B. Rockstroh, "Statistical control of artifacts in dense array EEG/MEG studies," *Psychophysiology*, vol. 37, pp. 523-532, 2000.

[50] B. Kemp and J. Olivan, "European data format 'plus'(EDF+), an EDF alike standard format for the exchange of physiological data," *Clinical Neurophysiology*, vol. 114, no. 9, pp. 1755-1761, 2003.

[51] S. KG, K. Venugopal, and L. Patnaik, "Feature extraction using fuzzy c-means clustering for data mining systems," *IJCSNS*, vol. 6, p. 230, 2006.

[52] C.-W. Ko, Y.-D. Lin, H.-W. Chung, and G.-J. Jan, "An EEG spike detection algorithm using artificial neural network with multi-channel correlation," in *Engineering in Medicine and Biology Society, 1998. Proceedings of the 20th Annual International Conference of the IEEE*, 1998, pp. 2070-2073.

[53] X. Kong and P. S. Yu, "Brain network analysis: a data mining perspective," *ACM SIGKDD Explorations Newsletter*, vol. 15, pp. 30-38, 2014.

[54] E. Kroupi, A. Yazdani, J.-M. Vesin, and T. Ebrahimi, "EEG correlates of pleasant and unpleasant odor perception," *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, vol. 11, p. 13, 2014.

[55] S. S. Labib, "A novel brain computer interface based on Principle Component Analysis and Fuzzy Logic," in *Digital Information Processing and Communications (ICDIPC), 2016 Sixth International Conference on*, 2016, pp. 31-36.

[56] S. Laxman and P. S. Sastry, "A survey of temporal data mining," *Sadhana,* vol. 31, pp. 173-198, 2006.

[57] J.-H. Lee, S. Oh, F. A. Jolesz, H. Park, and S.-S. Yoo, "Application of Independent Component Analysis for the Data Mining of Simultaneous Eeg–fMRI: Preliminary Experience on Sleep Onset," *International Journal of Neuroscience,* vol. 119, pp. 1118-1136, 2009.

[58] S. H. Lee, B. Abibullaev, W.-S. Kang, Y. Shin, and J. An, "Analysis of attention deficit hyperactivity disorder in EEG using wavelet transform and self organizing maps," in *Control Automation and Systems (ICCAS), 2010 International Conference on*, 2010, pp. 2439-2442.

[59] L. Lhotská, V. Krajca, J. Mohylová, S. Petránek, and V. Gerla, "EEG Data Mining Using PCA," in *Data Mining and Medical Knowledge Management: Cases and Applications*, ed: IGI Global, 2009, pp. 161-180.

[60] Y.-P. Lin, C.-H. Wang, T.-P. Jung, T.-L. Wu, S.-K. Jeng, J.-R. Duann, *et al.*, "EEG-based emotion recognition in music listening," *IEEE Transactions on Biomedical Engineering,* vol. 57, pp. 1798-1806, 2010.

[61] K. Markopoulos, C. Mavrokefalidis, K. Berberidis, and E. Daskalopoulou, "BCI-based approaches for real-time applications," in *Proceedings of the 20th Pan-Hellenic Conference on Informatics*, 2016, p. 25.

[62] P. E. McSharry, L. A. Smith, and L. Tarassenko, "Comparison of predictability of epileptic seizures by a linear and a nonlinear method," *IEEE transactions on Biomedical Engineering,* vol. 50, no. 5, pp. 628-633, 2003.

[63] F. Mörchen, "Time series feature extraction for data mining using DWT and DFT," ed: Univ., 2003.

[64] B. I. Morshed and A. Khan, "Journal of Bioengineering & Biomedical Science," 2014.

[65] K.-R. Muller, J. Kohlmorgen, A. Ziehe, and B. Blankertz, "Decomposition algorithms for analysing brain signals," in *Adaptive Systems for Signal Processing, Communications, and Control Symposium 2000. AS-SPCC. The IEEE 2000*, 2000, pp. 105-110.

[66] A. K. Ngugi, C. Bottomley, I. Kleinschmidt, J. W. Sander, and C. R. Newton, "Estimation of the burden of active and life-time epilepsy: a meta-analytic approach," *Epilepsia,* vol. 51, no. 5, pp. 883-890, 2010.

[67] H. Nolan, R. Whelan, and R. Reilly, "FASTER: fully automated statistical thresholding for EEG artifact rejection," *Journal of neuroscience methods,* vol. 192, pp. 152-162, 2010.

[68] E. Nurse, B. S. Mashford, A. J. Yepes, I. Kiral-Kornek, S. Harrer, and D. R. Freestone, "Decoding EEG and LFP signals using deep learning: heading TrueNorth," in *Proceedings of the ACM International Conference on Computing Frontiers*, 2016, pp. 259-266.

[69] U. Orhan, M. Hekim, and M. Ozer, "EEG signals classification using the K-means clustering and a multilayer perceptron neural network model," *Expert Systems with Applications,* vol. 38, pp. 13475-13481, 2011.

[70] Y. Pan, C. Guan, J. Yu, K. K. Ang, and T. E. Chan, "Common frequency pattern for music preference identification using frontal EEG," in *Neural Engineering (NER), 2013 6th International IEEE/EMBS Conference on*, 2013, pp. 505-508.

[71] A. Panuccio, M. Bicego, and V. Murino, "A Hidden Markov Model-based approach to sequential data clustering," *Structural, Syntactic, and Statistical Pattern Recognition,* pp. 734-743, 2002.

[72] H. Peng, B. Hu, F. Zheng, D. Fan, W. Zhao, X. Chen, *et al.*, "A method of identifying chronic stress by EEG," *Personal and ubiquitous computing,* vol. 17, pp. 1341-1347, 2013.

[73] P. C. Petrantonakis and L. J. Hadjileontiadis, "EEG-based emotion recognition using hybrid filtering and higher order crossings," in *Affective Computing and Intelligent Interaction and Workshops, 2009. ACII 2009. 3rd International Conference on*, 2009, pp. 1-6.

[74] R. D. Pinto and H. A. Ferreira, "Development of a Non-invasive Brain Computer Interface for Neurorehabilitation," in *Proceedings of the 3rd 2015 Workshop on ICTs for improving Patients Rehabilitation Research Techniques*, 2015, pp. 126-130.

[75] R. Polikar, "The wavelet tutorial," 1996.

[76] M. Poulos, M. Rangoussi, and E. Kafetzopoulos, "Person identification via the EEG using computational geometry algorithms," in *Signal Processing Conference (EUSIPCO 1998), 9th European*, 1998, pp. 1-4.

[77] M. Pregenzer and G. Pfurtscheller, "Frequency component selection for an EEG-based brain to computer interface," *IEEE Transactions on Rehabilitation Engineering*, vol. 7, pp. 413-419, 1999.

[78] K. C. Pullaiah and T. Pitchaiah, "A low powr high speed implementation of ECG feature extraction using hybrid algorithm," in *Inventive Computation Technologies (ICICT), International Conference on*, 2016, pp. 1-5.

[79] P. D. Purnamasari, A. A. P. Ratna, and B. Kusumoputro, "Artificial Neural Networks Based Emotion Classification System through Relative Wavelet Energy of EEG Signal," in *Proceedings of the Fifth International Conference on Network, Communication and Computing*, 2016, pp. 135-139.

[80] T. Rakthanmanon, B. Campana, A. Mueen, G. Batista, B. Westover, Q. Zhu, *et al.*, "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*, 2012, pp. 262-270.

[81] P. RamaRaju, N. AnogjnaAurora, and V. M. Rao, "Relevance of wavelet transform for taxonomy of EEG signals," in *Electronics Computer Technology (ICECT), 2011 3rd International Conference on*, 2011, pp. 466-470.

[82] J. Rasekhi, M. R. K. Mollaei, M. Bandarabadi, C. A. Teixeira, and A. Dourado, "Preprocessing effects of 22 linear univariate features on the performance of seizure prediction methods," *Journal of neuroscience methods*, vol. 217, no. 1-2, pp. 9-16, 2013.

[83] I. Rejer and K. Lorenz, "Genetic algorithm and forward method for feature selection in EEG feature space," *Journal of Theoretical and Applied Computer Science*, vol. 7, pp. 72-82, 2013.

[84] D. Rivero, E. Fernandez-Blanco, J. Dorado, and A. Pazos, "Classification of two-channel signals by means of genetic programming," in *Proceedings of the Companion Publication of the 2015 Annual Conference on Genetic and Evolutionary Computation*, 2015, pp. 1319-1325.

[85] V. Robinson and E. B. Varghese, "A novel approach for ensuring the privacy of EEG signals using application-specific feature extraction and AES algorithm," in *Inventive Computation Technologies (ICICT), International Conference on*, 2016, pp. 1-6.

[86] V. Sakkalis *et al.*, "Absence seizure epilepsy detection using linear and nonlinear EEG analysis methods," in *Engineering in Medicine and Biology Society (EMBC),*

2013 35th Annual International Conference of the IEEE, 2013, pp. 6333-6336: IEEE.

[87] R. Santana, L. Bonnet, J. Legény, and A. Lécuyer, "Introducing the use of model-based evolutionary algorithms for EEG-based motor imagery classification," in *Proceedings of the 14th annual conference on Genetic and evolutionary computation*, 2012, pp. 1159-1166.

[88] A. Schlögl, C. Keinrath, D. Zimmermann, R. Scherer, R. Leeb, and G. Pfurtscheller, "A fully automated correction method of EOG artifacts in EEG recordings," *Clinical neurophysiology*, vol. 118, pp. 98-104, 2007.

[89] M. Schroder, M. Bogdan, T. Hinterberger, and N. Birbaumer, "Automated EEG feature selection for brain computer interfaces," in *Neural Engineering, 2003. Conference Proceedings. First International IEEE EMBS Conference on*, 2003, pp. 626-629.

[90] T. Schuster, S. Gruss, H. Kessler, A. Scheck, H. Hoffmann, and H. Traue, "EEG: pattern classification during emotional picture processing," in *Proceedings of the 3rd international conference on pervasive technologies related to assistive environments*, 2010, p. 67.

[91] B. Şen, M. Peker, A. Çavuşoğlu, and F. V. Çelebi, "A comparative study on classification of sleep stage based on EEG signals using feature selection and classification algorithms," *Journal of medical systems*, vol. 38, p. 18, 2014.

[92] A. S. Shirazi, M. Hassib, N. Henze, A. Schmidt, and K. Kunze, "What's on your mind?: mental task awareness using single electrode brain computer interfaces," in *Proceedings of the 5th Augmented Human International Conference*, 2014, p. 45.

[93] A. H. Shoeb and J. V. Guttag, "Application of machine learning to epileptic seizure detection," in *Proceedings of the 27th International Conference on Machine Learning (ICML-10)*, 2010, pp. 975-982.

[94] D. R. Simkin, R. W. Thatcher, and J. Lubar, "Quantitative EEG and neurofeedback in children and adolescents: anxiety disorders, depressive disorders, comorbid addiction and attention-deficit/hyperactivity disorder, and brain injury," *Child and adolescent psychiatric clinics of North America*, vol. 23, pp. 427-464, 2014.

[95] K. Su and K. A. Robbins, "Subject specific parameter selection for the EEG classifier using common spatial patterns," in *Proceedings of the 2012 ACM Research in Applied Computation Symposium*, 2012, pp. 68-69.

[96] K. Su and K. A. Robbins, "Content-based EEG database retrieval using a multiclass SVM classifier," in *Global Conference on Signal and Information Processing (GlobalSIP), 2013 IEEE*, 2013, pp. 17-20.

[97] A. Subasi and M. I. Gursoy, "EEG signal classification using PCA, ICA, LDA and support vector machines," *Expert Systems with Applications*, vol. 37, pp. 8659-8666, 2010.

[98] H.-I. Suk and S.-W. Lee, "Data-driven frequency bands selection in EEG-based brain-computer interface," in *Pattern Recognition in NeuroImaging (PRNI), 2011 International Workshop on*, 2011, pp. 25-28.

[99] A. Swetapadma and B. R. Swain, "A data mining approach for sleep wave and sleep stage classification," in *Inventive Computation Technologies (ICICT), International Conference on*, 2016, pp. 1-6.

[100] A. Takajyo, M. Katayama, K. Inoue, K. Kumamaru, and S. Matsuoka, "Time-frequency analysis of human sleep EEG," in *SICE-ICASE, 2006. International Joint Conference*, 2006, pp. 3303-3307.

[101] L. Tarassenko, Y. Khan, and M. Holt, "Identification of inter-ictal spikes in the EEG using neural network analysis," *IEE Proceedings-Science, Measurement and Technology*, vol. 145, no. 6, pp. 270-278, 1998.

[102] M. Teplan, "Fundamentals of EEG measurement," *Measurement science review*, vol. 2, pp. 1-11, 2002.

[103] T. N. Thieu, H.-J. Yang, S.-H. Kim, and A. Oh, "Fast Fourier Transform and Tucker Decomposition for Feature Extraction in EEG Signals," in *Proceedings of the 2015 International Conference on Big Data Applications and Services*, 2015, pp. 78-83.

[104] W. Ting, Y. Guo-zheng, Y. Bang-hua, and S. Hong, "EEG feature extraction based on wavelet packet decomposition for brain computer interface," *Measurement*, vol. 41, pp. 618-625, 2008.

[105] A. T. Tzallas, M. G. Tsipouras, and D. I. Fotiadis, "Epileptic seizure detection in EEGs using time–frequency analysis," *IEEE transactions on information technology in biomedicine*, vol. 13, no. 5, pp. 703-710, 2009.

[106] A. T. Tzallas *et al.*, "Automated epileptic seizure detection methods: a review study," in *Epilepsy-histological, electroencephalographic and psychological aspects*: InTech, 2012.[107] P. Valenti, E. Cazamajou, M. Scarpettini, A. Aizemberg, W. Silva, and S. Kochen, "Automatic detection of interictal spikes

using data mining models," *Journal of neuroscience methods,* vol. 150, pp. 105-110, 2006.

[108] W. N. Venables, D. M. Smith, and R. D. C. Team, "An introduction to R," ed: Network Theory Limited, 2004. [96] R. Vigário, J. Sarela, V. Jousmiki, M. Hamalainen, and E. Oja, "Independent component approach to the analysis of EEG and MEG recordings," *IEEE transactions on biomedical engineering,* vol. 47, pp. 589-593, 2000.

[109] B. Wang, L. Jun, J. Bai, L. Peng, G. Li, and Y. Li, "EEG recognition based on multiple types of information by using wavelet packet transform and neural networks," in *Engineering in Medicine and Biology Society, 2005. IEEE-EMBS 2005. 27th Annual International Conference of the*, 2006, pp. 5377-5380.

[110] D. Wang, D. Miao, and C. Xie, "Best basis-based wavelet packet entropy feature extraction and hierarchical EEG classification for epileptic detection," *Expert Systems with Applications,* vol. 38, pp. 14314-14320, 2011.

[111] J. Wang, G. Xu, L. Wang, and H. Zhang, "Feature extraction of brain-computer interface based on improved multivariate adaptive autoregressive models," in *Biomedical engineering and informatics (bmei), 2010 3rd international conference on*, 2010, pp. 895-898.

[112] S. Wang, C.-J. Lin, C. Wu, and W. A. Chaovalitwongse, "Early detection of numerical typing errors using data mining techniques," *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans,* vol. 41, pp. 1199-1212, 2011.

[113] G. Wiechert, M. Triff, Z. Liu, Z. Yin, S. Zhao, Z. Zhong, *et al.*, "Identifying users and activities with cognitive signal processing from a wearable headband," in *Cognitive Informatics & Cognitive Computing (ICCI* CC), 2016 IEEE 15th International Conference on*, 2016, pp. 129-136.

[114] W. Xu, C. Guan, C. E. Siong, S. Ranganatha, M. Thulasidas, and J. Wu, "High accuracy classification of EEG signal," in *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on*, 2004, pp. 391-394.

[115] J.-Z. Xue, H. Zhang, C.-X. Zheng, and X.-G. Yan, "Wavelet packet transform for feature extraction of EEG during mental tasks," in *Machine learning and cybernetics, 2003 international conference on*, 2003, pp. 360-363.

[116] C. Yaacoub, G. Mhanna, and S. Rihana, "A Genetic-Based Feature Selection Approach in the Identification of Left/Right Hand Motor Imagery for a Brain-Computer Interface," *Brain sciences,* vol. 7, p. 12, 2017.

[117] B. Yahya-Zoubir, M. Bentlemsan, E.-T. Zemouri, and K. Ferroudji, "Adaptive Time Window for EEG-based Motor Imagery Classification," in *Proceedings of the International Conference on Intelligent Information Processing, Security and Advanced Communication*, 2015, p. 83.

[118] C. N. Yalung, S. Al Majeed, and J. Karam, "Analysis and Interpretation of Brain Wave Signals," in *Proceedings of the International Conference on Internet of things and Cloud Computing*, 2016, p. 67.

[119] T. Yamagutchi, K. Nagata, and P. Q. Truong, "Pattern recognition of EEG signal during motor imagery by using SOM," in *Innovative Computing, Information and Control, 2007. ICICIC'07. Second International Conference on*, 2007, pp. 121-121.