FEATURE SELECTION FOR PERSONALIZED EEG-BASED SEIZURE MONITORING

by

Genchang Peng

APPROVED BY SUPERVISORY COMMITTEE:

______________________________
Dr. Mehrdad Nourani, Chair

______________________________
Dr. Issa Panahi

______________________________
Dr. Joseph Friedman
This thesis is dedicated to my parents,
who always stand there as my spiritual support.
FEATURE SELECTION FOR PERSONALIZED EEG-BASED SEIZURE MONITORING

by

GENCHANG PENG, B.Eng

THESIS
Presented to the Faculty of
The University of Texas at Dallas
in Partial Fulfillment
of the Requirements
for the Degree of

MASTER OF SCIENCE IN
ELECTRICAL ENGINEERING

THE UNIVERSITY OF TEXAS AT DALLAS
December 2020
ACKNOWLEDGMENTS

I am deeply indebted to my research advisor, Dr. Mehrdad Nourani, for his insightful guidance and mentoring during my master’s program. His enthusiasm and endeavors towards interdisciplinary research will always encourage me to overcome the next challenges in the unknown field.

I thank Dr. Issa Panahi and Dr. Joseph Friedman for serving as committee members of my thesis defense. Their invaluable feedback helped me to polish this work. I thank Dr. Jay Harvey and Dr. Hina Dave of UTSW for their clinical expertise and suggestions during my research. Besides, I am grateful to all the instructors of the core and elective courses in our department, from which I have built a strong foundation for my research work.

I would like to thank my labmates in the Predictive & Analytics Technologies Lab for their novel ideas and viewpoints in various disciplines during the daily talks. I also thank all the peers and friends I have met in UTD for their caring help in my daily life.

Last but not least, I would like to sincerely thank my parents for their unconditional and unreserved love and support.

November 2020
Epilepsy refers to a group of neurological disorders characterized by neurons in brain that misfire and cause recurrent seizures. Epilepsy affects millions of people around the world. Epileptic patients who have intractable seizures may suffer from injuries or even sudden death. It is of great value and significance to provide continuous, reliable and timely seizure monitoring for epileptic patients. Scalp electroencephalogram (EEG) signal is widely considered the golden marker of epileptic seizures. A low-cost, wearable EEG-based seizure monitoring platform using machine learning technology will not only improve patients’ quality of lives, but also assist physicians for clinical diagnosis. Clinical EEG signals are collected from multiple channels (e.g., 10 to 30 electrodes), and usually have highly complex patterns that vary among different patients. To have a wearable monitoring device for a patient, it is important to determine a limited number of seizure-relevant channels (e.g., no more than 5) and select a few discriminative features of that subject.

This thesis presents two methodologies on personalized feature selection methodology to enhance EEG classification performances for wearable seizure monitoring. First, a feature-ranking approach using F-statistic value is presented to select discriminative features from fusion feature set consisting of spectral power and entropy features. The selection results of top-ranked features are visualized for personalized analysis. Second, to design a wearable EEG monitoring platform, a two-step feature selection strategy is proposed. Step 1 uses
linear discriminant analysis (LDA) to find seizure-indicative channels. Step 2 employs the least absolute shrinkage and selection operator (LASSO) method to select a subset of spectral features. What’s more, a personalization scheme is proposed to choose the best selection parameters of each subject. The proposed methods have been evaluated on public domain CHB-MIT EEG database, and the results are better or comparable to state-of-art works reported in literature.
TABLE OF CONTENTS

ACKNOWLEDGMENTS ................................................................. v

ABSTRACT ..................................................................................... vi

LIST OF FIGURES ........................................................................... x

LIST OF TABLES ........................................................................... xi

CHAPTER 1 INTRODUCTION ...................................................... 1

1.1 Background ........................................................................... 1

1.1.1 Epileptic Seizures ......................................................... 1

1.1.2 EEG ............................................................................... 1

1.2 EEG-Based Seizure Monitoring ............................................ 3

1.2.1 Motivation ................................................................. 3

1.2.2 Procedure ................................................................. 4

1.2.3 Challenges ............................................................... 4

1.3 Contribution ................................................................. 5

1.4 Related Work ............................................................... 6

1.5 Thesis Organization ....................................................... 7

CHAPTER 2 FEATURE SELECTION USING F-STATISTIC VALUES FOR EEG SIGNAL ANALYSIS .................................................... 8

2.1 Overview ........................................................................... 8

2.1.1 Motivation ............................................................... 8

2.1.2 Main Contribution ..................................................... 9

2.2 Background on EEG Features ........................................... 9

2.2.1 Power Spectral Density .............................................. 9

2.2.2 Sample Entropy ......................................................... 10

2.2.3 Permutation Entropy .................................................. 11

2.2.4 Spectral Entropy ......................................................... 12

2.3 Proposed Methodology .................................................. 12

2.3.1 Feature Extraction .................................................... 13

2.3.2 Feature Selection ...................................................... 13
<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Electrode position of international 10-20 system. The electrode distance are marked between two possible pairs.</td>
</tr>
<tr>
<td>1.2</td>
<td>An example of EEG waveform containing seizure period from 23 channels</td>
</tr>
<tr>
<td>2.1</td>
<td>Flowchart of feature selection using F-statistic values</td>
</tr>
<tr>
<td>2.2</td>
<td>Separating and grouping non-seizure data</td>
</tr>
<tr>
<td>2.3</td>
<td>Selection results for Subject-chb01 (a) and Subject-chb03 (b). The colored grids denote selected features using F-statistic values (darker color reflects larger F value).</td>
</tr>
<tr>
<td>3.1</td>
<td>Flowchart of two-step selection method.</td>
</tr>
<tr>
<td>3.2</td>
<td>Channel ranking of subjects: (a) chb02 and (b) chb05</td>
</tr>
<tr>
<td>3.3</td>
<td>Channel ranking of chb01 using: (a) PSD values, (b) ENT values and (c) PSD + ENT.</td>
</tr>
<tr>
<td>3.4</td>
<td>T-SNE visualization of LASSO selected subset for subject chb01.</td>
</tr>
<tr>
<td>3.5</td>
<td>Boxplot of feature values of subject chb17: (a) $\alpha$ PSD in T8-P8, (b) $\alpha$ PSD in P8-O2, (c) PermEn in T8-P8 and (d) PermEn in P8-O2.</td>
</tr>
<tr>
<td>3.6</td>
<td>Best features for seizure detection of subject chb10: (a) Raw EEG waveform of seizure event in channel F7-T7, 10 second epoch of (b) $\alpha$-PSD, (c) $\beta$-PSD, (d) SampEn and (e) SpecEn.</td>
</tr>
<tr>
<td>3.7</td>
<td>Barplot of classification results of ten subjects using SVM, KNN and RF based on average (a) F-1 score, (b) Sensitivity, (c) Specificity and (d) Accuracy. In each case the score is expressed as mean $\pm$ std.</td>
</tr>
</tbody>
</table>
LIST OF TABLES

2.1 Results of classification and feature selection ........................................ 16
2.2 Comparison with other works .................................................................... 17
3.1 Comparing Top-K Channels With full channels using fusion features and SVM classifier ................................................................. 27
3.2 Comparing Limited Channel Results using PSD, ENT and both (PSD + ENT) .......................................................... 30
3.3 Top-5 Channels of chb01 Using PSD, ENT and PSD+ENT ......................... 31
3.4 Ten features selected by LASSO for chb01 for $K = 3$ and $4$ ....................... 32
3.5 Classification of LASSO’s Selected Features and Full Features .................... 33
3.6 Best Selection and Classification Results .................................................. 36
3.7 Comparing Selected Features for chb02 and chb03 .................................... 38
3.8 Average run-time of Each Part on three subjects ....................................... 39
3.9 Comparison with other works on CHB-MIT Database .................................. 41
CHAPTER 1
INTRODUCTION

1.1 Background

1.1.1 Epileptic Seizures

Epilepsy is one of the most common neurological disorder that affects over 50 millions of people of all ages, races, social classes and geographical locations [1]. Epilepsy is seen in both the children and adults at the age 65-70 [2]. The prevalence of epilepsy in resource-poor countries was nearly two times higher than that in developed countries [3], and nearly 80-90% of epileptic patients from developing countries are believed to receive inadequate or no treatment at all [4]. Epilepsy inserts huge medical, social and economic burden on people around the world. In the United States, it costs over 15 billion dollars directly and indirectly due to epilepsy [5].

Epilepsy is characterized by repeated and recurrent seizures generated from excessive neuron misfirings or discharges in the brain. Such nerve-cell discharges may remain in a small area of the brain, giving rise to partial (focal) seizures, or starting immediately in the whole brain or spread from a small area to the whole brain, leading to generalized seizures. Patients who are experiencing seizures are often seen with sudden abnormal functions of their bodies and with losses of their consciousness [6]. Note that, a patient may have a combination of different seizure types. Additionally, the types, symptoms severity of seizure may change over the years, based on the age and maturation of his or her brain.

1.1.2 EEG

Electroencephalogram (EEG) measures the electrical potentials generated by the neuron activities of the brain. There are two types of EEG signal categorized by its acquisition method. The first is scalp EEG, which is a non-invasive approach that are collected through
electrodes glued on the patient’s scalp. The second is intracranial EEG (iEEG), an invasive method to place electrodes inside the brain. The iEEG can be further divided into 1) the electrocorticography (ECoG) which places electrode grids or strips on the exposed surface of the brain, 2) the stereoelectroencephalogram (stereo-EEG or SEEG) implanting small incisions in the scalp and skull [7]. Scalp EEG is efficient for long-term recordings, but is sensitive to electrical noise and motion artifacts during its recordings. Intracranial EEG provides clean and artifact-free recordings deeply inside the brain, but the required craniotomical operation is time-consuming and even risky. This thesis focuses on scalp EEG and its analysis and applications.

The electrode position of scalp EEG follows the international 10-20 system, as shown in Figure [1.1]. Each electrode is named as “letters+ numbers”. The letters are used to identify the lobe positioning, where F, T, C, P and O stand for Frontal, Temporal, Central, Parietal and Occipital, respectively [8]. Two letters can be combined to identify the hemisphere location, e.g., Fp denotes front-temporal lobe. The odd numbers (1, 3, 5, 7) and even numbers (2, 4, 6, 8) refer to left and right hemisphere, respectively, and z (zero) is the middle line. The values in percent between adjacent electrodes are the distance measured by the ratio of the brain size. For instance, the 20% between electrodes Fz and Cz equals to 20% of the total front-back distance of the skull. In real scenario, the bi-polar reference strategy is employed by taking the difference between two electrodes.

EEG is suggested as the golden marker of epileptic seizures [9]. Figure [1.2] gives an example of EEG waveform containing one seizure collected from 23 channels. The signal amplitude of different channels are relatively lower. When a seizure kicks in at 25 seconds, the signals ‘burst out’ with higher amplitudes, showing high fluctuation and chaos in waveform, and reflecting strong discharges inside the brain.

1The term EEG refers to scalp EEG for the rest of this thesis, unless noted otherwise
Figure 1.1. Electrode position of international 10-20 system. The electrode distance are marked between two possible pairs.

Figure 1.2. An example of EEG waveform containing seizure period from 23 channels

1.2 EEG-Based Seizure Monitoring

1.2.1 Motivation

In clinical environment, neurologists and doctors visually check the EEG waveform for seizure diagnosis. Such a manual work is expensive and time-consuming, and the visual inspection
might be subjective to the expertise and experience of different doctors. In daily life, without timely alert of seizure events, epileptic patients may suffer from injuries or even death when unexpected seizures kicks in. Therefore, a reliable, low-cost seizure monitoring platform is highly in demand both for clinical and daily use.

With advances in sensing devices and machine learning, EEG-based analysis has been proposed to achieve efficient seizure monitoring. This technology offers benefits for both patients and doctors. For epileptic patients, a portable and wearable seizure monitoring platform, which can timely alert the seizure events, will significantly improve qualities of epileptic patients’ lives. For doctors and clinicians, a seizure monitoring system will greatly assist their diagnosis by automatic classification of seizure and non-seizure across long-time EEG signals with high accuracy. Besides, an electronic profile of seizure recordings can assist care-givers for treatment planning.

1.2.2 Procedure

Typically, EEG-based seizure monitoring consists of several steps. First, the continuous EEG recordings are pre-processed into frames (e.g., 10 seconds). Then, different features (e.g., in time or frequency domain) are extracted from each frame to characterize signal patterns. Next, a classifier (e.g., neural networks or support vector machine) is built to classify the feature frames into multiple classes (e.g., seizure, non-seizure and pre-seizure). Based on different classification targets, the learning task can be categorized into seizure detection and seizure prediction. In seizure detection, a binary classification task is performed to classify each frame into seizure and non-seizure. Seizure prediction, on the other hand, involves multi-target classification, for example pre-ictal, inter-ictal or post-ictal. This thesis focuses on seizure detection and not prediction.

1.2.3 Challenges

To achieve a reliable and low-cost seizure monitoring, several challenges need to be addressed:
1. Feature selection for seizure pattern analysis: EEG signals are highly complex and non-stationary. To fully characterize seizure patterns, different types of features are extracted, such as statistical values, spectral powers and complexity parameters. Not all features are necessarily informative to classification tasks. To avoid the curse of dimensionality [10] in high-dimensional feature space, an efficient feature selection approach is required to find a compact and informative feature subsets for enhanced classification performances.

2. Personalization of learning process: Epilepsy patterns, including seizure locations and severity, vary among patients. For different subjects, their seizure-indicative features are different. To achieve a reliable monitoring results for a particular patient, the machine learning process, including feature selection, classifier learning, parameter tuning, need to be decided in a subject-dependent style.

3. Imbalanced data distribution: In EEG dataset, the seizure data of one patient last only a few (e.g., 1-30) minutes, whereas the non-seizure recordings are usually longer than ten hours. The classification results will be biased toward the dominated class. To accurately detect seizure events, this imbalance problem needs to be tackled.

1.3 Contribution

This thesis concentrates on feature selection methods for personalized seizure monitoring. Two different works are presented to enhance seizure classification performances. The main contributions are described as below:

In the first work, a feature ranking scheme using F-statistic values is proposed. A hybrid feature set containing spectral and entropy features is extracted from full-channel EEG recordings. Then, to select discriminative feature subset, the F-statistic value is applied to
rank the discriminative capacity of feature values. The selected features yield high classification scores, and the selection results are visualized for personalized analysis.

In the second work, for wearable applications, a two-step feature selection method is proposed. In the first step, to reduce the number of required electrodes, a channel ranking scheme using linear discriminant analysis (LDA) is applied to rank the discriminative capacity of EEG channels. After determining a limited number of channels (no more than 5), in the second step, the least absolute shrinkage and selection operator (LASSO) is employed to choose a few feature subset (i.e., less than 10) to further improve the classification performances. Additionally, a personalization strategy is adopted to find the best selection parameters of each subject. The impact of each step is extensively evaluated.

1.4 Related Work

EEG signal is highly complex and non-stationary. Therefore, various types of linear or non-linear features from time or frequency domain are extracted from multiple EEG channels (usually 16 to 24 for scalp EEG) to characterize the signal patterns. Such a full-channel analysis is expensive, complex and time-consuming. Besides, not all features are necessarily informative of seizure events. In fact, the redundant and/or irrelevant features from less informative channels are detrimental to classification performances. It is theoretically and practically significant to select seizure-indicative features from a few EEG channels.

In machine learning field, feature selection can be categorized into three categories [10, 11]:

1. Filter Method. Filter method is independent of any learning algorithms and relies only on characteristics of data to assess importance of each feature [10]. Authors in [12] applied the mutual information method to find the best five channels to extract multivariate time-frequency features for seizure detection. In [13], the analysis of variance
method was employed to select 18 discriminative nonlinear features from intracranial EEG data.

2. **Wrapper Method.** Wrapper method generate feature subsets by search algorithms and evaluate feature qualities based on their learning performances of a given classifier. The wrapper methods are computationally intensive compared to filter methods, but can achieve a higher predictive accuracy [14]. Authors in [13] employed the forward sequential selection technique to search for the optimal nonlinear feature subset for seizure detection using intracranial EEG data. In [15], the recursive feature elimination was adopted with support vector machine (SVM) to find 30 seizure-indicative features for patient-independent seizure detection.

3. **Embedded Method.** In this approach, the feature selection process is embedded into model learning procedure. Embedded method is more computational efficient than wrapper methods, since they avoid the repetitive execution of classifier and exhaustive search of every feature subset [16]. Authors in [17] applied a regression tree based feature selection strategy to select spectral power and spectral ratio features. In [18], a random forest (RF) model was adopted to determine the top three channels by evaluating the contributions of features of each channel.

### 1.5 Thesis Organization

This thesis is organized as follows: Chapter 2 presents an approach using F-statistic values for feature ranking and selection. Chapter 3 proposes a two-step feature selection method to choose discriminative feature subsets from a limited number of channels. Chapter 4 summarizes our findings, and highlights some possible improvements and potential directions for future work.
CHAPTER 2

FEATURE SELECTION USING F-STATISTIC VALUES FOR EEG SIGNAL ANALYSIS [1]

2.1 Overview

This Chapter presents an approach to enhance EEG classification performance by selecting discriminative features from a combined feature set consisting of frequency domain and entropy based features. For each EEG channel, nine different features are extracted, including six sub-band spectral powers and three entropy values (sample, permutation and spectral entropy). Features are then ranked across all channels using F-statistic values and selected for SVM classification. Experiments on public-domain dataset show that our method achieves high performances. The selection result of each patient is visualized for personalized analysis.

2.1.1 Motivation

An important issue on EEG analysis and especially seizure detection is to extract informative features from multi-channel EEG signals. Frequency domain features, or spectral power features, have been widely applied in many seizure detection works on either intracranial EEG dataset [17] or scalp EEG dataset [19]. Spectral power features such as absolute band powers, relative band powers in specific bands, or spectral power ratios between band powers are all known to be good features for seizure detection [17]. On the other hand, entropy, which measures the randomness and information complexity of EEG signal, has been less utilized. Different entropy based features, including sample entropy [20], permutation entropy [21], are recently adopted for seizure detection task. Researchers have also combined different

type of entropy features [22, 13] or combined entropy with spectral power features [23] to improve detection performance.

### 2.1.2 Main Contribution

The main contribution of this Chapter is twofold. First, we combine both spectral- and entropy-based features for seizure detection task. Because EEG signals are highly non-stationary and complex, we expect that a joint feature set will enhance classification performance since different features contain various patterns from different domains. Second, since not all features are necessarily discriminative for classification, we use F-statistic value to rank each type of features across all channels and select the most informative features in a patient-specific way.

The rest of this Chapter is organized as follows. The background knowledge on spectral and entropy features are described in Section 2.2. Our proposed methodology are outlined in Section 2.3. Experimental results are reported in Section 2.4. Summarized remarks are made in Section 2.5.

### 2.2 Background on EEG Features

Two main categories of features are used in our work, i.e., power spectral density and entropy based features. All together, they provide nine features as explained in this Section.

#### 2.2.1 Power Spectral Density

Power spectral density (PSD) are widely adopted for seizure detection since spectral features in different frequency bands can discriminate seizure and non-seizure events. For a $L$-sample length signal $X = \{x_1, x_2, \ldots, x_L\}$, the PSD coefficient at frequency $f$ (in Hz) is calculated as:

$$P(f) = \frac{1}{f_s L} \left| \sum_{n=1}^{L} x_n e^{-j2\pi f n} \right|^2 , \ |f| \leq f_s/2$$  \hspace{1cm} (2.1)
Then, for a given frequency band $\omega = [\omega_1, \omega_2]$ where $\omega_1 \leq \omega_2 \leq f_s/2$, the normalized PSD is the sum of PSD coefficients over $\omega$ divided by sum over total frequency range:

$$P(\omega) = \frac{\sum_{f=\omega_1}^{f=\omega_2} P(f)}{\sum_{f=0}^{f=f_s/2} P(f)}$$

(2.2)

Six different frequency bands: $\delta$ (0.1-4 Hz), $\theta$ (4-8 Hz), $\alpha$ (8-12 Hz), $\beta$ (12-30 Hz), $\gamma_1$ (30-57 Hz) and $\gamma_2$ (63-90 Hz), are used in this work.

### 2.2.2 Sample Entropy

Sample entropy (SampEn) is first introduced in [24] to assess the complexity of physiological time series signals and diagnose disease states. SampEn has the advantages of length independence and relative consistency, as it provides a good evaluation of time series regularity and dynamics [24].

Two parameters are used in calculating SampEn: the embedding dimension $m$ and tolerance $r$. Signal $X$ is divided into $L - m + 1$ sample vectors of $m$-dimension by one point shift. Let $x_m(i) = \{x_i, x_{i+1}, \ldots, x_{i+m-1}\}$ and $x_m(j) = \{x_j, x_{j+1}, \ldots, x_{j+m-1}\}$ be the $i$-th and $j$-th sample where $i \neq j$, the Chebyshev Distance $d_m(i,j)$ between $x_m(i)$ and $x_m(j)$ is:

$$d_m(i,j) = \max\{|x_{i+k} - x_{j+k}| \ni 0 \leq k \leq m-1, i \neq j\}$$

(2.3)

which is the maximum absolute difference of their corresponding elements. We define counting number $C_m(i)$ as:

$$C_m(j) = \# \text{ of } x_m(j), \ni d_m(i,j) \leq r$$

(2.4)

Then, the probability of having $C_m(j)$ is:

$$p_m(i) = \frac{1}{L - m} C_m(i)$$

(2.5)

where $L - m$ suggests that self match is not considered. We average $p_m(i)$ for the first $L - m$ sample of dimension $m$:

$$B_m = \frac{1}{L - m} \sum_{i=1}^{L-m} p_m(i)$$

(2.6)
Note that, choosing $L - m$ sample is to ensure both $x_m(i)$ and $x_{m+1}(i)$ can be defined when $1 \leq i \leq L - m$. Increasing $m$ to $m + 1$, we get $B_{m+1}$, and SampEn is defined as:

$$\text{SampEn} = -\log \left( \frac{B_{m+1}}{B_m} \right)$$

Equation (2.7) is always non-negative since $B_{m+1} \leq B_m$. When $X$ reflects strong regularity SampEn approaches to zero. As empirically suggested in [24], we choose $m = 2$ and $r = 0.2 \times \text{Std}$, where Std denotes the standard deviation of original signal $X$.

### 2.2.3 Permutation Entropy

Permutation entropy (PermEn) is first introduced in [25] to calculate the complexity of signals based on comparison of neighboring values. It has advantages of simplicity, extremely fast calculation and robustness with respect to non-linear monotonous transformations.

The definition of embedding dimension $m$ in PermEn is similar to that of SampEn, by which signal $X$ is partitioned into $L - m + 1$ vectors of $m$-dimension. The ordinal pattern $\pi$ is the relative order of permutation $m!$. Using $m = 3$ for instance, $x_m(i)$ may have pattern $\pi_1$ when $x_i < x_{i+1} < x_{i+2}$, or pattern $\pi_6$ when $x_i > x_{i+1} > x_{i+2}$. In cases where several elements are equal, like $x_i = x_{i+1} < x_{i+2}$, the relative order is corresponding to their indices order and the vector matches pattern $\pi_1$. Let $N(j)$ be the number of vectors having $j$-th permutation order $\pi_j$ where $1 \leq j \leq m!$ and $0 \leq N(j) \leq L - m + 1$. The probability of having pattern $\pi_j$ is:

$$p(j) = \frac{N(j)}{L - m + 1}$$

Then, PermEn is defined as:

$$\text{PermEn} = -\sum_{j=1}^{m!} p(j) \log p(j)$$

where $p(j) \log p(j)$ is considered 0 if $p(j) = 0$. When $X$ is monotonously increasing or decreasing, PermEn becomes 0. In this work, we choose $m = 3$ for computation as empirically suggested in [25].
2.2.4 Spectral Entropy

Spectral entropy (SpecEn) measures the degree of signal’s disorder in the frequency distribution of the power spectrum amplitude components of a signal [26]. It can be viewed as the measurement of complexity in spectral domain. SpecEn is incorporated with PSD in aforementioned parts. After getting PSD coefficients $P(f_i)$ at $i$-th frequency using Equation (2.1), the probability that $P(f_i)$ distributes over total frequency range is:

$$ p(i) = \frac{P(f_i)}{\sum_{f=0}^{f_s/2} P(f)} $$

(2.10)

And SpecEn is defined as follows:

$$ \text{SpecEn} = -\sum_{i=0}^{f_s/2} p(i) \log p(i) $$

(2.11)

The greater SpecEn is, the more uniform the spectral power is distributed and more irregular the signal is.

2.3 Proposed Methodology

The flow chart of our proposed method is shown in Fig. 2.1. First, we extract PSD and entropy-based features (ENT) described in Section II from multi-channel input raw EEG signals. Second, we rank features across channels using F-statistic values and select features with high F-statistics. Then, we feed the selected feature subset into a binary support vector machine (SVM) classifier to classify data in seizure or non-seizure classes.

![Figure 2.1. Flowchart of feature selection using F-statistic values](image-url)
2.3.1 Feature Extraction

A ten seconds window with zero overlap is used to divide EEG signal of each channel into several segments. Each segment has sample points $L = 10 \times f_s$, where $f_s$ is sampling frequency. Features are then extracted segment by segment. Following Section II, six sub-band PSDs, SampEn, PermuEn and SpecEn are extracted from 23 channels, and a total number of $9 \times 23 = 207$ features are formed. These are the candidates for our feature selection strategy, as described in the next part.

2.3.2 Feature Selection

The significance of feature selection is twofold. First, it improves classification performance by discarding irrelevant features and alleviating the high dimension problem. Second, by selecting relevant feature set for each patient, we can get a better understanding of seizure patterns for a specific patient. Such personalization allows more accurate seizure analysis (e.g., automatic detection or pre-ictal alert). In this work, we use F-statistic value for feature ranking and selection. F-statistic measures the ratio of between-group and within-group variances. It is defined as \[27\]:

$$F = \frac{\sum_{k=1}^{K} (\bar{y}_k - \bar{y})^2 / (K - 1)}{\sum_{k=1}^{K} \sum_{j=1}^{n_k} (\bar{y}_{kj} - \bar{y}_k)^2 / (N - k)} \quad (2.12)$$

where $n_k$ is the number of samples in $k$-th group (e.g., $k = 1$ for seizure or $k = 2$ for non-seizure), $y_{kj}$ is the $j$-th observation belonging to $k$-th group, $\bar{y}_k$ is the mean of observations in $k$-th group and $\bar{y}$ is the mean of all observations. The numerator and denominator of Equation \[2.12\] calculate the between-group and within-group variance, respectively. A large F-statistic value indicates a significant difference between seizure and non-seizure features.

Let $F^c_i$ denotes the F-statistic value of $i$-th feature from channel $c$. We rank $i$-th feature across all channels and select subsets of $i$-th feature with high $F^c_i$. We discard those features that have low F-statistics and show no significant difference between seizure and non-seizure
classes ($p > 0.05$ chosen in our statistical hypothesis test). Then, we determine a threshold $\tau_i$ by averaging $F_{ci}^{i}$ among the remaining channels $C_i$:

$$\tau_i = \frac{1}{C_i} \sum_{c=1}^{C_i} F_{ci}^{i} \quad (2.13)$$

We then select features whose $F_{ci}^{i}$ are greater than $\tau_i$. This is repeated for all nine features described in Section II.

Equation (2.13) shows that we choose the selection threshold in a feature-specific style. Different features have various discriminative patterns within different channels, and thus, it is hard to predetermine a fixed number of best features. Also, using Equation (2.13), the arithmetic mean will not be affected by neglecting those small F-statistics where no significant difference exists between seizure and non-seizure groups.

### 2.3.3 Classification

For validation, after feature selection, we feed the selected feature subset into radial basis function (RBF) kernel support vector machine (SVM) classifier. Detailed description of SVM algorithm is beyond the scope of this paper and can be found in many sources such as [28].

An important issue in seizure detection task is the extremely high imbalance ratio between non-seizure and seizure recordings. The classification performance will be dominated by non-seizure class. On the other hand, we want to get full use of whole database since various information are included in non-seizure data. In this work, to alleviate the imbalanced data problem, we divide non-seizure recordings into $n$ groups such that each non-seizure data are no more than four hours. A copy of remaining seizure data is then added to each group, and the resultant structure would be $n$ groups containing seizure data and divided non-seizure data. Training and testing are conducted on each group by 10-fold cross validation (10-CV), and final result is the average from all groups. This process is illustrated in Fig. 2.2.

Four metrics are used for performance evaluation: sensitivity ($\text{Sen} = \frac{TP}{TP + FN}$), specificity ($\text{Spe} = \frac{TN}{TN + FP}$), F-1 score ($F-1 = \frac{2TP}{2TP + FP + FN}$) and accuracy
Figure 2.2. Separating and grouping non-seizure data

\[
\text{Acc} = \frac{(TP + TN)}{(TP + FP + TN + FN)},
\]
where TP, FN, TN, FP are the number of true positive, false negative, true negative and false positive in classification.

2.4 Experimental Results

2.4.1 EEG Dataset

In this work, we use a public scalp EEG dataset, the CHB-MIT database containing 23 subjects collected at Boston’s Children’s Hospital [29]. This database reported various number of channels in their recordings files. To keep number of channels consistent during our experiment, we chose fifteen subjects maintaining uniform 23 channels.

2.4.2 Results and Discussion

Table 2.1 lists the classification results and selection results on fifteen different subjects. In terms of classification performance, the average Sen, Spe, F-1 and Acc achieve 92.63%, 99.72%, 91.21% and 99.59%, respectively. In particular, we focus on Sen and F-1 because they indicate successful detection of seizure events. In most cases the Sen and F-1 are above 90%, which implies the effectiveness and robustness of our method. In terms of selection results, on average, 39 out of 138 (6 × 23) PSD features and 22 out of 69 (3 × 23) ENT features are selected, as shown in the last two columns of Table 2.1.
Table 2.1. Results of classification and feature selection

<table>
<thead>
<tr>
<th>Subject</th>
<th>Classification Results (%)</th>
<th>Features Selected (#)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sen</td>
<td>Spe</td>
</tr>
<tr>
<td>chb01</td>
<td>96.40</td>
<td>99.45</td>
</tr>
<tr>
<td>chb02</td>
<td>97.64</td>
<td>99.56</td>
</tr>
<tr>
<td>chb03</td>
<td>93.74</td>
<td>99.68</td>
</tr>
<tr>
<td>chb05</td>
<td>96.29</td>
<td>99.97</td>
</tr>
<tr>
<td>chb06</td>
<td>83.33</td>
<td>99.79</td>
</tr>
<tr>
<td>chb07</td>
<td>96.22</td>
<td>99.77</td>
</tr>
<tr>
<td>chb08</td>
<td>94.50</td>
<td>99.31</td>
</tr>
<tr>
<td>chb10</td>
<td>98.51</td>
<td>100.00</td>
</tr>
<tr>
<td>chb11</td>
<td>95.31</td>
<td>99.65</td>
</tr>
<tr>
<td>chb14</td>
<td>93.33</td>
<td>99.79</td>
</tr>
<tr>
<td>chb18</td>
<td>86.20</td>
<td>100.00</td>
</tr>
<tr>
<td>chb20</td>
<td>90.41</td>
<td>99.28</td>
</tr>
<tr>
<td>chb21</td>
<td>91.66</td>
<td>99.90</td>
</tr>
<tr>
<td>chb22</td>
<td>89.65</td>
<td>99.93</td>
</tr>
<tr>
<td>chb23</td>
<td>86.29</td>
<td>99.79</td>
</tr>
<tr>
<td>Average</td>
<td>92.63</td>
<td>99.72</td>
</tr>
</tbody>
</table>

We use chb01 and chb03 as two examples toillustrate the importance of patient-specific feature selection in Figure 2.3. The horizontal axis of Figure 2.3 denotes 23 channels and vertical axis denotes 9 different features. The selected features within specific channel are marked with color. In chb01, 13 $\beta$-PSD features are selected, but in chb03, only 5 $\beta$-PSD features remained. Also note that, for each subject, different channels have different discriminative features. For chb01, all features from channel #4 are selected, but no feature from channel #1, #2, #13-15 or #22 is chosen.

Table 2.2 compares our results with some prior works that analyzed CHB-MIT database. Our results outperform [18] in all metrics because we rank features across channels and choose them separately, whereas in [18] PSD features are extracted from three pre-determined channels. Compared with [22] in which fuzzy entropy and sample entropy are combined for
Figure 2.3. Selection results for Subject chb01 (a) and Subject chb03 (b). The colored grids denote selected features using F-statistic values (darker color reflects larger F value).

Table 2.2. Comparison with other works

<table>
<thead>
<tr>
<th>Ref.</th>
<th># of Subj.</th>
<th>Classifier</th>
<th>Sen (%)</th>
<th>Spe (%)</th>
<th>F-1 (%)</th>
<th>Acc (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>[18]</td>
<td>23</td>
<td>KNN</td>
<td>90.00</td>
<td>NA</td>
<td>61.50</td>
<td>93.00</td>
</tr>
<tr>
<td>[22]</td>
<td>18</td>
<td>SVM</td>
<td>98.27</td>
<td>98.36</td>
<td>NA</td>
<td>98.31</td>
</tr>
<tr>
<td>[30]</td>
<td>18</td>
<td>LDA+NN</td>
<td>87.50</td>
<td>99.90</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>[31]</td>
<td>7</td>
<td>RF</td>
<td>82.98</td>
<td>99.41</td>
<td>NA</td>
<td>99.36</td>
</tr>
<tr>
<td>Ours</td>
<td>15</td>
<td>SVM</td>
<td>92.63</td>
<td>99.72</td>
<td>91.21</td>
<td>99.59</td>
</tr>
</tbody>
</table>

SVM classification, our method yields higher specificity and accuracy, and we report F-1 to get more comprehensive results. In [30], the authors proposed a patient-nonspecific strategy based on Stationary Wavelet Transform of EEG signal, and our method shows higher sensitivity. In addition, our results outperform [31] in which features are extracted from both continuous EEG (cEEG) and amplitude integrated EEG (aEEG) signals.
2.5 Summary

In this Chapter, we combine frequency domain features and entropy based features for EEG classification and seizure detection. EEG signals are complex and non-stationary, so we extract sample, permutation and spectral entropy values to capture informative characteristics. These entropy features are combined with spectral power features to form a combined feature set. Also, we rank each feature across channels using F-statistic values and select discriminative features for classification. By selecting most informative features of different subjects we not only improve classification results but also provide patient-specific ranking of features.
CHAPTER 3
PERSONALIZED FEATURE SELECTION FOR WEARABLE EEG MONITORING PLATFORM

3.1 Overview

This Chapter presents a feature selection approach to design a wearable EEG monitoring platform for seizure detection. A two-step selection method is proposed to select a few discriminative features from limited number of channels. Step 1 uses linear discriminant analysis to determine a few seizure-relevant channels. Step 2 employs the least absolute shrinkage and selection operator to choose discriminative feature subset. Furthermore, to address the high variation among different patients, a personalization strategy is proposed to find the best number of channels and features of each subject. Experimental results on CHB-MIT database show that by choosing no more than 7 features from 3 channels or less, an average sensitivity, specificity and F-1 score of 87.43%, 99.83% and 89.48% can respectively be achieved.

3.1.1 Motivation

EEG-based seizure monitoring usually relies on a set of features which are extracted from EEG data collected by multiple scalp mounted electrodes (e.g., 16-24). Such a full-channel analysis is expensive, complex and time-consuming, and is only applicable in clinical setting. A more appealing way is to analyze EEG signals by using a small number of features from a limited number of channels. A limited-channel EEG analysis for seizure detection offers


several benefits. First, it enables the design of wearable EEG monitoring platform in daily use. A low-cost and reliable seizure monitoring and alert system can greatly enhance the quality of epileptic patient’s life. Second, it assists clinicians for decision making by offering a personalized EEG pattern analysis for each patient. Note that full-channel EEG data contain both informative and redundant features. Thus, the procedure of selecting informative channels and discriminant features for different subjects can improve the explicity and interpretability of our analysis.

3.1.2 Main Contribution

The main contribution of this Chapter is a personalized scheme to choose limited number of features for each subject from a few carefully-selected channels. A two-step hybrid feature selection strategy is proposed in this work. First, for wearability, a filter method is applied to select no more than five channels by using linear discriminant analysis (LDA) to rank channels. In the second step, the least absolute shrinkage and selection operator (LASSO) is adopted as an embedded technique to select a small subset (e.g., less than ten) of features. Furthermore, to address the high variety among different subjects, a personalization strategy is also proposed to choose the best number of channels and features.

This paper is organized as follows. The background knowledge on LDA and LASSO are provided in Section 3.2. Our proposed methodology are outlined in Section 3.3. Experimental results are reported in Section 3.4. Conclusion is in Section 3.5.

3.2 Background

3.2.1 Linear Discriminant Analysis (LDA)

In EEG classification tasks, different types of features are often extracted from each channel, and are combined to form a feature vector. Therefore, it is important to measure the
discrimination effectiveness of feature vectors. Linear discriminant analysis, abbreviated as LDA, is an important technique which has been applied for either classification [32], dimension reduction [33] or feature selection [34]. Specifically, for feature selection, LDA provides a linear separability criterion to examine the class separability of feature vectors [35, Ch. 5].

Let $C_1$ and $C_2$ represent class 1 (seizure) and class 2 (non-seizure), respectively, and $N_1, N_2$ be the number of instances in these classes. The within-class scatter matrix $S_w$ is computed as:

$$S_w = \sum_{i=1}^{2} p_i \Sigma_i = p_i E \left[ (f - \mu_i)(f - \mu_i)^\top \right] \tag{3.1}$$

where $p_i$ represents the a priori probability of class $C_i$, i.e., $p_i = N_i / N$ for total number of $N = N_1 + N_2$ instances, $\Sigma_i$ is the covariance matrix of feature vectors $f$ in class $C_i$, and $\mu_i$ denotes the mean vector of class $C_i$. The between-class scatter matrix $S_b$ is computed as:

$$S_b = \sum_{i=1}^{2} p_i (\mu_i - \bar{\mu})(\mu_i - \bar{\mu})^\top \tag{3.2}$$

where $\bar{\mu}$ is the global mean vector. Let the mixture scatter matrix $S_m = S_b + S_w$, then, the linear separation criterion is [35]:

$$J = \text{Tr} \left\{ S_w^{-1} S_m \right\} \tag{3.3}$$

where $\text{Tr}\{\cdot\}$ denotes matrix trace, and sometimes $S_m$ can be replaced by $S_b$ [35]. It can be found that Equation (3.3) measures class separability by taking the ratio of mixture- and within-class variances. A larger $J$ indicates higher separability since features are more separated between two classes than they are clustered within each class.

### 3.2.2 Least Absolute Shrinkage and Selection Operator (LASSO)

The least absolute shrinkage and selection operator (LASSO) was first proposed in [36] for estimation in linear regression models. In terms of feature selection, LASSO can be viewed
as an embedded method that forces coefficients of eliminated feature toward zero by adding a regularization term in regression models.

We consider an ordinary linear regression problem, where an $N \times 1$ feature vector $\mathbf{x}_d = [x^d_1, x^d_2, \ldots, x^d_N]^\top$ is the $N$ observed values of $d$-th feature. An $N \times D$ feature matrix $\mathbf{X}$ is formulated as:

$$
\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_D] =
\begin{bmatrix}
  x^1_1 & x^2_1 & \cdots & x^D_1 \\
  x^1_2 & x^2_2 & \cdots & x^D_2 \\
  \vdots & \vdots & \ddots & \vdots \\
  x^1_N & x^2_N & \cdots & x^D_N
\end{bmatrix}_{N \times D}
$$

(3.4)

Note that in this matrix, each row and column represents an instance and a feature, respectively. We have an $N \times 1$ label vector $\mathbf{y} = [y_1, y_2, \ldots, y_N]^\top$, where $y_i \subseteq \{\pm 1\}$. Specifically, $y = 1$ denotes seizure class and $y = -1$ denotes non-seizure class. Let a $D \times 1$ vector $\beta$ be the regression coefficients, the least square loss function using $\ell_2$ norm (Euclidean distance) is written as:

$$
\minimize_{\beta \in \mathbb{R}^D} \left\{ \frac{1}{2N} \| \mathbf{y} - \mathbf{X}\beta \|_2^2 \right\}
$$

(3.5)

Typically, all the $D$ entries of $\beta$ are nonzero. If we want to select a subset of features, we have to keep the coefficients of chosen features be nonzero, but shrink the coefficients of eliminated features into zero. In LASSO method, this is achieved by adding a regularization term $\lambda \| \beta \|_1$ in Equation (3.5), where the nonzero $\lambda$ is the LASSO regularization parameter, and $\| \cdot \|_1$ denotes the $\ell_1$ norm (Manhattan distance). By doing so, we have the standard LASSO equation [37]:

$$
\minimize_{\beta \in \mathbb{R}^D} \left\{ \frac{1}{2N} \| \mathbf{y} - \mathbf{X}\beta \|_2^2 + \lambda \| \beta \|_1 \right\}
$$

(3.6)

As seen in Equation (3.6), the regularization parameter $\lambda$ yields constraints on regression model. The larger $\lambda$ is, the higher constraints it exerts, and more coefficients will be shrunk into zero. The set of features having non-zero coefficients is defined as active set [38], i.e.,
\[ \mathcal{A} = \{d \mid \beta_d \neq 0 ; 1 \leq d \leq D \}. \] An important property of LASSO is the uniqueness of its solutions \[39\], which leads to the uniqueness of active set \( \mathcal{A} \). Therefore, we can incrementally add new features into active set as selection result by gradually decreasing \( \lambda \) values.

### 3.3 Two-Step Feature Selection Methodology

The basic workflow of our proposed method is shown as Figure 3.1. First, we extract two types of features from raw multi-channel EEG signals, i.e., power spectral density (PSD) values and three different entropy values (sample, permutation and spectral entropy). Then, a two-step feature selection strategy is employed to find a small subset of informative features from limited number of channels. In Step 1, we use linear discriminant analysis (LDA) to determine top-\( K \) seizure-relevant channels. In Step 2, we use the least absolute shrinkage and selection operator (LASSO) method to choose \( M \) discriminative features. To further enhance the accuracy for each patient, we employ a personalization method to find the most informative and discriminative \( K^* \) channels and \( M^* \) features for that particular subject based on support vector machine (SVM) evaluation.

#### 3.3.1 Feature Extraction

In this Chapter, two types of features, i.e., PSD values of six frequency bands and three entropy values (SampEn, PermEn and SpecEn), are extracted from one EEG channel, which
have been introduced in Chapter 2. All together, for each channel, a feature vector of
dimension nine is formed, and a total $9 \times C$ feature values from $C$ channels are extracted
per frame for our feature selection method.

### 3.3.2 Two-Step Feature Selection

#### Step 1

For a specific channel $c$, after feature extraction, we follow the procedures of LDA compu-
tation from Equation (3.1) to Equation (3.3), and get discriminant criterion value $J^c$ of this
channel. We rank different $J^c$ in descending order, and pick up the top-$K$ channels having the
highest criterion values. The top-$K$ ranked channel set are denoted as $S^K = \{c_1, c_2, \ldots, c_K\}$
where $J^{c_1} > J^{c_2} > \ldots > J^{c_K}$, and which provide $9 \times K$ features for LASSO subset selection
in the next step.

#### Step 2

After determining the Top-$K$ channels from Step 1, an $N \times 9K$ feature matrix is formulated,
as in Equation (3.4). The LASSO method is applied to gradually select the $M$ features by
tuning different values of $\lambda$. In our work, we use the warm starts strategy in [37, Ch.5] to
generate a sequence of values $\{\lambda_\ell\}_1^\mathcal{L}$ decreasing from $\lambda_1$ down to $\lambda_\mathcal{L}$ on a log scale. The
largest value $\lambda_1$ is determined as [37]:

$$\lambda_1 = \frac{1}{N} \max_d |\langle x_d, y \rangle| \quad (3.7)$$

where $\langle x_d, y \rangle$ denotes the inner product of $d$-th feature vector $x_d$ and label vector $y$, pre-
viously defined in Equation (3.4) and (3.5), respectively. The value of $\lambda_1$ yields an empty
selection set, and by decreasing $\lambda_\ell$ we incrementally add new feature into selection subset.
The sequential number $\mathcal{L}$ is empirically chosen as 100 in our work.
3.3.3 Personalization

In our two-step selection method, the number of utilized channels and selected features ($K$ and $M$, respectively) are not known \textit{apriori}. For some subjects, these parameters may be very low (e.g., 1 channels and 4 features) to achieve a satisfactory performance. For others, the required values of $K$ and $M$ may be higher (e.g., 5 channels and 10 features). Therefore, a personalization procedure is needed to determine the best selection parameter $K^*$ and $M^*$ in a patient-specific style to achieve the highest classification result for that subject.

Let $A^k_m$ be the selection subset containing $m$ features from Top-$k$ channels, where $1 \leq m \leq M$ and $1 \leq k \leq K$. When fixing $k$ channels, the value of $m$ is the selection order, and we have $A^k_m \subseteq A^k_{m+1}$. Then, for different $m$, we feed $A^k_m$ into SVM, and report the corresponding classification scores $F^k_m$. Different performance metrics, like sensitivity, accuracy or F-1 score, can be applied. In this work, we choose F-1 score (harmonic mean of sensitivity and precision, which is effective for imbalanced learning scenarios [40]) as evaluation metric. This means that we only consider positive instances (i.e., seizure samples). This process is repeated for all $k'$s, and we sort $F^k_m$ in descending order to get the best subset $A^*$ that produces the highest $F^*$ together with corresponding $K^*$ and $M^*$. This procedure is summarized in Algorithm 1.

\begin{algorithm}
\caption{Personalizing feature selection parameters}
\begin{algorithmic}[1]
\Statex \textbf{Input:} Feature subset $A^k_m$, $1 \leq k \leq K$, $1 \leq m \leq M$
\Statex \textbf{Output:} Best personalized parameter ($K^*$ and $M^*$) and best subset $A^*$
\For{$k = 1, 2, \ldots, K$} \Do
\For{$m = \min\{9 \times k, M\}$} \Do
\State Feeding $A^k_m$ into SVM classifier
\State Reporting $F^k_m$
\EndDo
\EndDo
\State $\{K^*, M^*\} = \arg\max_{k,m} \{F^k_m\}$
\State $A^* \leftarrow A^{K^*}_{M^*}$
\end{algorithmic}
\end{algorithm}
3.4 Experimental Results

3.4.1 Experimental Data and Settings

In this Chapter, we use a public-domain scalp EEG dataset, the CHB-MIT database for evaluation, which is consistent with previous work in Chapter 2.4.1. The radial basis function (RBF) kernel SVM classifier is chosen for evaluation. The training and testing are performed by 10-fold cross validation (10-CV), and the main evaluation metric is the average F-1 score of ten folds.

EEG data are heavily imbalanced, since the seizure recordings of each subject last 1 to 33 minutes whereas the non-seizure data may exceed to tens of hours. For uniformity of our analysis, for each subject, we consider a close imbalance ratio of each subject’s data by choosing all seizure data together with four hours non-seizure recordings that are close to seizure events.

3.4.2 Channel Ranking

Results on Limited Channel Analysis

To rationalize the efficacy of limited channel analysis, Table 3.1 compares the results of using fused features from top-K channels with full channels. Having nine features from different modalities, i.e., six sub-band PSDs and three different entropy values, results in $9 \times K$ features. The highest F-1 score of each patient is denoted in boldface. For all subjects, on average, using only four channels can yield the highest F-1, whereas using full channels would actually result in a lower performance. For subjects chb04, chb07, chb14, chb17 and chb19, using one channel will yield the best F-1 scores. Only in chb05, the full channel performance is the highest. Intuitively, the irrelevant or redundant features in low-ranked channels tend to bring more ambiguity and noise into the feature space. An improved classification results for most cases can be achieved by using $K = 4$ or fewer channels compared to full channels.
Table 3.1. Comparing Top-\(K\) Channels With full channels using fusion features and SVM classifier

<table>
<thead>
<tr>
<th>Subj.</th>
<th>Seizure (mm:ss)</th>
<th>F-1 Score for Top-(K) Channels (%)</th>
<th>Top-1</th>
<th>Top-2</th>
<th>Top-3</th>
<th>Top-4</th>
<th>Top-5</th>
<th>Full</th>
</tr>
</thead>
<tbody>
<tr>
<td>chb01</td>
<td>07:22</td>
<td></td>
<td>56.20</td>
<td>67.21</td>
<td>81.90</td>
<td>80.76</td>
<td>60.74</td>
<td>55.75</td>
</tr>
<tr>
<td>chb02</td>
<td>02:52</td>
<td></td>
<td>47.02</td>
<td>45.71</td>
<td>49.50</td>
<td>54.58</td>
<td>58.31</td>
<td>38.03</td>
</tr>
<tr>
<td>chb03</td>
<td>06:42</td>
<td></td>
<td>56.14</td>
<td>60.19</td>
<td>76.54</td>
<td>79.01</td>
<td>72.94</td>
<td>52.38</td>
</tr>
<tr>
<td>chb04</td>
<td>06:18</td>
<td></td>
<td>71.16</td>
<td>71.02</td>
<td>65.58</td>
<td>66.98</td>
<td>67.08</td>
<td>68.42</td>
</tr>
<tr>
<td>chb05</td>
<td>09:18</td>
<td></td>
<td>90.54</td>
<td>89.52</td>
<td>84.74</td>
<td>90.09</td>
<td>89.09</td>
<td>94.12</td>
</tr>
<tr>
<td>chb06</td>
<td>02:33</td>
<td></td>
<td>47.61</td>
<td>57.14</td>
<td>62.00</td>
<td>65.71</td>
<td>60.81</td>
<td>45.15</td>
</tr>
<tr>
<td>chb07</td>
<td>05:25</td>
<td></td>
<td>61.38</td>
<td>49.57</td>
<td>48.27</td>
<td>61.22</td>
<td>53.60</td>
<td>54.32</td>
</tr>
<tr>
<td>chb08</td>
<td>15:46</td>
<td></td>
<td>92.87</td>
<td>97.26</td>
<td>83.41</td>
<td>83.00</td>
<td>85.42</td>
<td>79.42</td>
</tr>
<tr>
<td>chb09</td>
<td>04:36</td>
<td></td>
<td>89.28</td>
<td>83.33</td>
<td>90.91</td>
<td>94.54</td>
<td>94.33</td>
<td>80.00</td>
</tr>
<tr>
<td>chb10</td>
<td>07:27</td>
<td></td>
<td>98.76</td>
<td>100.00</td>
<td>100.00</td>
<td>100.00</td>
<td>98.76</td>
<td>89.54</td>
</tr>
<tr>
<td>chb11</td>
<td>13:26</td>
<td></td>
<td>70.31</td>
<td>83.60</td>
<td>75.75</td>
<td>67.52</td>
<td>70.18</td>
<td>64.51</td>
</tr>
<tr>
<td>chb12</td>
<td>16:29</td>
<td></td>
<td>95.23</td>
<td>100.00</td>
<td>100.00</td>
<td>100.00</td>
<td>98.76</td>
<td>76.74</td>
</tr>
<tr>
<td>chb13</td>
<td>08:55</td>
<td></td>
<td>87.75</td>
<td>93.18</td>
<td>87.35</td>
<td>85.39</td>
<td>76.00</td>
<td>68.42</td>
</tr>
<tr>
<td>chb14</td>
<td>02:49</td>
<td></td>
<td>100.00</td>
<td>100.00</td>
<td>96.77</td>
<td>91.79</td>
<td>88.09</td>
<td>77.27</td>
</tr>
<tr>
<td>chb15</td>
<td>33:12</td>
<td></td>
<td>82.45</td>
<td>85.05</td>
<td>80.00</td>
<td>76.26</td>
<td>76.31</td>
<td>71.39</td>
</tr>
<tr>
<td>chb16</td>
<td>01:24</td>
<td></td>
<td>54.28</td>
<td>47.24</td>
<td>54.45</td>
<td>50.40</td>
<td>45.50</td>
<td>38.00</td>
</tr>
<tr>
<td>chb17</td>
<td>04:53</td>
<td></td>
<td>68.57</td>
<td>67.61</td>
<td>60.73</td>
<td>57.73</td>
<td>51.07</td>
<td>49.30</td>
</tr>
<tr>
<td>chb18</td>
<td>05:17</td>
<td></td>
<td>71.23</td>
<td>83.58</td>
<td>83.88</td>
<td>76.82</td>
<td>79.25</td>
<td>77.76</td>
</tr>
<tr>
<td>chb19</td>
<td>03:56</td>
<td></td>
<td>86.27</td>
<td>65.62</td>
<td>72.72</td>
<td>75.41</td>
<td>79.24</td>
<td>66.06</td>
</tr>
<tr>
<td>chb20</td>
<td>04:55</td>
<td></td>
<td>82.61</td>
<td>81.60</td>
<td>84.57</td>
<td>89.51</td>
<td>85.38</td>
<td>78.84</td>
</tr>
<tr>
<td>chb21</td>
<td>03:19</td>
<td></td>
<td>64.28</td>
<td>75.00</td>
<td>71.75</td>
<td>66.77</td>
<td>59.91</td>
<td>53.43</td>
</tr>
<tr>
<td>chb22</td>
<td>03:22</td>
<td></td>
<td>80.00</td>
<td>89.47</td>
<td>74.48</td>
<td>75.50</td>
<td>66.67</td>
<td>63.04</td>
</tr>
<tr>
<td>chb23</td>
<td>07:04</td>
<td></td>
<td>93.02</td>
<td>97.56</td>
<td>96.38</td>
<td>98.76</td>
<td>98.76</td>
<td>95.65</td>
</tr>
<tr>
<td>Avg.</td>
<td>07:43</td>
<td></td>
<td>76.16</td>
<td>76.09</td>
<td>77.46</td>
<td>77.71</td>
<td>74.62</td>
<td>66.94</td>
</tr>
</tbody>
</table>

To illustrate the importance of channel ranking and efficiency of LDA for personalization, we choose subjects chb02 and chb05 as two examples, and show their LDA values as in Figure 3.2. For each subject, the EEG patterns and feature properties are quite different, leading the channel ranking results to be very patient-specific. The top three channels of chb02 are F7-T7, P7-T7 and T7-P7, respectively, whereas for chb05 channels CZ-PZ, P7-O1 and T7-FT9 are the best three. Also note that, the highest LDA value of chb02 is 0.42
in channel F7-T7, much lower than that of chb05 (3.3 in channel CZ-PZ). Because larger LDA values corresponds to higher separability of feature vectors, the F-1 scores of chb02 are overall poorer than those of chb05 (see Table 3.1).

Impact of Feature Fusion

Both PSD and entropy values have been widely adopted in seizure monitoring and detection tasks [18, 41, 42]. To investigate whether the fusion of multi-modal features can improve the performance, Table 3.2 compares the classification results of all 23 subjects using PSD, entropy and both features from Top-1 to Top-5 channels. Here, the Top-K channels refer to those having the highest LDA values calculated by different feature vectors (PSD, ENT or PSD + ENT). For each patient, the discriminant capabilities of PSD and entropy are
different. For instance, for chb05, the ENT values have the highest performances when single channel is used. For the same case, the fused features yield the best result with two channels, and PSDs have the best performances when three channels are exploited. For chb01, with only one channel, the ENT values have the highest F-1, but PSDs yield the best when $K = 2$. On average, the fusion of PSD and ENT features have the strongest indicative capacity of seizure events in all cases. This demonstrates that feature fusion (PSD + ENT) can improve classification performance under limited channels.

Also note that, though different features show various discriminative patterns from multiple modalities, the top-ranked channels are relatively consistent. Figure 3.3 gives an example of LDA channel ranking results for subject chb01 using (a) PSD values, (b) ENT values and (c) PSD + ENT values. Table 3.3 summarizes the top five channels in these three cases, where $k$ in the first column denotes the ranking order (from 1 to 5), and different channels are denoted in boldface. When ENT features are added, channel FZ-CZ becomes one of top-five channels replacing P7-O1 of using only PSDs. However, the other four channels, though different in ranking order, maintain the same. And even in Figure 3.3(b) and (c), the rankings of FZ-CZ and P7-O1 are very close. Both PSD and ENT are informative features, and for some patients, their seizure events are originated within specific locations of the brain. Therefore, the top seizure-indicative channels are consistently identified when using different types of features. This also demonstrates the efficiency of the limited channel analysis.

3.4.3 Subset Selection

LASSO Selection Results

In Step 2, we use LASSO method to select subset $\mathcal{A}_m^k$ from pool of $9 \times K$ features. When $K$ increases, more features are added into candidates pool, and LASSO selection might be different. Using subject chb01 as an example, Table 3.4 list the 10 features from Top-$K$
Table 3.2. Comparing Limited Channel Results using PSD, ENT and both (PSD + ENT)

<table>
<thead>
<tr>
<th>Subj</th>
<th>Top-1</th>
<th>Top-2</th>
<th>Top-3</th>
<th>Top-4</th>
<th>Top-5</th>
</tr>
</thead>
<tbody>
<tr>
<td>chb01</td>
<td>58.13</td>
<td>60.13</td>
<td>56.20</td>
<td>75.92</td>
<td>81.90</td>
</tr>
<tr>
<td>chb02</td>
<td>50.00</td>
<td>46.64</td>
<td>47.02</td>
<td>69.87</td>
<td>77.06</td>
</tr>
<tr>
<td>chb03</td>
<td>64.58</td>
<td>59.16</td>
<td>56.14</td>
<td>51.67</td>
<td>60.74</td>
</tr>
<tr>
<td>chb04</td>
<td>50.04</td>
<td>54.59</td>
<td>71.16</td>
<td>90.26</td>
<td>90.43</td>
</tr>
<tr>
<td>chb05</td>
<td>87.93</td>
<td>90.59</td>
<td>91.58</td>
<td>87.75</td>
<td>84.61</td>
</tr>
<tr>
<td>chb06</td>
<td>68.96</td>
<td>50.61</td>
<td>47.61</td>
<td>50.94</td>
<td>62.92</td>
</tr>
<tr>
<td>chb07</td>
<td>48.64</td>
<td>70.58</td>
<td>61.38</td>
<td>47.70</td>
<td>75.60</td>
</tr>
<tr>
<td>chb08</td>
<td>67.83</td>
<td>94.68</td>
<td>92.87</td>
<td>78.30</td>
<td>97.29</td>
</tr>
<tr>
<td>chb09</td>
<td>89.28</td>
<td>78.68</td>
<td>90.28</td>
<td>88.13</td>
<td>78.79</td>
</tr>
<tr>
<td>chb10</td>
<td>100.00</td>
<td>90.47</td>
<td>98.76</td>
<td>100.00</td>
<td>81.79</td>
</tr>
<tr>
<td>chb11</td>
<td>76.32</td>
<td>84.32</td>
<td>70.31</td>
<td>71.09</td>
<td>79.20</td>
</tr>
<tr>
<td>chb12</td>
<td>88.61</td>
<td>87.24</td>
<td>95.23</td>
<td>95.23</td>
<td>89.59</td>
</tr>
<tr>
<td>chb13</td>
<td>86.86</td>
<td>74.07</td>
<td>87.75</td>
<td>79.02</td>
<td>83.33</td>
</tr>
<tr>
<td>chb14</td>
<td>80.00</td>
<td>96.65</td>
<td>100.00</td>
<td>96.77</td>
<td>96.77</td>
</tr>
<tr>
<td>chb15</td>
<td>82.96</td>
<td>78.00</td>
<td>82.45</td>
<td>78.96</td>
<td>72.91</td>
</tr>
<tr>
<td>chb16</td>
<td>50.40</td>
<td>40.00</td>
<td>54.28</td>
<td>50.53</td>
<td>43.24</td>
</tr>
<tr>
<td>chb17</td>
<td>79.43</td>
<td>63.82</td>
<td>68.57</td>
<td>78.10</td>
<td>73.14</td>
</tr>
<tr>
<td>chb18</td>
<td>69.39</td>
<td>53.92</td>
<td>71.23</td>
<td>69.21</td>
<td>52.28</td>
</tr>
<tr>
<td>chb19</td>
<td>65.45</td>
<td>57.27</td>
<td>86.27</td>
<td>79.54</td>
<td>83.01</td>
</tr>
<tr>
<td>chb20</td>
<td>55.56</td>
<td>66.66</td>
<td>82.61</td>
<td>75.14</td>
<td>69.69</td>
</tr>
<tr>
<td>chb21</td>
<td>61.58</td>
<td>76.00</td>
<td>64.28</td>
<td>76.66</td>
<td>75.00</td>
</tr>
<tr>
<td>chb22</td>
<td>91.89</td>
<td>79.06</td>
<td>80.00</td>
<td>89.47</td>
<td>86.48</td>
</tr>
<tr>
<td>chb23</td>
<td>81.63</td>
<td>93.02</td>
<td>93.02</td>
<td>90.90</td>
<td>97.56</td>
</tr>
</tbody>
</table>

Avg.  71.97  71.57  76.16  74.97  75.87  76.09  71.45  69.68  77.46  71.59  71.72  77.71  71.30  68.49  74.62
channels when $K = 3$ and 4. In both cases, seven out of ten selected features are the same. The three uniquely selected features are denoted in boldface. Such homogeneous selection from different channels is that, for some patients, their seizure patterns are strongly reflected by specific features, either from single frequency band or from spectral signal complexity.

To illustrate the discriminant capacity of LASSO’s selection, we choose seven commonly selected features in Table 3.4 and use t-SNE method [43] to visualize the original feature space of seven dimension in a new 2D space. As a nonlinear data embedding technique, t-SNE visualizes the high dimensional data by preserving the distance and distribution relationships of original space in the translated low dimension (e.g., 2D) [43]. Figure 3.4 shows the 2D scatter plot by t-SNE method, where the two axes are t-SNE components, and each point corresponds to a feature sample calculated from a 10-second windowed EEG segment. Visually, the distinguishable points in t-SNE plot correspond to well-separated seizure and
Table 3.4. Ten features selected by LASSO for chb01 for $K = 3$ and 4

<table>
<thead>
<tr>
<th>$m$</th>
<th>$K = 3$</th>
<th>$K = 4$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Features</td>
<td>Channels</td>
</tr>
<tr>
<td>1</td>
<td>$\theta$-PSD</td>
<td>CZ-PZ</td>
</tr>
<tr>
<td>2</td>
<td>SpecEn</td>
<td>CZ-PZ</td>
</tr>
<tr>
<td>3</td>
<td>$\theta$-PSD</td>
<td>P3-O1</td>
</tr>
<tr>
<td>4</td>
<td>$\alpha$-PSD</td>
<td>P3-O1</td>
</tr>
<tr>
<td>5</td>
<td>SpecEn</td>
<td>P3-O1</td>
</tr>
<tr>
<td>6</td>
<td>$\alpha$-PSD</td>
<td>C4-P4</td>
</tr>
<tr>
<td>7</td>
<td>SpecEn</td>
<td>C4-P4</td>
</tr>
<tr>
<td>8</td>
<td>$\theta$-PSD</td>
<td>C4-P4</td>
</tr>
<tr>
<td>9</td>
<td>$\gamma_2$-PSD</td>
<td>CZ-PZ</td>
</tr>
<tr>
<td>10</td>
<td>$\gamma_2$-PSD</td>
<td>P3-O1</td>
</tr>
</tbody>
</table>

non-seizure instances, which imply that the selected features are discriminative. The dominant non-seizure data are loosely interspersed among 2D space, but are still well separated from seizure data clustered in the left bottom. This demonstrates that the selected feature subsets are quite discriminatory for our classification.

**Impact of Subset Selection**

To evaluate the effectiveness of subset selection, we choose six subjects (chb01 to chb03, chb06, chb08 and chb17) whose best F-1 score is yielded with $K \leq 5$ channels. For these subjects, we compare their classification results using 10 LASSO-selected features from their respective Top-$K$ channels with all features ($9 \times K$). As shown in Table 3.5, on average, the LASSO selection yield higher F-1 scores than unselected features. Such improvement is achieved for 5 out of 6 subjects. Only for chb08, the LASSO selection result is slightly lower than that of full channels (96.73% against 97.26%).

Note that, even in top-ranked channels, there are still some features that are less discriminatory. We chose subject chb17 as an example. Figure 3.5(a) and (b) show the boxplot of $\alpha$
Figure 3.4. T-SNE visualization of LASSO selected subset for subject chb01.

Table 3.5. Classification of LASSO’s Selected Features and Full Features

<table>
<thead>
<tr>
<th>Subj.</th>
<th>K</th>
<th>F-1 (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>LASSO</td>
<td>All</td>
</tr>
<tr>
<td>chb01</td>
<td>3</td>
<td>86.41</td>
</tr>
<tr>
<td>chb02</td>
<td>5</td>
<td>64.00</td>
</tr>
<tr>
<td>chb03</td>
<td>4</td>
<td>83.54</td>
</tr>
<tr>
<td>chb06</td>
<td>4</td>
<td>85.71</td>
</tr>
<tr>
<td>chb08</td>
<td>2</td>
<td>96.73</td>
</tr>
<tr>
<td>chb17</td>
<td>2</td>
<td>84.00</td>
</tr>
<tr>
<td>Avg.</td>
<td>3.6</td>
<td>83.27</td>
</tr>
</tbody>
</table>

PSD from the corresponding Top-2 channels (T8-P8 and P8-O2). The $\alpha$ values have similar distributions between non-seizure and seizure classes. Using one-way analysis of variance for statistical test [27], and setting $p = 0.05$ as significance level, we find that the $\alpha$ PSDs of both channels do not show significant difference between non-seizure and seizure groups ($p = 0.64$ and 0.49, respectively). For PermEn, as shown in Figure 3.5(c) and (d), their values of seizure group are significantly higher than that of non-seizure group ($p < 0.01$), which
Figure 3.5. Boxplot of feature values of subject chb17: (a) $\alpha$ PSD in T8-P8, (b) $\alpha$ PSD in P8-O2, (c) PermEn in T8-P8 and (d) PermEn in P8-O2.

reflects stronger nonlinearity during seizure periods. Therefore, by using LASSO method for subset selection, the irrelevant features are discarded and the classification results are enhanced (see Table 3.5).

3.4.4 Personalization

Best $K^*$ and $M^*$

In personalization stage, we aim at finding the best selection parameters (i.e., $K^*$ channels and $M^*$ features) for each patient. Table 3.6 tabulates the values of $K^*$ and $M^*$ for each subject together with their classification results. For all subjects, on average, using no more than $\lceil2.34\rceil = 3$ channels and $\lceil6.78\rceil = 7$ features will yield $F-1=89.48\%$ and $Sen=87.43\%$, respectively. In terms of selection results, only subject chb07 requires $K^* = 5$ channels, and three subjects (chb06, chb18 and chb21) need $M^* = 10$ features. For classification results, in most cases, both $F-1$ and $Sen$ are higher than $80\%$. For subjects chb10, chb12,
chb14 and chb23, a 100% result of all metrics is achieved. Only for chb02 and chb16 whose seizure recordings are relatively short (see Table 3.1), the performance is relatively lower (F-1=69.23% and 68.88%, Sen=52.94% and 59.89%, respectively). High classification performance is important because in practical seizure monitoring, we need: a) a small number of channels ($K^* \leq 5$) for wearability, b) a small subset of discriminative features ($M^* \leq 10$) for real-time processing on a mobile device, and c) a reliable detection performance of each subject.

We choose subject chb10 as an example for illustration. Figure 3.6(a) shows a raw EEG waveform of Top-1 channel F7-T7 recorded during 6788-7058 seconds. In this case, the seizure starts at 6888 second (red line) and ends at 6958 second (black line), respectively. When seizure kicks in, the amplitude of raw EEG wave gradually increases. Four features, $\alpha$-PSD, $\beta$-PSD, SampEn and SpecEn, are extracted from 27 epochs divided by 10-second non-overlapping window. These feature values are shown in Figure 3.6(b)-(e), respectively, and the non-seizure and seizure epochs are distinguished by blue and red points. For $\alpha$ and $\beta$ PSDs, their values increase during seizure period. For sample and spectral entropy values, their increased values reflect stronger nonlinearity and irregularity either in time or spectral domain. It is expected that some features for frames close to seizure start and end may still carry seizure-like patterns. Overall, the fluctuations of feature values for seizure epochs reflect their discriminative abilities.

Impact of Personalization

Our proposed method is patient-specific, and therefore the final results are expectedly subject-dependent. To clarify the impact of personalization, we chose subjects chb02 and chb03, both having $K^* = 3$ and $M^* = 8$ as in Table 3.6, and list their respective selected channels and features in Table 3.7. For chb02, 6 of 8 selected features are entropy values (i.e., three SpecEn, two SampEn and one PermEn). On the other hand, the selection
Table 3.6. Best Selection and Classification Results

<table>
<thead>
<tr>
<th>Subj.</th>
<th>Personalization (#)</th>
<th>Classification Metrics (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$K^*$</td>
<td>$M^*$</td>
</tr>
<tr>
<td>chb01</td>
<td>2</td>
<td>8</td>
</tr>
<tr>
<td>chb02</td>
<td>3</td>
<td>8</td>
</tr>
<tr>
<td>chb03</td>
<td>3</td>
<td>8</td>
</tr>
<tr>
<td>chb04</td>
<td>2</td>
<td>9</td>
</tr>
<tr>
<td>chb05</td>
<td>4</td>
<td>9</td>
</tr>
<tr>
<td>chb06</td>
<td>3</td>
<td>10</td>
</tr>
<tr>
<td>chb07</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>chb08</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>chb09</td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>chb10</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>chb11</td>
<td>3</td>
<td>8</td>
</tr>
<tr>
<td>chb12</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>chb13</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>chb14</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>chb15</td>
<td>2</td>
<td>9</td>
</tr>
<tr>
<td>chb16</td>
<td>3</td>
<td>7</td>
</tr>
<tr>
<td>chb17</td>
<td>2</td>
<td>8</td>
</tr>
<tr>
<td>chb18</td>
<td>2</td>
<td>10</td>
</tr>
<tr>
<td>chb19</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>chb20</td>
<td>4</td>
<td>7</td>
</tr>
<tr>
<td>chb21</td>
<td>2</td>
<td>10</td>
</tr>
<tr>
<td>chb22</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>chb23</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Avg.</td>
<td>2.34</td>
<td>6.78</td>
</tr>
</tbody>
</table>
Figure 3.6. Best features for seizure detection of subject chb10: (a) Raw EEG waveform of seizure event in channel F7-T7, 10 second epoch of (b) $\alpha$-PSD, (c) $\beta$-PSD, (d) SampEn and (e) SpecEn.
results of chb03 mostly consist of PSDs, and only one ENT value is chosen. This again illustrates the necessity of personalization, since for different subjects, their seizure patterns are characterized by different types of features.

Table 3.7. Comparing Selected Features for chb02 and chb03

<table>
<thead>
<tr>
<th>m</th>
<th>Features</th>
<th>Channels</th>
<th>Features</th>
<th>Channels</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>SpecEn</td>
<td>F7-T7</td>
<td>γ1 PSD</td>
<td>FT10-T8</td>
</tr>
<tr>
<td>2</td>
<td>SpecEn</td>
<td>P7-T7</td>
<td>γ1 PSD</td>
<td>P8-O2</td>
</tr>
<tr>
<td>3</td>
<td>SpecEn</td>
<td>T7-P7</td>
<td>β PSD</td>
<td>P8-O2</td>
</tr>
<tr>
<td>4</td>
<td>SampEn</td>
<td>F7-T7</td>
<td>β PSD</td>
<td>FT10-T8</td>
</tr>
<tr>
<td>5</td>
<td>SampEn</td>
<td>P7-T7</td>
<td>γ2 PSD</td>
<td>P8-O2</td>
</tr>
<tr>
<td>6</td>
<td>θ PSD</td>
<td>T7-P7</td>
<td>θ PSD</td>
<td>T8-P8-0</td>
</tr>
<tr>
<td>7</td>
<td>PermEn</td>
<td>F7-T7</td>
<td>γ2 PSD</td>
<td>T8-P8-0</td>
</tr>
</tbody>
</table>

Note that, for a specific subject, the personalization procedure also enhances the classification performances. For example, the Top-1 channel classification result for subject chb04 using \(9 \times 1 = 9\) features is 71.16\%, which is the highest performance in Table 3.1. However, after personalization stage, the selected 9 features from Top-2 channels yield higher F-1 of 73.01\%. This means that by adding only one channel, we can select more discriminative features while maintaining the same feature size. The validity of limited channels and features has been verified by the number of utilized channels (≤ 5) and selected features (≤ 10). In summary, we strongly advocate such personalization strategy to achieve both wearability and an enhanced classification performance.

3.4.5 Run-Time Analysis

To evaluate the feasibility of real-time wearable EEG monitoring, we measure the run-time of our proposed method for three subjects (chb01, chb02 and chb03). The utilized seizure
data of each subject is 440, 172 and 558 seconds, respectively, resulting in 44, 17 and 55 seizure frames of 10 second length. The non-seizure data of each subject is uniformly 4 hours equivalent to 1440 frames. The main computational procedure consists of three parts: feature extraction (PSD and ENT), feature selection (LDA and LASSO) and SVM classification (training and testing). The average run-time (wall-clock) of each part conducted on our laptop platform (2.3GHz CPU with 8GB RAM) is tabulated in Table 3.8. For feature extraction, the entropy calculation takes longer time than PSD since entropy involves probability estimation whereas for PSD only FFT is required. On average, for 23 full channels, the mean time of calculating 9 features per frame is $29.7 + 376.7 = 406.4$ msec per frame. For feature selection, LDA method requires only the covariance calculation to get scatter matrix. The LASSO subset selection, which is based on linear regression, has efficient numerical solutions with optimized implementations [41, 28]. As for classifier, the SVM has well-established theory and solution even for nonlinear RBF kernels, and the training and testing process of 10-fold is conducted around 19.58 and 4.12 seconds, respectively. In continuous EEG monitoring, after getting a 10-second frame of EEG, we expect all the aforementioned analysis (4.12 sec) done before the next frame comes (10 sec).

<table>
<thead>
<tr>
<th>Computation</th>
<th>Total Time (Sec.)</th>
<th>Mean Time (mSec./frame)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Feature Extraction</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PSD</td>
<td>46.12</td>
<td>29.7</td>
</tr>
<tr>
<td>ENT</td>
<td>559.48</td>
<td>376.7</td>
</tr>
<tr>
<td>Feature Selection</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LDA</td>
<td>16.39</td>
<td>11.4</td>
</tr>
<tr>
<td>LASSO</td>
<td>5.64</td>
<td>3.9</td>
</tr>
<tr>
<td>SVM Classification</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Training†</td>
<td>19.58</td>
<td>13.6</td>
</tr>
<tr>
<td>Testing</td>
<td>4.12</td>
<td>2.9</td>
</tr>
</tbody>
</table>

† Offline
3.4.6 Comparison

To evaluate the consistency of our feature selection and personalization method, we employed two widely-adopted classifiers, the k-nearest neighbors (KNN) and random forest (RF). Then, we evaluated the classification performances using the personalized best feature subset $\mathcal{A}^*$ for ten subjects (chb01 to chb10). Figure 3.7 compares the barplots of F-1, Sen, Spe and Acc using three classifiers, where in each case the classification score is expressed as mean ± std. The results are quite consistent among three classifiers. In terms of F-1 and Sen, RF shows slightly higher scores, whereas KNN has relatively lower performances, but the overall results are close to each other. We argue that the final classification scores expectedly vary among different classifiers. The fine tuning of parameters for each classifier is beyond the scope of this work. The intention here is to demonstrate the consistency and effectiveness of our feature selection and personalization results independent of classifier.

We compared our work with other papers reported on CHB-MIT dataset, and tabulated the corresponding selection methods and classification results in Table 3.9. Authors in [18] used RF to select three top-contributed channels, and applied t-SNE method to reduce the original feature dimension from 15 to 3. However, such dimension reduction would inevitably cause information loss during transformation, which caused relatively lower classification performances. In [22], the Kolmogorov-Smirnov (K-S) test was employed to filter out fuzzy entropy (FuzzENT) features that were significantly different, but the required number of channels remained high. Researchers in [41] employed the minimum redundancy maximum relevance (mRMR) method to select the optimal feature from multiple-domain feature pools, and achieved accuracy of up to 98.33% using one-dimensional feature. However, only three subjects had been tested. For these three works, none of them had addressed the issue of personalized feature selection.
Figure 3.7. Barplot of classification results of ten subjects using SVM, KNN and RF based on average (a) F-1 score, (b) Sensitivity, (c) Specificity and (d) Accuracy. In each case the score is expressed as mean ± std.

Table 3.9. Comparison with other works on CHB-MIT Database

<table>
<thead>
<tr>
<th>Ref.</th>
<th># of Subj.</th>
<th>Classifier</th>
<th>Main Feature</th>
<th>Selection Method</th>
<th>Metrics (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>F-1  Sen  Spe  Acc</td>
</tr>
<tr>
<td>[18]</td>
<td>23</td>
<td>KNN</td>
<td>PSD</td>
<td>RF+t-SNE</td>
<td>61.50 90.00 NA 93.00</td>
</tr>
<tr>
<td>[22]</td>
<td>18</td>
<td>SVM</td>
<td>FuzzENT</td>
<td>K-S test</td>
<td>NA 98.27 98.36 98.31</td>
</tr>
<tr>
<td>[41]</td>
<td>3</td>
<td>SVM</td>
<td>Multiple</td>
<td>mRMR</td>
<td>NA NA NA 98.33</td>
</tr>
<tr>
<td>This work</td>
<td>23</td>
<td>SVM</td>
<td>PSD+ENT</td>
<td>LDA+LASSO</td>
<td>89.48 87.43 99.83 99.40</td>
</tr>
</tbody>
</table>

3.5 Summary

In this Chapter, we advocate a personalized feature selection method to choose the best number of $K$ channels and $M$ features for each patient. Instead of globally searching the best
feature subset, we divide our selection procedure into two steps. The complexity of our two-step selection approach is $O(K) + O(M)$, whereas the exhaustive searching requires $O(2^{KM})$. When high number of EEG channels are involved, e.g., 30-90 electrodes for intracranial EEG (iEEG) recordings [45], our two-step method is computationally efficient. Also note that, for wearability, we bound $K \leq 5$ and $M \leq 10$, which makes our personalization process to be computationally reasonable. Our technique is specifically effective for patients who have specific type of seizures (e.g., focal seizure) that are usually located in a few areas of the brain.
CHAPTER 4
CONCLUSION AND FUTURE WORK

4.1 Conclusion

In this thesis, we proposed two feature selection approaches for personalized seizure monitoring and classification:

1. To improve seizure classification performances, we used F-statistic values to rank the discriminatory ability of PSD and entropy features. Different subjects have their individual seizure patterns, which lead the selection results to be subject-dependent. The filter-based F-statistic ranking scheme is computationally efficient for selecting discriminative values from large set of multi-domain features.

2. To design a wearable seizure monitoring platform, we presented a two-step feature selection method to select a few features from limited number of channels. Seizure symptoms may be reflected by specific types of features from a few channel locations, which is the reason for the effectiveness of our limited-channel analysis and subset selection. This two-step strategy is efficient when multiple channels (e.g., hundreds of channels in iEEG) are involved, since we can 1) determine relevant channels and 2) choose informative features from these channels.

Both approaches had been personalized to find the best selection results of each patient. Bio-signal patterns vary among different patients, and therefore our learning process should be patient-specific. In future, when other types of biosignals are involved, such as electrocardiogram (ECG) or magnetoencephalogram (MEG), this personalization strategy is important to yield the best learning performances of each patient.
4.2 Future Work

There are several challenges that remain to be addressed for future improvements:

1. Imbalance data: In the two studies, seizure data is much less than non-seizure part. Dealing with imbalanced EEG data effectively is itself challenging. Numerous information are contained in non-seizure recordings, and by learning from non-seizure data we can study the evolution and progress of the EEG patterns. Machine learning approaches, like down-sampling strategy, data clustering, may help us address the imbalance issue. Besides, the non-seizure data include various information about brain activities, and learning from the majority class may provide a better insight to EEG data [46].

2. Poor signal qualities of raw EEG data: In these two works, we do not apply any pre-processing method for signal enhancement. Raw EEG recordings are contaminated by motion artifacts and noise. In practice, algorithms for signal enhancement and noise suppression are required to provide high-quality data.

3. Prediction of seizure events: This thesis focuses on seizure detection tasks. In real scenarios, a timely alert of pre-ictal period (e.g., 103 minute before seizure start) is essential for patients and care-givers. Different stages of seizure events need to be analyzed, including pre-ictal, ictal, inter-ictal and post-ictal phases. A multi-class classification model will be established for this task.

Intracranial EEG (iEEG) has been recently investigated for seizure monitoring and analysis. Intracranial data provide new approaches to study seizure origins, to predict seizure onset, and to evaluate effectiveness of seizure therapies. The correlation study on both scalp and intracranial data is noteworthy, since the varying properties of iEEG signal inside the brain may also be reflected in scalp recordings. In future, we plan to collaborate with medical
institutions, and collect our own neurosignal database containing both EEG (scalp and intracranial) and non-EEG signal (e.g., heart rate) for correlation study.
REFERENCES


BIOGRAPHICAL SKETCH

Genchang Peng started his Master of Science program in electrical engineering in Fall of 2018, concentrated in signal and system track. Before entering UT Dallas, he received his B.Eng of Automation and M.Eng of System Engineering in 2015 and 2018, respectively, all from Huazhong University of Science and Technology, China. His main interests are signal processing, machine learning, statistical analysis, control system engineering, with special emphasis on medical applications and healthcare technologies. His master’s research, supervised by Dr. Mehrdad Nourani, concentrates on EEG data analysis and feature selection for epileptic seizure monitoring. He will continue his PhD program in Electrical Engineering in Spring of 2021.
Curriculum Vitae

Genchang Peng
November 4, 2020

Contact Information:
Predictive Analytics and Technologies Lab
Department of Electrical and Computer Engineering
The University of Texas at Dallas, Richardson, TX 75080-3021, U.S.A.
E-Mail: gxp170004@utdallas.edu

Educational History:
B.Eng., Automation, Cornell University, Huazhong University of Science and Technology, Wuhan, China, 2015
M.Eng., System Engineering, Huazhong University of Science and Technology, Wuhan, China, 2018
M.Sc., Electrical Engineering, University of Texas at Dallas, Richardson, TX, USA, 2020
Feature Selection for Personalized EEG-Based Seizure Monitoring
M.Sc. Thesis
Department of Electrical and Computer Engineering, University of Texas at Dallas
Advisor: Dr. Mehrdad Nourani

Publications:

Journal

Conference


Honors:
Master’s Research Fellowship Program, UTD, 2020
Pathway’s to Research Scholarship, ECE Dept., UTD, 2019