國立臺灣大學電機資訊學院生醫電子與資訊學研究所

碩士論文

Graduate Institute of Biomedical Electronics and Bioinformatics

College of Electrical Engineering and Computer Science

National Taiwan University

Master Thesis

多通道腦波特徵抽取及分析之癲癇預測系統

Epileptic Seizure Detection System Using Multi-Channel EEG as Basis for Classification

劉時廷

Shih-Ting Liu

指導教授：賴飛羆教授

Advisor: Prof. Feipei Lai

中華民國 101 年 6 月

June 2012

誌謝

本論文能夠順利完成，是因為有許多人在背後的支持、鼓勵、與指導。首先，我要特別感謝我的指導教授賴飛羆老師，在課業上老師給予我技術及知識上的指導，在生活上老師教導我們做人處事的方法，同時也擴展了我的視野，讓我學會從不同角度看事情，用創新的方法解決問題。

特別感謝沈家平學長在研究過程中給予相關技術指導，在我研究遇到瓶頸時給予我方向。也特別感謝邱銘章醫師常常忙中抽空教導我們腦波專業知識使得本論文能順利完成。

兩年的研究生活充滿了許多快樂的經驗，不論是學術上或是生活上，都要感謝實驗室的夥伴們還莒學長、憲政學長、凱平學長、啟煌學長、鎮宇學長、偉昕學長、意儒學姐、經閔學長、峰生學長、麒文學長、浩準學長、逸帆學長、釭昇學長、亞霖學長、沅庭學長、景崴學姐、煌仁同學、冠仲同學、偉志同學、子華同學、凱傑同學、智閔同學、及恩羽、曉亞、文斌、翰平、相茹等學妹學弟們。

最後，我還要感謝我的父母，在我挫折和失意的時候不斷的給我鼓勵，以及給予我許多援助和資源，讓我在求學的路上不虞匱乏和不必擔心生活上的需求，讓我能專心地做研究，讓我無時無刻心中有個心靈支柱可以支持我。

中文摘要

癲癇是一種常見的慢性神經疾病，並且會有不定時的發作情形。顛顯發作時病人會短暫失去肢體控制並導致生命危險。目前有關癲癇之研究及診斷多數利用腦波圖(Electroencephalogram)。腦波圖可以用不同的顯示方法被呈現，其中兩種為單極點訊號 (Unipolar)和雙極點訊號 (Bipolar)。傳統腦波訊號分析大多利用單極點訊號作為基礎，但醫師在診斷顛癇時時常利用雙極點訊號來呈現腦波圖。因此我們也把雙極點訊號拿來作為辨識系統之參考數據。我們設計了一列對於雙極點訊號之訊號處理及特徵抽取方法希望能夠改善目前現有之自動化癲癇診斷系統。在訊號處理方面我們利用了小波轉換(Wavelet Transform)將主要不同腦波頻帶抽取出來。在特徵抽取上我們利用似熵 (Approximate entropy)及種總變差(Total variation)來顯示出規則與不規則之腦波現象。在特徵排序及選擇我們採用了基因演算法(Genetic Algorithm)和費雪分數法 (Fisher Score)。最後再利用支持向量機(Support Vector Machine)來當我們的分類器。

關鍵詞: 小波轉換、心電圖、支持向量機

ABSTRACT

Epilepsy is a common chronic neurological disorder characterized by recurrent unprovoked seizures. Seizure episodes can cause temporal paralysis of the body, which can lead to severe injuries. Electroencephalogram (EEG) is a tool commonly used for analyzing brain activity and diagnosing brain disorders. EEG can be presented under different montage schemes. This study focuses on two of the montage schemes; unipolar montage and bipolar montage. Traditionally, the most commonly used montage for automated EEG analysis is unipolar. We experiment with incorporating bipolar EEG montage for creating a classification system to classify different epileptic wave forms. A series of functions were designed for bipolar EEG montage. We used wavelet transform (WT) to decompose EEG signal into its primary sub-bands. We use Approximate Entropy and Total Variation as features designed specifically for spike and seizure detection. We used Genetic Algorithm and Fisher Score to rank and selected most influential features for classifier. Finally we use multi-class Support Vector Machine as our classifier.

Keywords: Genetic Algorithm, Fisher Score, Support Vector Machines

CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter 1 Introduction

Epilepsy is a chronic neurological disorder characterized by recurrent unprovoked seizures. The disorder affects almost 60 million people around the world [1]. Prevalence rates in the US and Europe are 0.52% and 0.68% respectively [2]. In developing countries the prevalence rates go as high as 1.5%. Epilepsy is the third most common neurological disorder after Alzheimer's disease and stroke [3]. In Taiwan, there are 192,369 people with epilepsy [4], which is about 0.8% of Taiwan's population. Seizure happens when clusters of brain neurons signal abnormally, which may temporarily cause anomaly in a person's consciousness, behavior, movements, and actions. Medicinal and surgical remedies for seizure prevention are available. However, in 25% of patients neither medication nor surgery completely controls epileptic seizure occurrence [5]. One way to diagnose and analyze epilepsy is through careful studies of patient's electroencephalogram (EEG.) EEG measures voltage changes from ionic current flows within the neurons of the brain [6], shows temporal and spatial information of the brain, and is useful in the diagnosis of epilepsy. Studies have shown that EEG provides high sensitivity and specificity for the diagnosis of epilepsy. For example, the 3-Hz spike-and-wave in EEG is unique to petit mal and a particular form of absence seizure of childhood onset. As a result, several studies 0-[14] have proposed forewarning algorithms for epileptic seizures using EEGs. If successful, it would improve the quality of life and safety for patients with epilepsy. Thus, real-time EEG analysis system is and will be a vital part of seizure forewarning system.

Various approaches for providing automatic seizure detection have been proposed. Weng and Khorasani [7] used amplitude, duration, and coefficients of variation and

frequency as inputs for a neural network. Güler and Übeyli [8]-[10] proposed a method for seizure detection based on wavelet coefficients, eigenvectors, and a support vector machine (SVM). A system proposed by Srinivasan et al. [11] utilizes approximate entropy as the feature classification for seizure detection. Recently, Adeli et al. [12] performed a principal component analysis on enhanced cosine radio basis function neural network (RBFNN) to detect seizures, while Tzallas et al. [13] demonstrated the suitability of time–frequency analysis to classify EEG segments for epileptic seizures. Some experts have also used specific shapes of interictal spikes as features in the hippocampus in terms of cell- and network-related parameters of neuronal circuits [14]. Although promising results have been reported in these studies, the accuracy of recognition is still low, and the performance of such systems needs to be improved. Furthermore, most of these studies tested the performance on open source data and lack validation with clinical data.

Multichannel EEG provides higher spatial resolution of brain activity than single channel EEG [15]. More often than not single channel EEG is not enough for extracting meaningful features for recognition of epilepsy. However, due to large computational and storage requirement, especially in the case of long-term EEG monitoring, many studies only focus on processing single channel EEG signals. Long term EEG records can last from 24 hours to several days, amounting large bodies of data. Therefore, features must be carefully selected in order to reduce the size of data for faster and more efficient classification. Also, considering physiological aspects of various phenomena of epilepsy, it is necessary to use multichannel EEG to extract physiologically meaningful features. Only through multichannel EEG can researchers get a fully understanding of different states of different parts of the brain. The aim of the research is to develop an

analytic algorithm for multichannel EEG signals to facilitate the detection of epileptic seizure warning signs. For this thesis we have indeed designed a preliminary system with promising results for the detection of various epileptic phenomena. The rest of the thesis will elaborate more on our work; some background knowledge related to this research is listed in Chapter 2 and details of this analytic system are discussed in detail in Chapter 3. Performance and experiment results of the proposed system are discussed in Chapter 4. Conclusion and discussion are discussed in Chapter 5.

Chapter 2 Background

2.1 Biosignal

2.1.1 Electroencephalogram (EEG)

Biological beings emit signals of different kinds pertaining to different physiological meanings. The term biosignal refers to all kinds of signals that can be measured and monitored from living organisms. Biosignals can be either non-electrical or electrical, but most commonly the latter, which are called bio-electrical signals. Electrical biosignals are usually taken to be electric currents produced by the sum of electrical potential differences across a specialized tissue, organ or cell system like the nervous system. Some best-known bio-electrical signals are the electroencephalogram (EEG), Magnetoencephalogram (MEG), Galvanic skin response (GSR), Electrocardiogram (ECG), Electromyogram (EMG), and heart rate variability (HRV). Among these bio-electrical signals, EEG is most commonly used to study activities and anomalies of the brain. In analyzing epileptic-related phenomenon, we focus on processing the EEG. The following section briefly introduces EEG and how it is used in analyzing brain-related disorders.

EEG is the recording of the electrical activity along the scalp produced by the firing of neurons within the brain. The voltage amplitudes are small, typically in the range of tens of microvolts. They are thought to be caused by synchronized activity in very large numbers of synapses in the cerebral cortex. In practice, EEG recordings usually last between 20 to 40 minutes, usually with multiple electrodes placed on the scalp. Typically, meaningful EEG signal frequencies fall between 1-60Hz. Activities

outside this range are likely to be artifactual. Different frequency bands (EEG bands) reflect different types of activities (Figure 1). The section below lists different EEG bands:

(1) Delta wave: has a frequency of 3 Hz or below. It tends to be the highest in amplitude and the slowest wave. It is normally the dominant rhythm in infants up to one year old and in stages 3 and 4 of sleep. It may occur focally with sub-cortical lesions and in general distribution with diffuse lesions, metabolic encephalopathy hydrocephalus, or deep midline lesions. Delta waves are usually most prominent frontally in adults and posterior in children.

(2) Theta wave: has a frequency of 3.5 to 7.5 Hz and is classified as "slow" activity. It is perfectly normal in children up to 13 years and in sleep but abnormal in adults who are awake. It can be seen as a manifestation of focal sub-cortical lesions and can also be considered in generalized distribution in diffuse disorders such as metabolic encephalopathy or some instances of hydrocephalus.

(3) Alpha wave: has a frequency between 7.5 and 13 Hz. It is usually best seen in the posterior regions of the head on each side, being higher in amplitude on the dominant side. It appears when an individual closes the eyes and relaxes, and disappears when the eyes are open or if there is an alert by any mechanism (e.g., thinking, calculating, etc.). It is the major rhythm observed in normal relaxed adults, and is present throughout most of an individual’s life, especially after the thirteenth year.

(4) Beta wave: denotes “fast” activity. It has a frequency higher than 14 Hz. It is usually found on both sides in symmetrical distribution and is most evident frontally. It is

accentuated by sedative-hypnotic drugs, especially the benzodiazepines and the barbiturates. It may be absent or reduced in areas of cortical damage. It is generally regarded as a normal rhythm. It is the dominant rhythm in patients who are alert or anxious or have their eyes open.

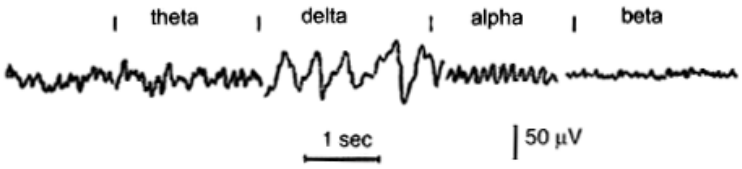

Figure 1 The various frequency bands of EEG signals

For most clinical and research applications, the placement and names of electrodes for EEG follows the International 10-20 system* (Figure 2). The purpose for using standardized scalp electrode placement guide is so that subject's test sessions could be compared and different subjects' experimental results could be compared as well. The "10" and "20" refer to the distances between adjacent electrodes to be either 10% or 20% of the total front-back or right-left distance of the skull. Each electrode site has a letter to identify lobe and a number to identify hemisphere. The letters for lobe identification are *F*, *T*, *C*, *P*, and *O*, which stand for frontal, temporal, central, parietal, and occipital lobes, respectively. Even numbers 2, 4, 6, and 8 refer to right hemispherical electrode positions. Odd numbers 1, 3, 5, and 7 refer to left hemispherical electrode positions. Between the two hemispheres are the midline electrodes marked with subscript z, which stands for zero.

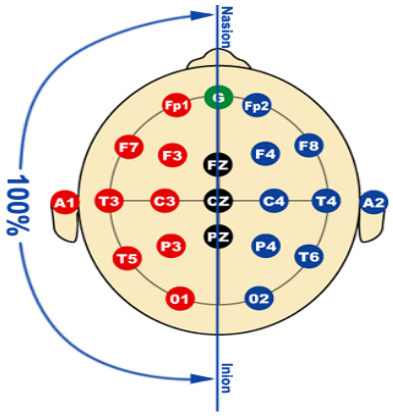

Figure 2 The scalp electrodes for a classical EEG recording

EEG signals can be displayed in one of the four ways, also known as montages: unipolar, bipolar, Laplacian, or average reference montage. In unipolar EEG montage scheme, a reference and a ground electrode are used and for each recording site an EEG amplifier electrode is placed. In Laplacian montage each channel represents the difference between an electrode and a weighted average of the surrounding electrodes. In average reference montage the outputs of all of the amplifiers are summed and averaged, and this averaged signal is used as the common reference for each channel. Bipolar EEG signal measures the action potential difference between pairs of electrodes, and can be calculated by subtracting between unipolar measurements. For n electrodes there are $n!/(r!(n\text{-}2)!)$ possibilities for electrode pairs. However, for practical purposes of this study, only neighboring electrode pairs are compared. Bipolar EEG montage showing neighboring potential differences allows us to spot locales where neurons are firing differently than neurons at other locales. This could be very useful in detecting various epileptic waveforms. One of our focuses is the usage of bipolar EEG montage for extracting useful features for analysis of epilepsy.

2.1.2 Functional Magnetic Resonance Imaging (fMRI)

Since EEG signal is derived directly from action potential of neurons, it offers very high temporal resolution. The drawback of EEG, due to superficial placement of electrodes, is poor spatial resolution. This is especially true when the locale of interest is deep inside the brain. Signals originating far from the surface where electrodes are placed decade fast and overlap with other signal origins, making it difficult to get a full understanding of brain activities. It is mathematically impossible to reconstruct a unique intracranial current source for a given EEG signal, as some currents produce potentials that cancel each other out. To complement for lack of spatial resolution, fMRI can be used synchronously alongside EEG.

In essence, fMRI measures metabolic activities of blood cells in the brain. It does not directly measure neuron activities; rather it measures the residual effects of neuron activities. The flow of blood within the brain differs according to the activity level of the individual sites; active sites use more energy and require more blood flowing through. From this knowledge, activities of the brain can be observed by getting readings on oxygen level of blood cells. We can map neural activity to changes in blood flow. fMRI scans provide 3 dimensional information of the brain activity and provides good spatial resolution that ranges from 4 to 5 mm to 1 mm per voxel*, which can give us detailed information of activities going on at different locales of the brain. Due to technology limitations and the time delay in metabolic activities, temporal resolution of fMRI is poor; often seconds pass by between scans. It is sometimes useful to use fMRI and EEG synchronously to achieve high spatial and temporal image of brain activity. However, the ultimate goal of this project is to develop a portable epilepsy seizure detection system and therefore we did not consider using fMRI as the fMRI machines

are not and most likely will not be portable in the near future. For future researches on the brain, the usage of fMRI, in combination with EEG could be considered.

2.2 Machine Learning and Classification

Classification is a common task of machine learning. The idea is to determine where new observations or instances, of any phenomena, belong. Machine learning could be either supervised or unsupervised. Supervised learning involves the usage of labeled instances for training, and as a result of training, classification models are constructed. Training sets are often paired with desired output result, which help guide training process. Classification models (also called classifiers) are inferred functions that should produce correct output from any valid input. Unsupervised learning, on the other hand, provides no labeled training guides to the learner. Unsupervised learning is often used to find hidden structures in data. Since we already know the different types of patterns we are looking for, for the classification of different epileptogenic patterns we use supervised training scheme.

Among many machine learning techniques, one of the latest learning methods is the Support Vector Machine (SVM). SVM was developed by Vladimr N. Vapnik in the 1990's. SVM is a linear learning system aimed for classification and regression. The basic idea of SVM is to transform non-linear points (data instances) to higher dimension spaces so that the non-linear points become linear in the higher dimension space. Once the points are linear, they can be separated linearly by using a hyper plane. Since only one hyper plane is used, standard SVM can only classify between two classes. For our system, we attempt to classify three different epileptogenic wave forms. Therefore we

use a modified version of SVM for classification which can classify multiple classes. Details of SVM are discussed in Chapter 3.

Chapter 3 Methods

3.1 System Architecture

Our EEG epilepsy classification system consists of four major components: data preprocessing, feature extraction, classification, and post-classification spike matching. The first 3 three components have the same system architecture as [16]. Figure 3 shows the system architecture for the EEG epileptogenic pattern classifier. Components are designed separately and modularized. Within each component further modularization takes place. This is done so that each component can be used for other researches, and also this allows for easy upgrades and changes in the future. For instance, the feature extraction module contains various sub-modules that are useful for signal processing in general; sub-modules such as wavelet transformation can be used for purposes other than of this research. Another advantage for modularized approach is so that we can replace modules easily, which could allow us to experiment the impact of different classification techniques. For example, wavelet transformation module could be replaced with Fourier transformation module for comparison between the impacts of using different signal transformation schemes. Some of the more general modules that can be used for other applications are written as libraries and can be used for other research purposes.

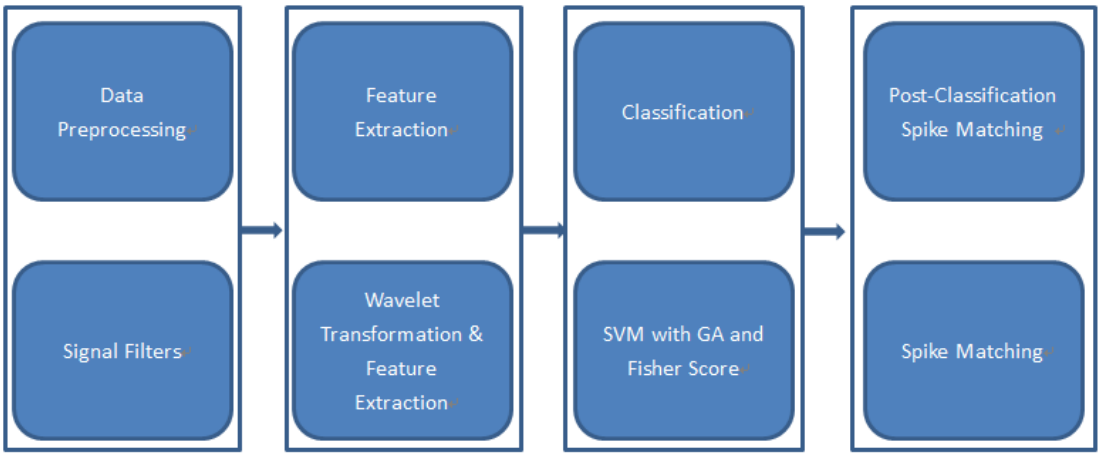

Figure 3 System Architecture

3.2 Data Acquisition

The Department of Epileptology at the University Hospital of Bonn offers five open source EEG datasets for research. The complete dataset consists of five sets (denoted A-E) each containing 100 single-channel EEG segments. These segments were selected and cut out from continuous multichannel EEG recordings after visual inspection for artifacts such as eye movements. Sets *A* and *B* consists of EEG recordings carried out on five healthy volunteers using the International 10-20 electrode placement scheme. Volunteers were in a relaxed awake-state with eyes open (A) and eyes closed (B), respectively. Sets *C*, *D*, and *E* originated from EEG archive of pre-surgical diagnosis. Segments in set *D* were recorded from within the epileptogenic zone, and those in set *C* from the hippocampal formation of the opposite hemisphere of the brain. Sets *C* and *D* contained only activity measured during seizure free intervals, set E only contained seizure activity. All EEG signals were recorded with the same 128-channel

amplifier system, using an average common reference. The EEG was acquired at a sampling rate of 173.61 Hz and a 12-bit analog-to-digital conversion.

Aside from the open source data available, we also use data collected from subjects receiving routine EEG examinations and from long-term EEG monitoring at the Department of Neurology at National Taiwan University Hospital (NTUH). The research has been approved by the ethical committee of the NTUH. The subjects were divided into two groups. EEG data from 22 participants (11 women, and 11 men) whose ages range from 23 to 86 years were included in our research. Participants are either in the control group or in the patient group. The patient group consisted of 4 women and 6 men who were diagnosed with temporal lobe epilepsy with abnormal focal or regional EEG signals. The control group consisted of 7 women and 5 men who were adults referred from outpatient clinic with normal EEG signals. These were subjects who complained about headaches or dizziness, but were not diagnosed with epilepsy or seizure disorders. The mean age for the patient group was 67.2 years and 43.58 years for the control group. The EEG signals were collected using 21 scalp Ag/AgCl electrodes placed according to the International 10-20 system. Signals were digitalized at a sampling rate of 200Hz with a dynamic range of 12 bits. The recorded EEG was classified into one of the three epileptogenic wave types, namely normal, ictal, and interictal.

The EEG was collected from 21-channel scalp Ag-AgCl electrodes according to the 10-20 International System and was digitalized at a sampling rate of 200 Hz and a dynamic range of 12 bits. The recorded EEG was classified into normal EEG, ictal and ictal discharges and was segmented into 2-second epochs. Inter-ictal epileptiform

discharges should meet the following conventional criteria [17]: (a) they must be paroxysmal; (b) they must include an abrupt change in polarity occurring over several seconds; (c) the duration of each transient should be less than 200 ms (spikes < 70 ms and sharp waves between 70 and 200 ms); (d) the discharge must have a physiology field. EEG abnormalities in patients with seizure disorders may be categorized as either specific or nonspecific patterns. The specific patterns include the spikes, sharp waves, spike-wave complexes, temporal intermittent rhythmic delta activity (TIRDA), and periodic lateralized epileptiform discharges (PLEDs), which are all potentially epileptogenic and provide diagnostically useful information [18] while the non-specific changes such as generalized or focal slow-wave activity do not. In this study we chose only those specific patterns for recognition.

The electrographic onset of a seizure is characterized by a sudden change of frequency and appearance of a new rhythm. Focal onset of the electrographic seizure may evolve through several phases: (1) focal desynchronization or attenuation of EEG activity ($\leqq 10\ \mu$ V); (2) focal, rhythmic, low voltage, fast activity ($\geqq 13$ Hz) discharges; and (3) progressive increase in amplitude with slowing that spreads to a regional anatomic distribution. Focal ictal discharges may be recorded as paroxysmal repetitive spikes, spike-waves (3 or more discharges in sequence) or rhythmic fast or theta activity [39, 40]. Since EEG experts may have different opinions on EEG classification. The EEG classifications of interictal and ictal activities were examined by two EEG experts (Chen and Chiu). We performed inter-rater reliability test between the two experts' ratings. We found that the agreement rate was 82% for inter ictal activities (epileptiform discharges) and the agreement rate for ictal activities (seizure discharges) was 94%. We

only use signal segments where there is consensus in the interpretation from both experts.

3.3 Data Preprocessing

Physiologically meaningful EEG frequencies are below 60Hz. Frequencies over 60Hz are usually generated from electromyographic signals (often generated from skeletal muscles) and electrical power lines. For the analysis of epilepsy, we are not concerned about signals originating from skeletal muscles. Therefore in the data preprocessing stage, we filter out artifacts generated from electromyographic signals and electrical power lines using a low-pass filter and a notch filter, respectively. The five primary EEG sub-bands: delta, theta, alpha, beta, and gamma span the 0–60 Hz frequency range.

A low-pass filter is a filter that allows low-frequency signals to pass but reduces (ideally eliminates) the amplitude of signals with frequencies higher than the cutoff frequency. The actual amount of attenuation for each frequency varies from filter to filter. Signal filters can also be adjusted to allow high-frequency signals to pass and reduce the amplitude of signals with frequencies lower than the cutoff frequency. In short, a low-pass filter smoothes out signals by removing short-term oscillations, leaving only the long term trend. A high-pass filter, on contrary, filters out the long trend and leaves short term oscillations. Low-pass and high-pass filters can be used concurrently to obtain signals from a specific desired range of frequency. For our experiment, we use a low-pass filter set to filter out frequencies over 60Hz. The filtered signal has frequencies ranging from 1Hz to 60Hz, which contains the five primary EEG sub-bands: delta, theta, alpha, beta, and gamma. The next section discusses the

procedure of extracting the different EEG sub-bands for feature extraction.

3.4 Wavelet Transformation

Wavelet transformation was developed to overcome the shortcomings of Fourier Transformation (FT). In general, FT only provides frequency information and not location information. This means that signal transformed using FT is susceptible to Heisenberg's Uncertainty principle; information about the frequencies present in a signal could be obtained, but not where and when the occurrences took place. Wavelet Transform (WT) transforms signals in the time domain to a joint time-frequency domain. This allows the capture of both frequency and location information which is useful in analyzing continuous signals such as EEG, EKG, and other bio signals. For the detection of spike waves in epilepsy analysis this is a very useful feature. It allows us to pinpoint the exact locale of epileptic spike occurrence.

Each discrete wavelet transformation operation takes input signal and decomposes the signal into low and high frequencies by passing the signal through a low-pass filter and high-pass filter convoluting with impulse response G. Filter masks vary depending on the purpose of application. The result of the operation yields two decomposed signals; high frequency signal and low frequency signal. The transform of a signal x (length n) with filters g (high-pass filter) and h (low-pass filter) is shown in Equations 1 and 2. The decomposition process halves the time resolution and removes half the frequencies of the signal, and therefore, according to Nyquist's rule, half the samples can be discarded. The subsampling operator is used after each decomposition process. Since each decomposition process halves the signal sample size, the input signal must be a multiple of 2^n where n is the number of levels. To further obtain lower frequency signal,

the low frequency output signal is used as input for next level of discrete wavelet transformation.

$$Y_{low}[n] = \sum_{k=-\infty}^{\infty} x[k]g[2n-k] \quad (1)$$

$$Y_{high}[n] = \sum_{k=-\infty}^{\infty} x[k]h[2n-k] \quad (2)$$

Our research EEG data are sampled at 200Hz discretely. We take 2 seconds of EEG signals as one epoch which translates to 400 samples per epoch. It is possible to increase the longitude of sample epoch. We choose to use 2-second intervals to comply with the sampling requirement for 4-level decomposition process. Discrete wavelet transformation was used to decompose the signals into the five primary EEG sub bands. In practice, cascaded decomposition scheme with frequency reduced by half with each stage of decomposition is used to sequentially decompose the original signal, as shown in Figure 4. We decompose the signal four times to obtain the five major EEG sub-bands. Figure 5 shows an example of decomposed signal sample into the wanted sub-bands using Daubechies filter.

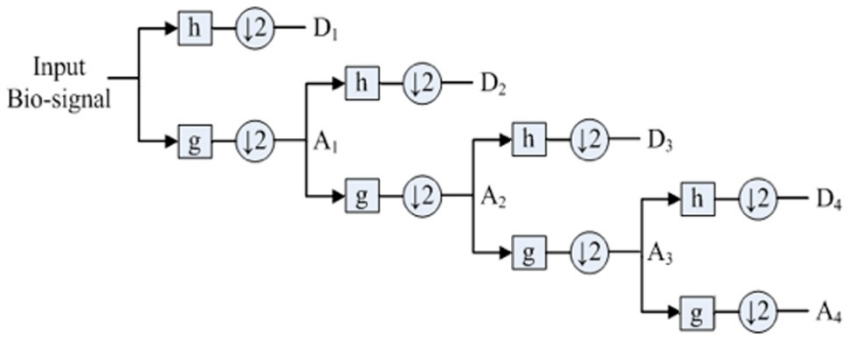

Figure 4 Wavelet Decomposition Scheme for EEG Signals

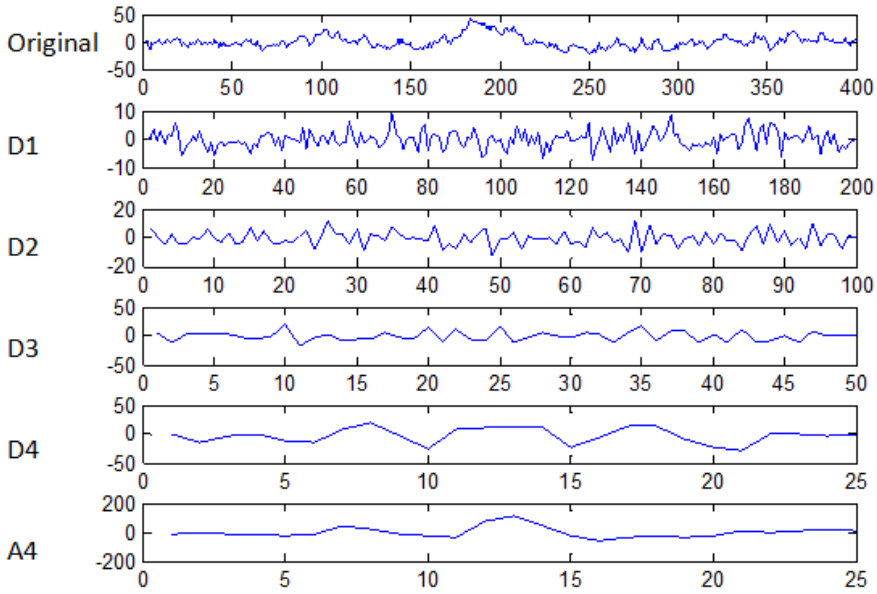

Figure 5 The decomposition results using the Daubechies filter

The input signal for wavelet transformation has been filtered to contain only frequencies below 60Hz. The decomposed bands D1, D2, D3, D4, and A4, have frequencies ranging from 30-60, 15-30, 8-15, 4-8, and 0-4Hz, respectively. These frequency ranges are the five primary EEG sub-bands that researchers use. The results D1, D2, D3, D4, and A4 are the Gamma, Beta, Alpha, Theta, and Delta sub-bands.

3.5 Feature Extraction

Once the five primary EEG sub-bands are obtained, we extract features from each of them. After consulting with Dr. Chiu of National Taiwan University Hospital's Neurology Department we devised series of features which doctors consider physiologically meaningful for the classification of epileptic waveforms. The goal of this research is to be able to distinguish between three classes of waveforms; normal, seizure, and spike. Ultimately, the aim is to be able to forecast seizure events. According to Dr. Chiu, the occurrences of spike events might shed light on understanding seizure

more in depth and may even be used to predict seizure events. So the focus for this research is to increase spike recognition rate. With this in mind, together with the help of Dr. Chiu, we designed features specifically targeting at increase of spike recognition rate. Some of standard statistical features are also used. The following section discusses the two features we use that we believe have the most impact on spike detection.

3.5.1 Approximate Entropy (ApEn)

In analyzing the regularity of time-series data, approximate entropy (ApEn) is often used. ApEn was initially developed for analyzing medical data such as heart rate and endocrine secretion. ApEn quantifies the regularity and predictability of time-series data. We believe this is the key to effectively distinguish between different EEG patterns, especially seizure. Seizure is defined as a transient symptom of "abnormal excessive or synchronous neuronal activity in the brain". The synchronicity of neuronal activity should be distinguishable from normal activity using ApEn. The calculation of ApEn process is listed in equations (3) to (5) with a signal S (finite length N) was performed by following step 1 through step 6. The parameter m represents the length of the sampling window, which was the dimension of the vector to be shifted, and r is the value of the threshold representing the noise filter level chosen in the range 0.1 to 0.9. Large ApEn values imply irregularity of a data sequence, whereas small values imply regularity. The section below describes the process of calculating ApEn for a vector of data sequence:

(1) $\boldsymbol{S} = [x(1), x(2), \ldots, x(N)]$ is the vector of data sequence.

(2) $\boldsymbol{x}^*(i)$ is a subsequence of $\boldsymbol{S}$ such that $\boldsymbol{x}^*(i) = [x(i), x(i+1), \ldots, x(i+m-1)]$ for $1 \leq i \leq N - m + 1$.

(3) Let $r = k \times \mathrm{SD}$ for $k = 0.1$ to 0.9, where SD is the standard deviation of $\boldsymbol{S}$.

(4) For each $1 \le \boldsymbol{x}^*(i), \boldsymbol{x}^*(j) \le N-m+1$, $i \neq j$, d [ ] is the Euclidean distance operator.

$$C_i^m(r) = \frac{\sum_{j=1}^{N-m+1} d[\boldsymbol{x}^*(i), \boldsymbol{x}^*(j)]}{N-m+1} \tag{3}$$

$$where\ d[\boldsymbol{x}^*(i), \boldsymbol{x}^*(j)] = \begin{cases} 1, \boldsymbol{x}^*(i) - \boldsymbol{x}^*(j) \le r \\ 0, otherwise \end{cases}$$

The quantity $\Phi^m(r)$ is calculated as

$$\Phi^m(r) = \frac{1}{N-m+1} \sum_{i=1}^{N-m+1} \ln C_i^m(r) \tag{4}$$

Finally, the ApEn is defined as follows:

$$ApEn = \Phi^m(r) - \Phi^{m+1}(r) \tag{5}$$

3.5.2 Total Variation

In mathematics, total variation can have meanings and interpretations depending on usage. Total variation is used mainly in de-noising image, and differential equation analysis. The concept of total variation for a real-valued continuous function can be viewed as an integral involving the function on a defined domain. For complex measures total variation has different definition. Equation 6 shows the definition of total variation for single-measure (real values) functions.

$$V_b^a(f) = \int_a^b |f'(x)| dx \tag{6}$$

The definition of total variation can be interpreted as the sum of “acceleration” of a given function. A large value of total variation implies faster value fluctuation of values over the defined interval, and vice versa. After examining real examples of different EEG activities, we believe that total variation could be used as an indicator for spike detection.

3.5.3 Feature Extraction Summary

Aside from ApEn and total variation, we also included three other commonly used features for the analysis of bio signals, namely, energy, skewness, and standard deviation. Skewness measures the asymmetry of a distribution, energy measures the total energy displacement of neurons, and standard deviation measures the dispersion variation of the EEG waves. Each 2-second epoch consists of signals from all 16 channels. One of our main focuses is to study the effectiveness of using bipolar montage in the detection of epileptic waveforms. Therefore, 16 bipolar montage signals are calculated using the 16 unipolar montage signal values. The 4-stage wavelet transformation decomposes the filtered signal into 8 sub frequency ranges, among them are the five primary EEG sub-bands. The 8 decomposed signal parts go through the feature extraction process. Entropy, total variation, standard deviation, skewness, and energy are the five feature types that can be extracted from each decomposed signal. Then each feature type extracted from each channel's decomposition bands are taken for calculation of statistical features; sum, max, min, and average. The statistical features of different feature types might help us distinguish different epileptic states. If there is any abnormal activity across all EEG sub-bands, then the statistical features might magnify the anomaly.

Since bipolar EEG shows the potential difference in neighboring electrodes. A clinically defined spike is more clearly shown to neurologists under bipolar EEG montage. Therefore, theoretically, it is possible to use bipolar EEG montage directly without using wavelet transform for detection of spike waves. We test this hypothesis by also directly extracting features from bipolar EEG signal values. Table 1 summarizes all the possible features our system can extract from a 2-second EEG epoch. In total 1700

features can be extracted. We choose all or a subset of these features for experiment which is discussed in detail in the experiment design section.

Table 1 Feature Extraction Summary

Feature type / Data Type	Total Variation	Standard Deviation	Approximate Entropy	Skewness	Energy
Unipolar Montage (With WT)	**160**	**160**	**160**	**160**	**160**
Bipolar Montage (With WT)	**160**	**160**	**160**	**160**	**160**
Bipolar (Without WT)	**20**	**20**	**20**	**20**	**20**

3.6 Feature Selection

3.6.1 Fisher Score

Classification of instances can be done very accurately if there exist enough differences in some feature between classes. Often we do not know the degree of difference that exist across different classes in different features. The fisher score ranks the difference rate of the feature between different classes. It can determine the most relevant features for classification. This is done using discriminative methods and generative statistical models. Fisher score uses the fisher function to rank the feature value and sets the importance of the features. Features are iteratively tested to achieve high accuracy.

For example, assume that there are *n* training samples for a class, and each training

sample has *h* types of features. The training sets of the two classes for $\Omega_{a,b}$ are denoted $C_a = \{V_{a,1}, V_{a,2}, V_{a,3}, \dots, V_{a,n}\}$ and $C_b = \{V_{b,1}, V_{b,2}, V_{b,3}, \dots, V_{b,n}\}$, where a training vector $V_{i,j}$ initially contains h feature values that can be represented as $V_{i,j} = [v_{i,j,1}, v_{i,j,2}, \dots, v_{i,j,h}]$. The discriminant ratio $I_{a,b,k}$ of the *k*-th feature can be evaluated using (6) for a hyper-plane $\Omega_{a,b}$. Equation (7) aims to evaluate the level of separability between the two classes (*a* and *b*) as well as the stability in the same class for the *k*-th feature value.

$$I_{a,b,k} = \frac{(\mu_{a,k} - \mu_{b,k})^2}{\sigma_{a,k}^2 + \sigma_{b,k}^2} \tag{7}$$

where $\mu_{i,k}$ and $\sigma_{i,k}$ denote the mean and standard deviation values, respectively, of the k-th feature for all training samples in class *i*, which can be evaluated using equations (8) and (9), respectively.

$$\mu_{i,k} = \frac{1}{n}\sum_{j=1}^{n} v_{i,j,k} \tag{8}$$

$$\sigma_{i,k} = \sqrt{\frac{1}{n-1}\sum_{j=1}^{n}(v_{i,j,k} - \mu_{i,k})^2} \tag{9}$$

For a hyperplane, the discriminant ratio of a feature type needs to be evaluated first. They are then sorted in descending order. The *k*-th feature in the sorted list corresponds to the *k*-th best feature for a hyperplane. The training problem thus becomes the finding the *m* best features from the feature set and then forming the feature vectors.

The process of feature selection for a hyperplane $\Omega_{a,b}$ is stated as follows: The discriminant ratios for all features are first arranged as an array $I_{a,b} = [I_{a,b,1}, I_{a,b,2}, \dots, I_{a,b,h}]$. The elements in array $I_{a,b}$ are first sorted in descending order. We select the feature according to the value of $I_{a,b}$, then evaluate the

accuracy of the two classes. The overall architecture of Fisher Score is shown in Figure 6.

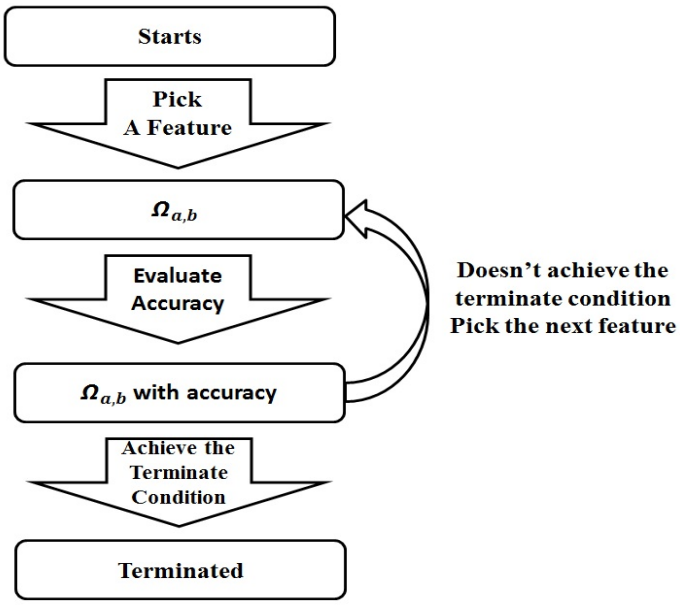

Figure 6 Fisher Score Architecture

3.6.2 Genetic Algorithm

Genetic algorithm (GA) is a search heuristic that aims to generate optimizations and solutions to search problems. The algorithm was proposed by Prof. John Holland. The idea behind GA was to mimic natural selection process, which involves mutations, inheritance, crossover, and selection. In nature, the selection process yields the fittest subjects; the survivors. For search problems, such as the problem of finding optimal parameters for an operation, the selection process tests the "fitness" of parameters and finds optimal parameters. GA simulates cells in nature, with its main component elements being genes, chromosomes, group, and fitness function. For each generation, fittest cells have the best current chromosomes and are the survivors. The surviving cells evolve generation to generation attempting to become better adapted to the

environment. GA starts by creating a population of randomly generated individuals represented in binary strings consisting of 0s and 1s. The fitness of each individual is evaluated for each generation and the fittest individual is selected. Surviving individuals can mutate, recombine, and mate with each other to generate the next genetic generation. Then the algorithm checks if the termination condition has been achieved or not. If the termination condition is not met, the selection process goes on until termination condition is achieved. Figure 7 illustrates the architecture of GA.

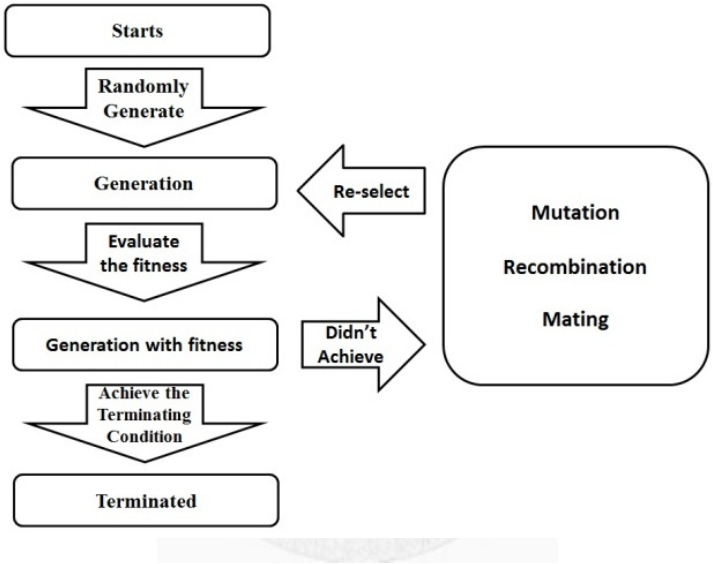

Figure 7 Genetic Algorithm flow diagram

3.1.1.1 Chromosome

A chromosome in a population represents a possible solution to the problem. Each chromosome is encoded with a binary string consisting of 0s and 1s. Each bit represents some characteristic to the solution. The longer the chromosomes, the more difficult the problem is as the number of possible solutions exist, leading to more generations required to obtain an optimal gene. In practice, it is possible for overly long chromosomes to lead to premature termination as sometimes, due to machine or time

limitations, convergence is forced and global solution is not obtained. In such case, accuracy of GA results would be negatively affected.

3.1.1.2 Fitness Function

A fitness function is a particular type of objective function that prescribes the optimality of a solution in a genetic algorithm so that the particular chromosome may be ranked against all the other chromosomes. It can evaluate whether the chromosomes are suitable for the environment. It can also select the best individual by ranking all chromosomes from the generation. Individuals are ranked and assigned expectation values depending only on their ranking, not on absolute fitness. Each individual is assigned a rank based on its fitness. Assuming that the best individual in a population is ranked as first, the probability of selecting an individual is calculated as follows in equation (6):

$$P_i = \frac{n_{max} - \frac{(n_{max} - n_{min})(i-1)}{N-1}}{N} \quad where \ n_{max} + n_{min} = 2 \ and \ n_{max} \geq n_{min} \geq 0 \qquad (6)$$

3.1.1.3 Operators

A. Crossover

A crossover is a genetic operator used to vary the configuration of a chromosome or chromosomes from one generation to the next. A crossover selects genes from parent chromosomes and creates a new offspring with new genes. Several types of crossover exist; the section below describes three commonly used crossover methods.

i. One-Point Crossover

A one-point crossover picks two mating individuals as parents, then picks a random locale in the chromosomes, known as crossover point and all data beyond that point in

either organism string is swapped between the two parent organisms.

ii. Two-Point Crossover

This is roughly the same as the one-point crossover, but it has two crossover points and swaps the middle part between two crossover points.

iii. Cut and Splice

Cut and splice means that parts with different lengths can be swapped between the two parents.

B. Mutation

In genetic algorithms, mutation is a genetic operator used to maintain genetic diversity from one generation of a population of algorithm chromosomes to the next. Mutation alters one or more gene values in a chromosome from its initial state. It can result in entirely new gene values being added to the gene pool. Mutation occurs during evolution according to a user-definable mutation probability. It can help to prevent the population from stagnating at any local optima.

3.7 Classification

3.7.1 Support Vector Machines

On the basis of characteristics, sometimes we would like to divide some of the data in a group of data into two groups. There are many methods that have good results for data classification. These include nearest neighbor, neural networks, decision trees, and support vector machines. The average accuracy rate of these methods is not much different; however, support vector machine has the advantage in that it is easy to use. The original idea of SVM classification is to use a linear separating hyperplane to create

a classifier. The vector that effects separation is called a “support vector”.

Support Vector Machines (SVMs) is a new technique for data classification. It is a relatively new learning process that is highly influenced by advances in statistical learning theory and a sufficient increase in computer processing power in recent years. Support vector machines use the given training sample sets to build the categories and train the model to predict unknown data. Each training sample data maps into the hyperplane and becomes the classification point. Support vector machines construct a hyperplane, or set of hyperplanes, in a high-dimensional or infinite-dimensional space, which is used for classification. We usually hope that the hyperplane will divide the data into two groups, with the same type of data being grouped onto the same side of the hyperplane. However, it is often the case that the data is far from linear and the datasets are inseparable. To allow for this, kernels are used to non-linearly map the input data to a high-dimensional space (feature space). The new mapping is then linearly separable. A simple illustration of this is shown in Figure 8.

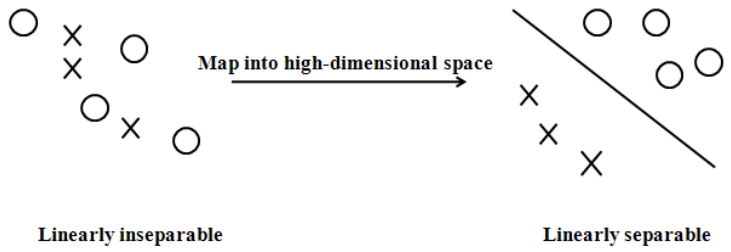

Figure 8 High-dimensional mapping

A classification task usually involves training and testing data consisting of some data instances. Each instance in the training set contains one class label and several values. The goal of an SVM is to produce a model that predicts the target value of data instances in the testing set given only the attributes. To use an SVM, we may need to

map the input space into a high dimensional feature space to realize the linearity of the classifier. By feeding the algorithm with a set of training data, it can determine a classifier or an optimal hyperplane. We can then use it to classify test data in order to observe the accuracy of the classification. If the accuracy is over a certain threshold, we use the classifier to classify the incoming unknown data. This is the goal of the development of the SVM. It has good performance in a wide variety of applications, such as text categorization, image recognition, hand-written digit recognition, and the determination of cancer cells.

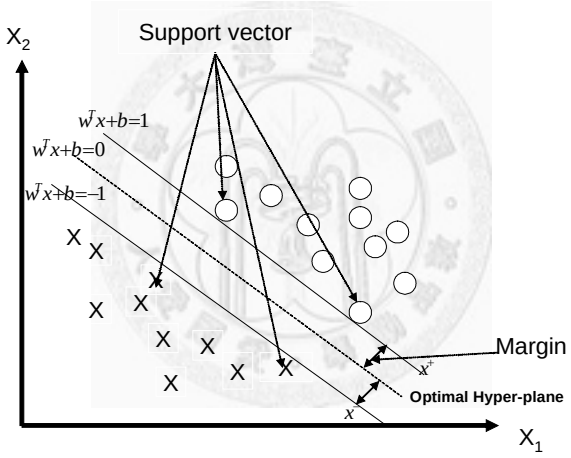

Figure 9 Support vector and separating hyperplane

Given a training set of instance-label pairs $(x_i, y_i), i = 1, \ldots, l, x_i \in R^n, y \in \{1, -1\}^l$

The hyperplane can be expressed as

$$< w^T \cdot x > + b = 0 \qquad \text{where} \quad w = [w_1, w_2, \ldots, w_n]^T$$
$$w_1 x_1 + w_2 x_2 + \ldots + w_n x_n + b = 0 \qquad x = [x_1, x_2, \ldots, x_n]^T$$

In conclusion, the definition of decision function is $f(\mathbf{x}) = sign(w^T \phi(x) + b)$

Then try to calculate margin $t(w)$ from Figure 12.

$w^T x^+ + b = +1$

$w^T x^- + b = -1$

$x^+ = x^- + tw$

$w^T(x^- + tw) + b = 1$

$-1 + w^T tw = 1$

$\mathrm{t} = {}^{2}/_{w^T w}$

We then try to maximize $\mathrm{t} = {}^{2}/_{w^T w}$, which is equal to minimize ${}^{w^T w}/_{2}$.

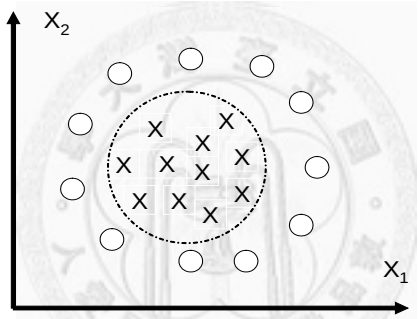

Figure 10 An example that is not linear separable

However, a practical problem may not be linearly separable, as illustrated by the example shown in Figure 10. In this situation, it is called “infeasible”.

3.8 Post-Classification Spike Matching

According to Dr. Chiu and our careful observation on real EEG data, spike occurrences often happen in “clusters”, that is, they appear to occur in groups with little time interval in between each other. This leads us to speculate that these spike clusters might be linked to seizure episodes. Our SVM classifier performs best at distinguishing between normal and seizure wave forms, which is not the main focus of this research.

We wish to improve spike recognition even further by examining normal outputs from our classifier. To do so we proposed the spike matching method. As far as we are concerned, no other research has added a post-classifier filter to further attempt to capture epileptic spikes.

Due to the idiosyncrasy of individuals with epilepsy, there is no one universally accepted definition for spike. Researches and neurologists generally have their own specific definitions for spike. In general, the definition of a spike is a short burst of electrical discharge, which is often followed by a slow “recovery” wave in which the neuron regenerates. Figure 11 shows an example of a classical definition spike. The trouble with defining a universally accepted spike is in defining 1) the magnitude of burst, 2) and the longevity of the burst and recovery. Figure 12 shows another real life example of spike which does not appear to conform to the general definition of a spike.

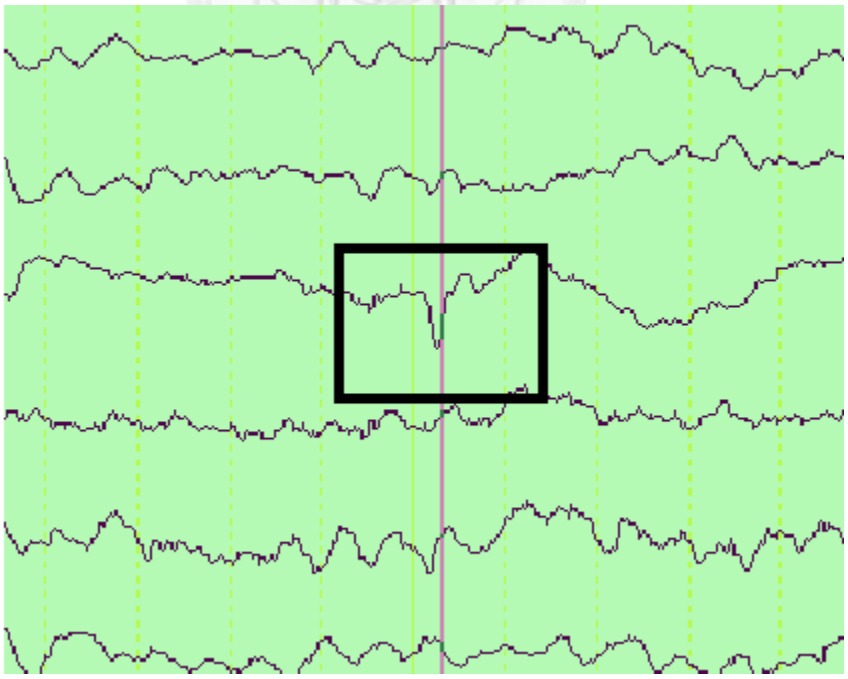

Figure 11 Real example of a classical spike.

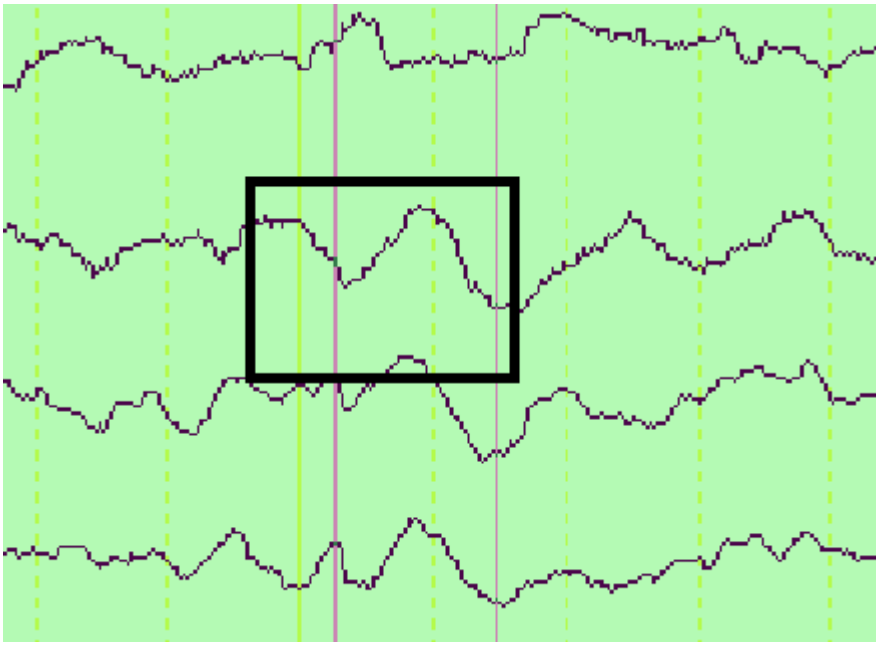

Figure 12 A case of a spike that is not so apparent and easy to spot.

Typically when a spike occurs, the opposite neighboring EEG readings show opposite signs. This is also called phase-reversal. Short pulse discharges alone does not imply spike, and neither does phase reversal. After consulting with Dr. Chiu, we designed the spike matching program to detect spike in two stages. The first stage detects short pulse of discharges. The second stage checks for phase-reversal. If an EEG segment passes these two criteria, then we consider it a spike.

The spike matching method alone is a poor tool to find spike segments given that real life EEG readings vary to a high degree from patient to patient and there are often other uncertainties such as signal noise and pure coincidences which would might lead a normal reading of EEG to be recognized as spike. This is why most researches turn away from using spike matching programs to spot spikes.

Since our main focus is to find clusters of spikes that occur within a short time frame, we are more meticulous in the short time frame following a spike confirmed by

SVM classifier. Once the SVM classifier classifies an EEG epoch as a segment containing spike(s) the following 10 seconds of EEG are not only screened by the classifier but also by the spike matching block. Spike matching is more lenient than the SVM classifier and therefore is only used right after a spike has been found by SVM. If the classifier classifies any 10-second segment after a spike normal, then the follow-up SVM-classified normal segment will get reexamined by the post spike matching block to ensure that no spike pass by undetected; this would improve spike detection rate. There is, however, a disadvantage where the overall normal recognition rate would drop; increasing false alarm rates. Tolerable rates of false forecast of epileptic seizure in exchange for higher sensitivity would not cause a big problem in patients with epilepsy. In a real-life scenario, a patient upon receiving a seizure warning only has to stop whatever he or she is doing and make sure that no potential danger is around should he or she lose control of his body temporarily due to a seizure episode. In contrast, if a seizure is not forecasted, the patient would be in danger of potential threats from seizure episodes. Activities even as simple as walking stairs could lead to severe injury.

3.9 Experiment Design

Once we had our system implemented, we test different feature configurations hoping to achieve an acceptable classification rate using the most effective feature extraction scheme. Since the ultimate goal of this research is to one day be able to forecast seizure occurrence, Dr. Chiu, the neurologist we collaborate with, suggests that it is more practical for the nature of this research to focus more on the recognition of epileptic spikes rather than seizures.

In theory, Fisher Score and Genetic Algorithm should be able to derive the best

features for classification and there is no need to use only a subset of all possible features. However, due to time, complexity, and memory limitations of computing, often premature convergence is forced and only local optimum of a solution is found instead of a global optimum. It is possible to obtain better results using only a subset of all features. Therefore, we conduct a series of experiments using different feature subsets for training and prediction and try to find an efficient feature extraction configuration.

For all our experiments, we use the same set of raw EEG data obtained from NTUH. The data consists of 83 minutes of selected EEG data which were annotated by neurologists. We first test with all the 1700 features described in feature extraction summary section. The results from using all features were used as benchmark for comparison.

From the benchmark classification run we obtain the list of top most important feature types provided by Fisher score and genetic algorithm. Using that list, we then proceed to test classification using only the feature types within. The first experiment is to compare classification results between using all features and using only the most important feature types.

The second experiment we conducted was to test our hypothesis about the effectiveness of using bipolar EEG montage for feature extraction. For this experiment, we use unipolar montage related features for classification.

Our third experiment was to test the effectiveness of the post-classification spike matching. We employ the post-classification spike matching module and assess the classification results. Ideally we want to increase sensitivity even further without suffering too much specificity.

Chapter 4 Experimental Results

4.1 Seizure and Spike Detection

4.1.1 All Features (Unipolar and Bipolar Montage)

The annotated short term EEG records we obtained includes 1939 2-second epochs of normal activity, 436 2-second epochs of spike activity, and 444 2-second epochs of seizure activity. Notice that the proportions of different waveforms do not reflect real-life occurrence rates of these waveforms. Spikes and seizure do not occur frequently out of the norm. However, these two waveforms are upmost detection priority and therefore more samples are needed for a strong prediction model.

Half of the 2-second epochs are taken for training, and the other half are taken for prediction. In total, our system currently can output 1700 features for each 2-second epoch. We use these 1700 features together to obtain an initial classifier. This classifier is also used as benchmark for other future configurations. Table 2 shows the recognition rates of classifier using all features.

Table 2 Recognition rates of classifier using all features

Input / Output	Normal Total: 969	Spike Total: 218	Seizure Total: 222
Normal	**908**	**35**	**0**
Spike	**61**	**181**	**0**
Seizure	**0**	**2**	**222**
Recognition Rate	**93.7049%**	**83.0275%**	**100%**

It can be seen from Table 2 that recognition rate for seizure is high. Seizure activities usually last for a few seconds to a few minutes and display wave patterns that are very distinguishable from normal wave patterns. The problem of classifying between normal and seizure waveforms has practically been solved by other previous researches. A research at MIT constructs patient-specific classifiers that detect the onset of an epileptic seizure through analysis of the scalp EEG shows promising results in detection of real-time EEG seizure detection applications. What is worth noticing is that there seems to be a difficulty in distinguishing between spike and normal waves. This is of no surprise since a typical spike only lasts for about 70 milliseconds with no noticeable signs ahead. Our feature extraction takes 2-second epochs; statistical features easily masquerades intrinsic characteristics of a spike, contributing to the lower recognition rate for spike.

4.1.2 Most Influential Features

From the first classifier with GA and Fisher Score we obtained a feature

importance list where the impact of feature is listed. The list not only reveals the effectiveness of features but as well as shows us what we should focus on in terms of feature extraction. Table 3 shows the top 5 ranked feature types. The table shows us the type of feature and which sub band the feature type was used. A feature type represents a range of features using same feature extraction function. The top five features types represent about 80% of top 100 features out of 1700 features.

Table 3 Top ranked feature types

Sub-band	**Feature Type**
Delta	**ApEn**
Theta	**Total Variation**
Alpha	**ApEn**
Alpha	**Total Variation**
Theta	**ApEn**

The top most common and influential features, as shown in Table 3, all have two things in common; 1) they all derived from lower frequency bands, 2) They are all either entropy or total variation related. We have yet to consult experts in epilepsy about the first phenomenon. The second phenomenon confirms our hypothesis about the effectiveness of using ApEn and total variation for detection of spikes and seizures.

Given the top ranked feature types are mostly ApEn and total variation related, we test the effectiveness of a classifier using solely these two types of features. A classifier with only these two feature types was created. The total number of features for this classifier was 740. Results are shown in Table 4.

Table 4 Recognition rates of classifier using only ApEn and total variation

Input / Output	Normal Total: 969	Spike Total: 218	Seizure Total: 222
Normal	**914**	**34**	**1**
Spike	**55**	**184**	**1**
Seizure	**0**	**0**	**220**
Recognition Rate	**94.324%**	**84.4037%**	**99.0991%**

The classifier using only ApEn and total variation showed slightly better recognition rates for spike and normal. Although the result is in accordance with our hypothesis about ApEn and total variation being the major feature types that could be used to tell apart spike and other waveforms, the improvement is still not enough for spike recognition rate to increase to a good level. Since spike recognition is the main focus of this study, we employ other methods to increase spike recognition rate which is described in detail in the post-classification spike matching at section 4.1.4.

4.1.3 Feature Extraction Using Unipolar Montage Values

As far as we are concerned, not many researches use bipolar montage for analysis. Mathematically, different EEG montages are merely different ways of presenting the same data. From unipolar montage, values of other montage types can be calculated easily by either obtaining differences between channels or averaging neighboring channels. However, the doctors we collaborate with use bipolar montage when they screen patients' EEG. We try to design algorithm for classification by understanding

how doctors screen EEG. Figure 13 shows the same segment of EEG under unipolar and bipolar montages. The figure on the left is the segment of EEG displayed using unipolar montage. The figure on the right is the same segment of EEG but displayed using bipolar montage. Notice the spike present under bipolar montage is much clearer than under unipolar montage due to the visibility of phase reversal wave pattern under bipolar montage.

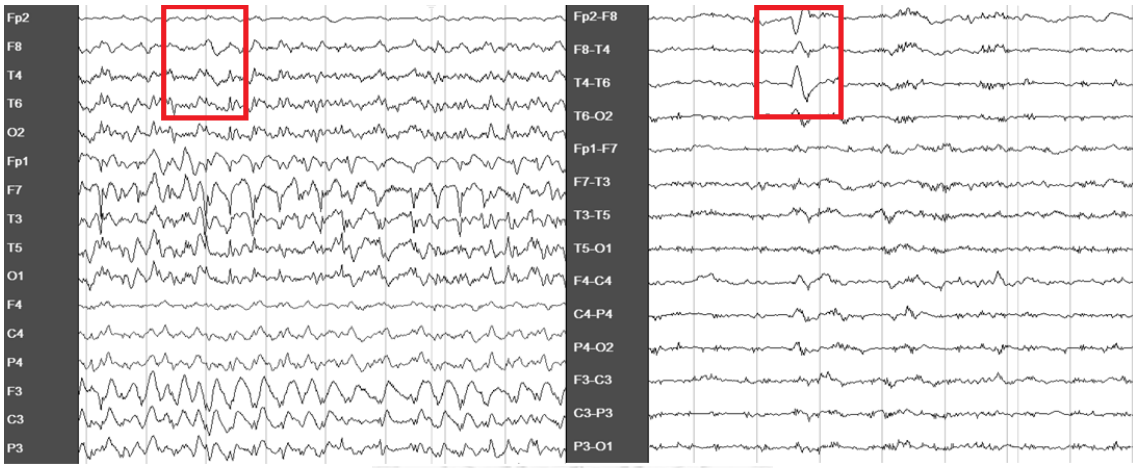

Figure 13 Example of EEG segment displayed under different montage configurations

The improvement of incorporating bipolar montage values for feature extraction and for classification is shown in Table 5. The table shows recognition rates of a classifier using only unipolar values. We can see that recognition rates for spike and normal are lower than the other classifiers we used for this research, which all included bipolar montage values. From the table we can also see that recognition rate for seizure is high. For classifiers only needing to distinguish between seizure and normal, using unipolar montage values is good enough. The main reason for usage of bipolar montage is try to magnify intrinsic characteristics of spike.

Table 5 Recognition rates of classifier using unipolar montage values only

Input / Output	Normal Total: 969	Spike Total: 218	Seizure Total: 222
Normal	**898**	**44**	**0**
Spike	**71**	**174**	**0**
Seizure	**0**	**0**	**222**
Recognition Rate	**92.6729%**	**79.8165%**	**100%**

4.1.4 Increasing Spike Recognition Rate Using Post-Classification Spike Matching

The post spike matching program takes bipolar montage values as input (not features) and checks for phase reversal and also checks for any EEG segments with template spike waveforms. The matching algorithm is not used alone due to the existence of artifacts, noise, and the fact that spike waveforms differ from person to person. As mentioned before, the spike matching program offers high sensitivity but low specificity. Nevertheless, we decided to use a spike matching program to try to find spike clusters. When the classifier takes features of an EEG epoch and classifies the segment spike, the spike matching program is then used for the following 10 seconds of EEG input. The idea is that spikes occur in clusters and once one spike has been spotted, the probability of spikes occurring within the follow-up short time frame increases. Standalone spikes pose relatively lower threat than spike clusters. The 10 seconds of EEG following a classifier-labeled spike is screened by the spike matching program as

well as being classified by the classifier. If either the classifier or the spike matching program labels any of the following 5 EEG epochs (10 seconds) spike, then we consider the corresponding epoch(s) as segments containing spike.

We employ the post-classification spike matching program on our classifier with the best performance, which is the one using only ApEn and total variation-type features. Table 6 shows the recognition result of using our post-classification spike matching program to further increase recognition rate for spike. Indeed, the spike recognition rate increased. However, it only increased by 2% at a cost of 5% decrease in normal recognition rate. Normal recognition rate dropped to about 91%, which might not be practical since this might translate to too many false alarms, disturbing patients' daily life.

Table 6 Recognition rate of classifier with post-classification spike matching

Input / Output	Normal Total: 969	Spike Total: 218	Seizure Total: 222
Normal	**878**	**29**	**1**
Spike	**91**	**189**	**1**
Seizure	**0**	**0**	**220**
Recognition Rate	**90.6089%**	**86.6972%**	**99.0991%**

We designed the spike matching program to analyze bipolar montage EEG values. We used derivatives to measure the speed of ascend of a spike. Sometimes doctors look for phase-reversal when screening for spike. Figure 14 shows an ideal depiction of

phase reversal. When a spike occurs, the discharge spreads out to the neighbors much like water ripples caused by water droplet. Again, we use derivatives to look for phase reversal. Opposite but same magnitude derivatives between two electrodes imply a spike has occurred exactly in the middle between the two electrodes. Opposite but different magnitude derivatives between two electrodes imply that a spike took place between the two electrodes, but not at the midpoint of the two locations. In such case, a phase lag also occurs where the discharge travels to the closer electrode first then to the other neighbor electrodes. When a spike has occurred and phase lag is present due to the location of the spike, it is difficult to tell between existence of phase reversal or just a coincidence exist in which derivatives happen to be opposite. We have yet to solve the phase lag problem and to come up with a better post classification spike-matching algorithm. It can be said that our approach for spike matching is a failure. However, we believe that adding a post-classification spike matching program could be a good solution to finding spike clusters. More studies need to be conducted to design a good post-classification spike-matching system.

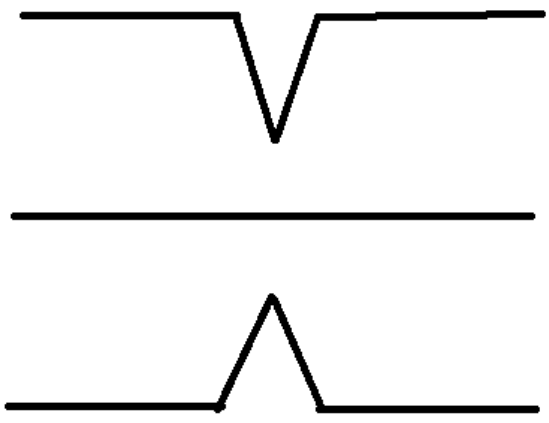

Figure 14 Phase reversal

Chapter 5 Conclusion and Future Work

5.1 Conclusion

The usage of EEG for diagnostic for epilepsy and other brain-related diseases has great potentials for helping people with brain disorders as EEG machines become cheaper and more portable. This thesis presents a classification system for seizure and spike detection, which might one day be made into products helping people with epilepsy in a real-time fashion. The system was tested using real data from NTUH and obtained good preliminary results. The results of this research mainly come from the collaboration of experts in neurology. This research, alongside with others of the same field, may serve as foundation to many brain-related applications.

5.2 Future Work

The first thing that we aim to improve is the design of post-classification spike matching system. We believe the overall system architecture for our classification system makes rational sense in that it complies with the way doctors screen EEG. We need to, however, find ways to overcome some fundamental problems such as the detection of phase reversal from bipolar montage. With a good post-classification spike matching module we can achieve high spike recognition rate. With high spike recognition rate, the next research will be finding correlation between spike and seizure, eventually leading to a seizure forecasting system; the ultimate goal of this research.

REFERENCE

[1] H. M. d. Boer, M. Mula, and J. W. Sander, “The global burden and stigma of epilepsy,” *Epilepsy & Behavior*, vol. 12, no. 4, pp. 540–546, February 2008.

[2] A. Strzelczyk, J. Reese, R. Dodel, and H. Hamer, “Cost of Epilepsy: A Systematic Review,” *PharmacoEconomics*, vol. 26, no. 6, pp. 463–476, June 2008.

[3] World Health Organization. International Classification of Functioning Disability and Health: ICF. Geneva: WHO; 2001.

[4] Health Grades Inc, “Statistics by Country for Epilepsy,” May 2003, http://www.cureresearch.com/e/epilepsy/stats-country.htm

[5] H. Witte, L. D. Iasemidis, and B. Litt, “Special issue on epileptic seizure prediction,” *IEEE Trans. Biomedical Engineering*, vol. 50, pp. 537–539, 2003.

[6] R. S. Fisher, W. V. E. Boas, W. Blume, C. Elger, P. Genton, P. Lee, and J. Engel Jr., “Epileptic seizures and epilepsy: Definitions proposed by the international league against epilepsy (ILAE) and the international bureau for epilepsy (IBE),” *Epilepsia*, vol. 46, no. 4, pp. 470–472, March 2005.

[7] W. Weng and K. Khorasani, “An adaptive structure neural network with application to EEG automatic seizure detection,” *Neural Network*, vol. 9, pp. 1223–1240, August 1996.

[8] I. Güler and E.D. Übeyli, “Adaptive neuro-fuzzy inference system for classification of EEG signals using wavelet coefficients,” *Neuroscience methods*, vol. 148, no. 2, pp. 113–121, April 2005.

[9] I. Güler and E.D. Übeyli, “Multiclass support vector machines for EEG signals classification,” *IEEE Trans. Information Technology in Biomedicine*, vol. 11, no. 2, pp. 117–126, March 2007.

[10] E.D. Übeyli and I. Güler, “Features extracted by eigenvector methods for detecting variability of EEG signals,” *Pattern Recognition Letter*, vol. 28, no.5, pp. 592–603, November 2007.

[11] V. Srinivasan, C. Eswaran, and N. Sriraam, “Approximate entropy-based epileptic EEG detection using artificial neural networks,” *IEEE Trans. Biomedical Engineering*, vol. 11 no. 3, pp. 512–518, May 2007.

[12] S. Ghosh-Dastidar and H. Adeli, “Principle component analysis-enhanced cosine radial basis function neural network for robust epilepsy and seizure detection,” *IEEE Trans. Biomedical Engineering*, vol. 55 no. 2, pp. 512–518, February 2008.

[13] A. T. Tzallas, M. G. Tsipouras, and D. I. Fotiadis, “Epileptic seizure detection in EEGs using time frequency analysis,” *IEEE Trans. Information Technology in Biomedicine*, vol. 13, no. 5, pp. 703–710, September 2009.

[14] S. Demont-Guignard, P. Benquet, U. Gerber, and F. Wendling, “Analysis of intracerebral EEG recordings of epileptic spikes insights from a neural network model,” *IEEE Trans. Biomedical Engineering*, vol. 56 , no. 12, pp. 2782–2794, December 2009.

[15] A. T. Tzallas, M. G. Tsipouras, and D. I. Fotiadis, “Epileptic seizure detection in EEGs using time frequency analysis,” *IEEE Trans. Information Technology in Biomedicine*, vol. 13, no. 5, pp. 703–710, September 2009.

[16] Shen, C.P., et al., A Multiclass Classification Tool Using Cloud Computing Architecture. International Symposium on Network Enabled Health Informatics, Biomedicine and Bioinformatics (HI-BI-BI 2012), 2012.

[17] Lawrence J Hirsch, and Hiba Arif, “Electroencephalography in the diagnosis of seizures and epilepsy,” *UpToDate*, January 2010.

[18] David P Moore, *Textbook of Clinical Neuropsychiatry*, Hodder Arnold Publishers, Louisville, Kentucky, USA ,June 2008.