Automatic Sleep Arousal Detection based on C-ELM and MRMR feature selection

by

Yuemeng Liang

B.E., Zhejiang University, 2012

A THESIS SUBMITTED IN PARTIAL FULFILMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF APPLIED SCIENCE

in

THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES

(Electrical and Computer Engineering)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

July 2015

Abstract

Sleep arousals are sudden awakenings from sleep which can be identified as an abrupt shift in EEG frequency and can be manually scored from various physiological signals by sleep experts. Frequent sleep arousals can degrade sleep quality, result in sleep fragmentation and lead to daytime sleepiness. Visual inspection of arousal events from PSG recordings is cumbersome, and manual scoring results can vary widely among different expert scorers. The main goal of this project is to design and evaluate the performance of an effective and efficient algorithm to automatically detect sleep arousals using a single channel EEG.

In the first part of the thesis, a detection model based on a Curious Extreme Learning Machine (C-ELM) using a set of 22 features is proposed. The performance was evaluated using the term Area Under the Receiver Operating Characteristic (ROC) Curve (AUC) and the Accuracy (ACC). The proposed C-ELM based model achieved an average AUC and ACC of 0.85 and 0.79 respectively. In comparison, the average AUC and ACC of a Support Vector Machine (SVM) based model were 0.69 and 0.67 respectively. This indicates that the proposed C-ELM based model works well for the sleep arousal detection problem.

In the second part of the thesis, an improved detection model is proposed by adding

a Minimum Redundancy Maximum Relevance (MRMR) feature selection into the C-ELM based model proposed in the first part. The efficiency of the model is improved by reducing dimensionality (reducing the number of features) of the dataset while the performance is largely unaffected. The achieved average AUC and ACC were 0.85 and 0.80 when a reduced set of 6 features were used, while the AUC and ACC were 0.86 and 0.79 for a full set of 22 features. The result indicates MRMR feature selection step is important for sleep arousal detection. By using the improved sleep arousal detection model, the system runs faster and achieves a good performance for the dataset utilized in our study.

Preface

I hereby declare that I am the author of this thesis. This thesis is an original, unpublished work under the supervision of Professor Cyril Leung. This work was supported in part by the Natural Sciences and Engineering Research Council (NSERC) of Canada under Grant RGPIN 1731-2013, and by the UBC Faculty of Applied Science.

Table of Contents

List of Tables

List of Figures

List of Symbols

Symbol	Description
$\mathcal{C}(\mathbf{x}^t)$	Conflict of an input vector $\mathbf{x}^t$
$\bar{f}$	Mean frequency
f_i	Center frequency of the band
$\boldsymbol{I}(h,i)$	The mutual information between feature i and classes h, $h = \{h_1, h_2, ..., h_K\}$
$\boldsymbol{I}(i,j)$	The mutual information between two different features i and j
$\max \boldsymbol{V_I}$	Maximum relevance condition
$\min \boldsymbol{W_I}$	Minimum redundancy condition
$\mathcal{N}(\mathbf{x}^t)$	Novelty of an input vector $\mathbf{x}^t$
p_i	The corresponding power of center frequency of the band
$|\boldsymbol{S}|$	The number of features in the selected feature set $\boldsymbol{S}$
$\mathcal{S}(\mathbf{x}^t)$	Surprise of an input vector $\mathbf{x}^t$

$\mathcal{U}(\mathbf{x}^t)$ Uncertainty of an input vector $\mathbf{x}^t$

$\mathbf{x}^t$ The tth M-dimensional input vector (or feature vector) of the training data $\{(\mathbf{x}^1, c_1), \cdots, (\mathbf{x}^t, c_t), \cdots\}$, $\mathbf{x}^t = [x_1^t, \cdots, x_M^t]^T \in \Re^M$

$\theta_\mathcal{C}$ Initialized neuron deletion threshold for conflict

$\theta_{\mathcal{N}^{add}}$ Initialized neuron addition threshold for novelty

$\theta_{\mathcal{N}^{del}}$ Initialized neuron deletion threshold for novelty

$\theta_\mathcal{S}$ Initialized neuron addition or deletion threshold for surprise

$\theta_\mathcal{U}$ Initialized neuron addition threshold for uncertainty

List of Acronyms

ACC   Accuracy

ANN   Artificial Neural Network

ASDA   American Sleep Disorders Association

AUC   Area Under the Curve

C-ELM   Curious Extreme Learning Machine

ECG   Electrocardiogram

EDS   Excessive Daytime Sleepiness

EEG   Electroencephalogram

EMG   Electromyography

EOG   Electroocculography

FT   Fourier Transform

FNR   False Negative Rate

FPR   False Positive Rate

FFT   Fast Fourier Transform

LIBSVM   Library of Support Vector Machine

MID   Mutual Information Difference

MIQ   Mutual Information Quotient

MLP   Multilayer Perceptron

MRMR   Minimum Redundancy-Maximum Relevance

NREM   Non Rapid Eye Movement

OSA   Obstructive Sleep Apnea

PAT   Peripheral Arterial Tone test

PAT-AAI   Peripheral Arterial Tone test based Autonomic Arousal In-
                       dices

PD   Parkinson's Disease

PPV   Positive Predictive Value

PSG   Polysomnography

PSQI   the Pittsburgh Sleep Quality Index

RBF   Radial Basis Function

ROC   Receiver Operating Characteristic

SAS   Sleep Apnea Syndrome

SBS   Sequential Backward Selection

SFS   Sequential Forward Selection

SHHS   Sleep Heart Health Study

SLFN   Single hidden Layer Feedforward neural Network

SVM   Support Vector Machine

Acknowledgements

Foremost, I would like to express my deepest gratitude to my research supervisor, Prof. Cyril Leung, for his encouragement, guidance and support during my graduate studies and research. I appreciate his patience, enthusiasm, and motivation. Without his help, this thesis would not have been possible.

I am thankful to my co-supervisors, Prof. Martin J. McKeown and Prof. Chunyan Miao (Nanyang Technological University) and their research group members, who have provided guidance and assistance during my research, especially in discussing the research idea and formulating the problem. I would also like to thank Dr. Qiong Wu (NTU) for her suggestions on my work and the source code for C-ELM.

Special thanks to Liang Zou for his expertise in machine learning and his precious assistance in my research. I also wish to thank Xinxin Zhang for his help in proofreading my thesis and discussions on mathematics. My sincere thanks also go to colleagues in our research group for sharing their knowledge and to all my friends who encouraged and supported me during my graduate studies.

of Applied Science.

Last but not least, I would like to express my gratitude to my beloved parents for their encouragement, support and greatest love of all. To them, I dedicate this thesis.

Chapter 1

Introduction

This chapter begins with an introduction to sleep arousals. In the first section, some background on sleep arousals including what the sleep arousals are, why detection of sleep arousals is important and the scoring rules and detection methods are discussed. The second section is a literature survey of previous works on this problem. The motivation and contributions of this thesis are then discussed. The organization of the thesis is outlined in the last part of the chapter.

1.1   Sleep Arousal Detection

Sleep problems are a frequent complaint among many people, especially the elderly, and have a substantial impact on quality of their lives. Sleep arousal conventionally refers to a temporary intrusion of wakefulness into sleep or at least a sudden transient elevation of the vigilance level due to arousal stimuli or to spontaneous vigilance level oscillations [5, 6]. Sleep arousals can be induced by various sleep disorders. Thus, arousals are a good marker of sleep disruption representing a detrimental and harmful feature for sleep [5].

1.1.1 What Is Sleep Arousal

Spontaneous arousal is a physiological component of normal sleep and is defined as "any increase in Electromyography (EMG) or any channel which is accompanied by a change in pattern on any additional channel" in Rechtschaffen & Kales criteria in 1968 [7]. This conventional assessment of sleep is performed in epochs of 30 seconds [7, 8]. However, in other clinical conditions, frequent transient arousals of a few seconds duration were studied [9, 10]. In 1992, the American Sleep Disorders Association (ASDA) developed scoring rules to determine arousals quantitively based on data collected from EEG and EMG channels of Polysomnography (PSG) [6]. This scoring rule is independent of R & K's 30-second scoring system and standardizes the assessment of arousals. It has since become the most widely used rule for manually scoring by sleep experts. The ASDA defined sleep arousals as "an abrupt shift in EEG frequency, which may include theta,alpha and/or frequencies greater than 16 Hz but not spindles" [6]. A sleep spindle is a burst of oscillatory brain activity visible on an EEG that occurs during sleep stage 2. It consists of 12–14 Hz waves that occur for at least 0.5 seconds. For definitions of sleep stages, please refer to Appendix A. An EEG arousal event lasting for 0.56 seconds is shown in Fig. 1.1.

Other scoring rules or arousal definitions have been proposed. Arousals are defined as a return of alpha or theta rhythm for at least 1.5 seconds associated with a transient (however brief) increase in EMG tone in [11]. In [12], movement arousal is defined as an abrupt appearance of an alpha rhythm in the EEG during a sleep epoch, accompanied
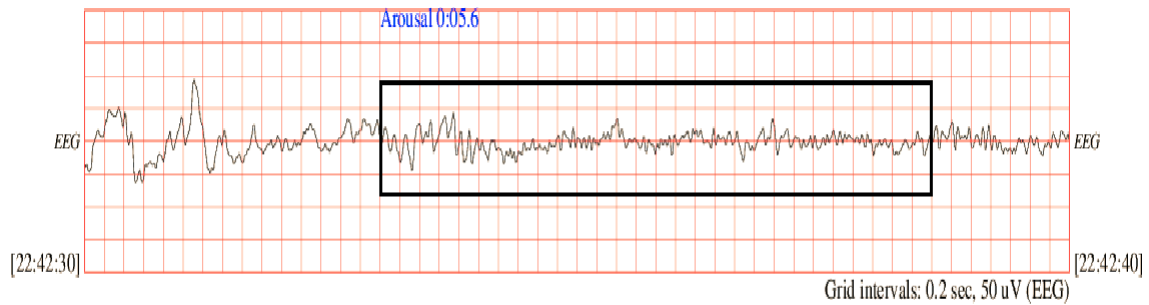
Figure 1.1: A 10 seconds' sleep EEG data from 22:42:30 to 22:42:40, collected from one EEG channel of a PSG recording downloaded from PhysioBank [1].The selected zone shows an EEG arousal event lasting 5.6 seconds which was manually scored by a sleep expert.

by an increase in EMG activity lasting at least 2 seconds. Simultaneous EEG and EMG changes must last at least 2 seconds. A return of theta rhythm and "K-arousals" are not counted. Some computer arousal detection methods have been developed which allow identification of arousals that are not visually scorable [13]. These methods make identification on arousal events which last less than 3 seconds achievable.

1.1.2 Sleep Arousal Identification

To identify sleep arousal events, one is traditionally asked to stay overnight in a hospital or a sleep laboratory to do a standard PSG test. It is an obtrusive test which requires the patient to wear a variety of sensors to collect several physiological signals. Arousal

events can be scored by sleep experts from one or several of the physiological changes recorded by PSG instrument.

Increases in heart rate, blood pressure during sleep are indicators for arousal identification [14]. It has been suggested that during room air breathing, arousals are strongly associated with periods of arterial oxygen desaturation [9].Thus, oxygen desaturation is another physiological signals for arousal identification. Changes in EEG and EMG activities are two important physiological indicators for arousal scoring which are used in ASDA standard arousal scoring rules [6].

According to ASDA arousal scoring rules, arousal events can be scored from one central EEG channel (two central EEG channels are C4/A1 and C3/A2 placement) obtained from C4/A1 or C3/A2 placement without accompanied EMG channel during non-REM (Rapid eye movement) sleep stages since arousals in NREM sleep may occur without concurrent increases in submental EMG amplitude (For EEG channel placement, please refer to Appendix C). However, arousal events should be scored only when accompanied by concurrent increases in submental EMG amplitude in REM sleep stages. "The presence of bursts of alpha or theata activity in REM sleep EEG are common phenomena. These events may or may not reflect physiological arousal from REM sleep". Thus, sleep arousals must be scored from both EEG and EMG activities during REM sleep. An arousal event is scored if the EEG frequency shift lasts for 3 seconds or greater in duration if scored manually since identification of events of shorter durations are difficult to achieve [6].

1.1.3 The Significance of Sleep Arousal Detection

Sleep problem is a common complaint even among healthy people, especially the elderly. Poor sleep quality may be indicated by reduced sleep time, increased sleep stage changes , and increased arousal frequency [9]. These short arousals are usually ignored in sleep analyses, but their impact is significant. Too many arousals during sleep can impair sleep continuity even when sleep efficiency is preserved [5]. Sleep efficiency can be indicated by the ratio between the number of hours slept and the number of hours spent in bed. Frequent sleep arousals degrade the quality of sleep and lead to sleep fragmentation. Sleep fragmentation is reported to influence the impairment of cognitive functions and is often associated with increased daytime sleepiness.[15]. Thus, sleep arousal frequency is a very important marker for sleep quality assessment.

The Pittsburgh Sleep Quality Index (PSQI) is a self-rated questionnaire which assesses sleep quality and disturbances over a 1-month time interval [16]. It is the most widely used method for sleep quality assessment as far. However, it is a relatively subjective approach. A more objective way based on detection of sleep arousals and other indicators such as the length of each sleep stage, etc. is worth developing.

Sleep arousal detection is also a key factor for research in sleep disorders including sleep apnea, periodic leg movement [5], snoring etc. and sleep of Parkinson's Disease (PD) patients [17].

1.2 Related Works

In this section, previous works on scoring or automatic detection of sleep arousals are reviewed and discussed. Data recordings, methodologies, and results of experiments are briefly described.

Currently, sleep arousal events are mostly diagnosed manually. Patients are asked to take an overnight PSG test which records several physiological signals. These recordings are then analysed and scored according to some rules by highly skilled sleep experts with specific domain knowledge. Various scoring rules and their reliabilities and validities have been developed and discussed in [13, 18, 19].

However, visual scoring of sleep arousals is time-consuming and cumbersome. Several automatic or semi-automatic detection methods based on computer algorithms have been proposed [15, 17, 20, 21, 22, 23, 24].

Detection methods in addition to analysing PSG recordings have been studied on. One of the detection methods is based on heart rate variability of electrocardiogram (ECG) and two other methods used peripheral arterial information to detect sleep arousals [20, 25, 26]. In [25, 26], patients are asked to take an overnight PSG test and a peripheral arterial tone (PAT) test simultaneously. The PAT signal and the pulse rate derived from it are then used to detect arousals from sleep. The total number of arousals scored by the PAT device is divided by the number of hours of sleep and termed PAT-based autonomic arousal indices (PAT-AAI). It is reported in [25, 26] that the sensitivity and specificity are 0.80 and 0.79 (for definitions of sensitivity and specificity, please refer to Appendix

B), respectively and area under ROC curve (AUC) is around 0.87. They only reported the results of patients with at least 20 arousals/hours [25, 26].

Concerning the EEG based detection methods, most of these adopt two or four EEG channels and one or two EMG channels [15, 17, 20, 21, 22, 23, 24]. Some of them also add other information such as heartbeat rate, oxygen carried by hemoglobin in the blood (SaO2%) [24], airflow pressure and airflow temperature [20].

In [24], an approach is developed based on statistical and data mining techniques. It first defined a set of general rules to detect arousals (termed meta-rules extraction step) with a training set of 6 adult patients' PSG recordings. The rules are then dynamically adjusted depending on the individual patient (called the actual-rules extraction step) and detected arousals. The correlations between occurrences of arousals and 2 central channel EEG (For the EEG 10-20 international system, please refer to Appendix C and [4, 27]), 1 channel chin-EMG, pulse (heart beat rate) and SaO2% signals were analysed on a test set of 20 patients' PSG recordings. The sensitivity and positive predictive values (PPV) were found to be 75.2% and 76.5% respectively (for definition of PPV, please refer to Appendix B). The sensitivity and PPV were found to be 49.4% and 82.5% when only EEG channels were used.

An automatic detection method of EEG arousals is described in [22]. The authors used two EEG channels (F4-C4 and C4-O2) and one EMG channel. In the first step of the study, a wavelet transform was used to process EEG signals and characterized the signal in the time-frequency domain. A set of indices were obtained after the first step. The indices

obtained from the first step was then used to estimate a linear discriminant function. Each 0.125-second epoch was evaluated with the function and arousals were marked when they last 3 or more seconds. Each possible arousal event was given a score. In the third step, the PSG recordings were inspected by two sleep experts independently. They then jointly examined the events scored by themselves and those scored by the computer's automatic algorithm for all the recordings. A reference set of arousal events named as definite arousals and uncertain arousals were obtained according to the two scorers' opinions. They defined a correctly detected arousal as an arousal event which overlapped with the reference set. They reported an overall sensitivity of 88.1% for the automatic method and 72.4% and 78.4% for the two experts with a selectivity of 74.4% for the automatic method and 83.0% and 82.0% for the experts when only definite arousals were considered. The sensitivity decreased to 84.5% , 67.9%, 73.2% and selectivity increased to 88.4%, 96.1%, and 94.6% for the computer, expert 1 and expert 2 respectively when all possible arousals were included in the reference set.

In [20], a study was conducted on the detection of respiratory-related arousals. In this work, a method for automatic detection of EEG arousals in sleep apnea syndrome (SAS) patients was proposed. PSG recordings including four channels of EEG (C3-A2, C4-A1, O1-A2, O2-A1), two channels of EMG, electroocculography (EOG), ECG, airflow pressure and temperature etc. were used. First, data were segmented into 2.56 seconds for EEG and EMG data. For respiratory data (airflow pressure and aiflow temperature), 10.24 s was adopted as the segmentation length according to the definition of the duration

of SAS [20]. Then, some fundamental parameters of amplitude, relative power and central frequency of EEG were calculated. Airflow pressure and temperature information were used for detecting pathological events, such as obstructive sleep apnea (OSA), which were then utilized for determining threshold values for EEG arousal detection. The authors reported an overall accuracy defined as the percentage of (TP + TN)/(TP + TN + FP + FN) (please see Appendix B) of 86%, a false negative rate (FNR) of 18% and a false positive rate (FPR) of 12%.

Another automatic detection method based on the idea of segmentation, spectral feature extraction, statistical methods and decisional rules is described in [21]. Two EEG channels of 2 patients' PSG recordings were utilized and three sleep experts were asked to score the sleep arousal events in this study. An automatic detection is assumed to be a valid arousal event if there is any overlap with the manually marked events. For one patient, the sensitivity and specificity were 82.2% and 72.4% when compared to score A. The sensitivities and specificities were found to be 66.4%, 81.8% and 74.5%, 67.3% respectively when compared to scorer B and scorer C. For the other patient, the sensitivity and specificity were 70.1% and 71.1% when compared to score A. The sensitivities and specificities were found to be 42.7%, 80.3% and 74.1%, 56.6% respectively when compared to scorer B and scorer C.

An approach to detect sleep EEG arousals based on signal processing and machine learning paradigm is presented in [23]. Two channels' EEG signals and one channel's chin-EMG signal of each of 10 patients' PSG sleep recordings were used. In the first phase, raw

data were segmented into one-second epochs. The energies of different sleep bands were measured using the Fourier Transform (FT). A set of 40 features of the 3 channels' signals were extracted in total to train classifiers. In the second phase, several models based on the classic Fisher's linear Discriminant, a quadratic discriminant, several configurations of Support Vector Machine (SVM) based on different parameters, and configurations of feed-forward Artificial Neural Networks (ANN) of different neurons in one hidden layer were tested. The SVM and ANN models achieved better performances than the other two classifiers and the best overall accuracy was reported to be 0.92 which was achieved by one model of ANN.

Two studies based on segmentation, feature extraction and machine learning techniques are reported in [15] and [17]. In [17], four channels of EEG (two central (C3-A2, C4-A1) and two occipital (O1-A2, O2-A1)) and one submental EMG channel of PSG recordings were used. Patients recruited in this study were patients with Parkinson Disease (PD). After data preprocessing, a titak if 14 features were extracted including sleep stages scored by sleep experts. Then, a two-layer feed-forward neural network with 9 neurons in a hidden layer was applied to classify the arousals with features extracted previously. In the last step, a postprocessing step was added to combine arousals classied in a certain proximity of each other. Arousals closer than 10 seconds from each other were combined to one arousal event. Arousals detected but lasting less than 3 seconds were removed. The authors assumed correctly detected arousal events as ones which overlapped with manually scored arousals. They reported an average sensitivity of 89.8% and PPV

of 88.8%.

In [15], only a single channel EEG (C3-A2) was used to automatic detect sleep arousals. Sleep data of non-REM sleep stages (wake stages and REM stages were excluded) of 9 PSG recordings of patients with sleep apnea, snoring and excessive daytime sleepiness (EDS) were used. After some preprocessing of the data, time-frequency analysis was used to extract several features. In the last step, the support vector machine (SVM) classifier was applied to features extracted based on 1-second epochs. The information of manually scored sleep stages was also included as one of the features. The authors reported that the proposed method achieved a sensitivity of 75.26% and specificity of 93.08% compared to the sleep expert's scores.

1.2.1 Discussion

To our knowledge, none of the works has reported a comparison between its own result and that of other works. Maybe it is because it's hard and unfair to do a comparison. Several possible reasons are listed as follows.

- First, different dataset were used in different works. Most of the studies collaborated with their own hospitals to recruit patients to collect PSG data. The devices used and patients participating in the test can vary a lot among different studies. In addition, various sleep experts involved in annotating the sleep arousal events in different studies. The results reported in [21] indicated the big difference between different scorers when doing annotation. The sensitivity can be as high as 70.1%

when compared with the annotations of scorer A and can be as low as 42.7% when compared with scorer B for the dataset of the same patient.

- Second, different physiological signals are used in different studies. For example, several studies used 2 channels of EEG signals and others used 4 channels of EEG signals. A few works added airflow pressure or temperature information while others utilized heartbeat rate etc.

- Last but not least, various performance evaluation methods are used in different works and there is no standard criteria for performance evaluation on this sleep arousal detection problem. For example, in [15], sensitivity and specificity were used to report research results. The annotations are segmented into 1-second epochs and the sensitivity and specificity were computed based on 1-second epochs. In another study [17], sensitivity and PPV were used and no specificity was calculated. In addition, they calculated the results based on arousal events lasting more than 3 seconds. And they assumed correctly detected arousal events as ones which overlapped with manually scored arousals. Based on their method, a high sensitivity could be achieved while the duration, start and end of an arousal event may vary a lot from the sleep expert's annotations. In study [22], sensitivity and selectivity were used as evaluation. In addition, they established a jointly reference set based on two sleep experts' annotations and the experiment result of the automatic detection algorithm and it was then used as the gold standard to compute sensitivity and selectivity instead of traditional annotation of one sleep expert. In [20], an

overall accuracy, false negative rate and false positive rate were used to present the experiment result.

In addition, most of the studies utilized imbalanced dataset to evaluate performances and they did not use imbalanced learning algorithms or make a balanced dataset. This may lead to a good overall accuracy while the real performance is bad.

1.3   Motivations

Sleep arousals are associated with various sleep disorders and can be a good indicator for sleep quality assessment. So far, sleep EEG arousals are mostly diagnosed by sleep experts with specific domain knowledge and the patient is required to take an overnight sleep test in the hospital or a sleep lab. There are several disadvantages for this kind of traditional sleep test. For example:

- It is very time consuming and cumbersome for a sleep expert to manually score sleep arousals because the expert needs to visually inspect the different channels of a PSG recording including EEG, EMG, EOG etc.

- Visual inspection is a relatively subjective way to diagnose sleep arousals. There can be large differences between individual sleep experts. For instance, in [22], the sensitivity of sleep expert A was 72.4% and the sensitivity of sleep expert B was 78.4% compared to the reference set which was jointly scored according to the computer algorithm's result, sleep expert A and sleep expert B's results. And

the agreement between the two experts was only 68%. In another study[21], the sensitivities of the automatic method varies from 70.1% to 42.7% when compared to two different scorers for the same patient. These results indicate less accuracy of manually scoring.

- It has been suggested that arousals of short durations (less than 3 seconds) may also be significant [13]. However, identification and agreement on events of such short durations are difficult to achieve, if scored manually.

- From a patient's perspective, the cost for a PSG test is high (ranging from $700 to $6000). The patient need to sleep in a sleep lab or hospital for a full night with a lot of electrodes attached to the patient's body. This may disturb the sleep process of the patient which makes the test less reliable. In addition, a sleep technician should always be in attendance and is responsible for attaching the electrodes to the patient and monitoring the patient during the study.

Due to the above–mentioned disadvantages, research on fast, accurate computer-aided automatic arousal detection approaches and portable, less obtrusive detection devices which allow patients to take the tests at home are of great significance. Several studies on automatic or semi-automatic sleep arousal detection are mentioned in Section 1.2. However, a number of issues still need to be solved.

- Most of the previous studies utilized various physiological information collected from a number of channels of PSG tests. For example, the method in [24] used 2

central channels of EEG (C4-A1, C3-A2), 1 channel of chin-EMG, pulse (heartbeat rate) and SaO2% signals. In [20], 4 channels of EEG (C3-A2,C4-A1, O1-A2, O2-A1), 2 channels of EMG, airflow pressure and airflow temperature, etc. were used. A large number of channels of information collected means more inconvenience for patients since more electrodes need to be placed on patients. In order to manufacture portable, less obtrusive devices, use of fewer electrodes is desirable. However, fewer channels of information lead to lower accuracy. For instance, the sensitivity was decreased from 75.2% to 49.4% when only 2 EEG channels were used in [24]. Thus, methods which can achieve relatively high accuracy with less physiological information collected should be studied.

- In sleep arousal detection, the amount of patient data is quite huge and takes long time to be processed, even by computer algorithms. In order to analyse data more effectively and even make real-time display achievable, a relatively fast algorithm with high accuracy is necessary.

- Several features need to be extracted to train the classifier in machine learning based algorithms. Features are chosen or added based on previous works or the researcher's own view in most of the studies. However, redundant or unimportant features may be added during the feature extraction process which will lower down the speed of the algorithm and may even decrease performance. Thus, a reliable feature selection algorithm is crucial.

1.4 Contributions

In this thesis, an algorithm to detect non-REM sleep EEG arousals using only 1 channel EEG(C4-A1/C3-A2) is developed. The main contributions are summarized as follows:

- In chapter 2, an automatic sleep arousal detection algorithm is proposed. Raw data from [1] is used in our study. A set of 22 features different from previous work is extracted based on 1-second segmentation from the preprocessed dataset. A recently proposed classifier named Curious Extreme Learning Machine (C-ELM), which is fast and easily implemented is adopted to do a binary classification on the whole feature set. The widely used Support Vector Machine (SVM) classifier is also used on the feature set. The information of the accuracy and the Area Under the Receiver Operating Characteristic (ROC) Curve (AUC) are calculated and compared for both our C-ELM-based and SVM-based detection algorithms. The speed of the two methods are also compared. During the process, 10-fold cross validation is used to avoid bias due to luckily/unluckily selected validation set, thus making the performance estimate less sensitive to the partitioning of the data. The result shows that our C-ELM based detection model has a better performance than SVM-based model.

- In Chapter 3, an improved automatic sleep arousal detection model based on Minimum Redundancy Maximum Relevance (MRMR) feature selection method and C-ELM are proposed. MRMR feature selection step is added to reduce the dimen-

sionality of the feature set and determine the subset of features which gives the best performance. It is shown that a subset of 17 features achieves the best performance and a set of 6 features can still have a similar performance to the 17 feature set. A low dimension feature set can increase the speed of the sleep arousal detection algorithm. By the result obtained, the improved model is found to achieve a good performance with a reduced system complexity. This result also indicates that the MRMR feature selection step plays an important role in designing an sleep arousal detection algorithm which is fast and accurate.

1.5 Structure of the Thesis

The remainder of the thesis is organized as follows.

In Chapter 2, we present our C-ELM classifier based algorithm and evaluate its classification performance. The model and process are first described. Then, data preprocessing and segmentation are introduced. Next, feature extraction and C-ELM, SVM classification are described . Finally, the performance of the algorithm is discussed (cross validation was utilized since the dataset is limit).

In Chapter 3, we propose an improved algorithm with MRMR feature selection. First, a few feature selection methods including MRMR feature selection method are described and compared. Then, binary classifications using C-ELM and SVM are performed on different feature subsets according to the MRMR feature selection ranking. The performances are also reported in this part. Finally, a brief summary of this chapter is

given.

Chapter 2

Automatic Sleep EEG Arousal Detection based on C-ELM

In this Chapter, we present our C-ELM classifier based algorithm and evaluate its classification performance. Support Vector Machine (SVM) has good performance on binary classification problems and it has been reported that it performs well when applied to sleep arousal detection problems[15, 23]. So, we also apply SVM on our feature set for comparison with our C-ELM based model. The overall sleep arousal detection model is described in Section2.1. In Section2.2, data preprocessing including the selection of raw data and band-pass filter process and segmentation are introduced. In Section2.3, various features are described and extracted from the preprocessed data. In Section2.4, theory of Curious Extreme Learning Machine (C-ELM) is studied. In Section2.5, the sleep arousal detection performances using models based on C-ELM and SVM binary classifications are compared. Cross validation is used to reduce the variance for different datasets when we do the performance estimate. The AUCs, ACCs and training times of C-ELM and SVM based algorithms are compared. In Section2.6, a summary is given.
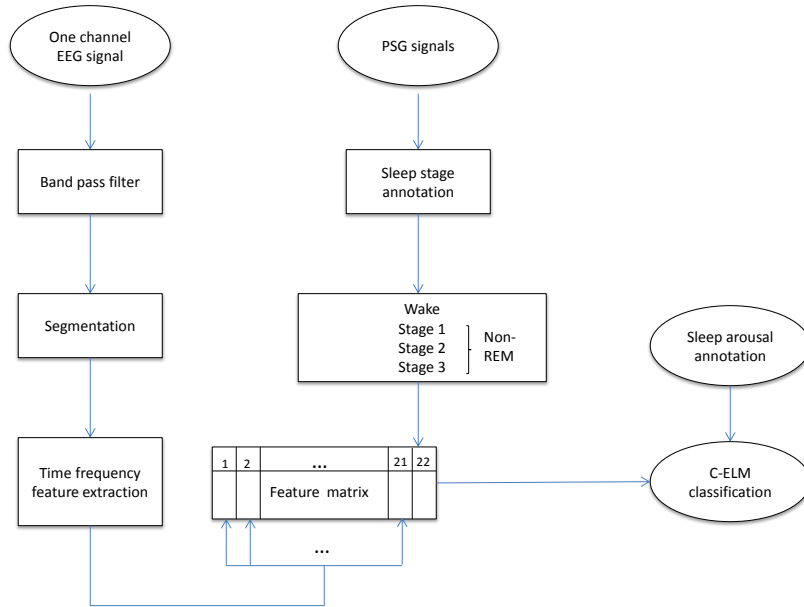
Figure 2.1: Sleep arousal detection model.

2.1 Sleep Arousal Detection Model

Our sleep arousal detection algorithm is based on segmentation and classification. First, raw sleep dataset which contains noise is obtained from Physiobank[1]. A band pass filter is used to remove artifacts and irrelevant information. Next, the preprocessed dataset is segmented into 1-second epochs in order to do the classification. Since the input data is too large to be processed, a feature extraction step is used to transform the raw dataset into feature vectors which contain the relevant information. Finally, the feature vectors of the dataset are input into the Curious Extreme Learning Machine (C-ELM) classifier and SVM classifier. The overall model is illustrated in Fig. 2.1.

2.2 Data Preprocessing and Segmentation

We now describe the Data acquisition, data preprocessing including band pass filter and segmentation.

2.2.1 Data Acquisition

One central EEG channel (C4-A1/C3-A2) with a sampling frequency of 250 Hz of a patient's single overnight PSG recording is utilized in this thesis. The EEG raw data is downloaded from Sleep Heat Health Study (SHHS) PSG DataBase of PhysioBank[1]. The SHHS is a prospective cohort study designed to investigate the relationship between sleep disordered breathing and cardiovascular disease. The age of the patient used in the study is over 40, without tracheostomy, without history of treatment of sleep apnea, without current home oxygen therapy. Other information, such as a sleep expert's annotations of arousal events and sleep stages, are downloaded from PhysioBank as well.

According to the ASDA manually scoring rules [6], arousal events during REM sleep stages must be scored when at least one EMG channel is used since the arousal events during REM sleep stages must be accompanied by an increase in submental EMG according to ASDA rules [6]. So only data of non-REM sleep stages (sleep stage 1, 2, 3) and wake stage are included in this study. Consequently, we have investigated a total of 1,920,000 samples (7680 seconds) for arousal detection.

2.2.2 Band Pass Filter and Segmentation

According to [15, 17, 20, 21, 22, 23, 24], sleep related frequencies can be divided into 6 bands: 0-0.5 Hz (gamma or slow delta), 0.5-4 Hz (delta), 4-8 Hz (theta), 8-12 Hz (alpha), 12-16 Hz (sigma), 16-30 Hz (beta). Some of the above-mentioned works define the beta band as 16-64 Hz [22], 16-40 Hz [21] or >13 Hz [24]. In [6], sleep EEG arousals are related to the theta, alpha and beta bands. Other sleep bands are related to sleep stages or sleep spindles. In order to remove noise and frequencies non-related to sleep, we band-pass filter the raw EEG signal from 0-50 Hz.

Analysis tools, such as Fast Fourier Transform (FFT), are widely used to process EEG signals. However, in this research, we want to identify sleep arousals based on 1-second epochs. Thus a time-frequency representation is performed which enables us to obtain time and frequency information simultaneously. It is useful in analyzing complex physiological signals [15]. In order to do time-frequency analysis and extract feature vectors from the signal every second, the band-pass filtered signal is segmented into 1-second epochs. Then, frequency analysis is performed for each epoch. A total of 7680 epochs of sleep data are thus obtained.

2.3 Feature Extraction

In this section, features extracted from the sleep EEG data are listed and described. In this stage, a total of 22 features are extracted from one single channel EEG to be used

in the classification stage. Some of the 22 features may be redundant or features with less predict power; however, in this stage, this is not of immediate concern. This will be discussed in the feature selection part in Chapter 3.

All the 22 features are now described. In the feature extraction step, Fast Fourier Transform (FFT) is used for the frequency and power analysis.

- Power Ratio: According to the ASDA scoring rules [6], sleep EEG arousals are abrupt frequency shifts of theta, alpha and beta sleep bands. The frequency shift can be represented by the changes of power in time. First, for each one second epoch, two temporal windows which contain the power information are chosen. A window of 10 seconds ending in the current epoch is used to represent prior power information and another window starting from the current epoch is chosen to provide the current or future power information. The changes of power can be represented by the power ratio between these two windows. According to [15], we make the "future" window 1 second in length and we also choose another "future" window of 3 seconds according to [23]. That is to say, we have 2 different power ratio frames. One is 1 second/10 second frame and the other one is 3 second/10 second frame. The duration of 10 seconds as the former window length also comes from the ASDA scoring rules [6]. In [6], the scoring rule suggests a minimum of 10 seconds of intervening sleep is necessary to score a second arousal once a previous arousal is detected. So we choose 10 seconds as the length of the former window. Next, each of the two windows are transformed to the frequency domain

using FFT, and the power of each window can be calculated. Each of the six sleep bands' power ratios including theta ratio (4-8 Hz), alpha ratio (8-12 Hz), beta ratio (16-30 Hz), gamma ratio (0-0.5 Hz), sigma ratio (12-16 Hz) and delta ratio (0.5-4 Hz). and the whole power ratio (0-50 Hz) are calculated and extracted as features. A total of 14 features (we have two frames of windows (3-second/10-second and 1-second/10-second)) are extracted based on the power ratios.

- Sleep Spindle: It is stated in the ASDA scoring rules [6] that arousals are abrupt EEG frequency shifts which are not sleep spindles. Hence, the power ratio between sigma and (alpha plus beta) using 3-second/10-second window frame is selected to indicate the presence of sleep spindles [15, 22].

- Mean Frequency: The signal's mean frequency of each 1-second epoch is extracted as a feature [15]. The mean frequency is computed as follows [22].

$$\bar{f} = \frac{\sum p_i \times f_i}{\sum p_i}, \tag{2.1}$$

where f_i is the center frequency of the band and p_i is its power.

- Power and Max Power Frequency: The power of 0-50 Hz band for each 1-second epoch is selected as a feature. Another feature is max power frequency, which is defined as the frequency corresponding to the maximum power or maximum amplitude in the FFT amplitude spectrum.

- Time Domain Based Features: The mean value and standard deviation of the signal in the time domain are selected as features for each 1 second. An abrupt shift in EEG frequency may be indicated by the number of zero-crossing, so the number of zero-crossing is another feature selected in the time domain. We choose the mean value of each second as "Zero" (the baseline). A large number of zero-crossing may indicate an abrupt shift in EEG frequency occurs, thereby an arousal may have happened. These three features are added from our own perspective based on the ASDA rules [6].

- Sleep Stages: Although it is widely accepted that the scoring of sleep arousals is independent of Rechtschaffen & Kales criteria [7], the selection of sleep stages as a feature is still necessary. First, an arousal event is easy to be incorrectly scored during the wake stage. Second, sleep stages are characterized by various sleep waves (theta wave, alpha wave etc.). In this thesis, the annotations of sleep stages are downloaded from Physionet [1] which is manually scored by sleep experts.

To have a brief summary of all the features extracted in this study, please see Table. 2.1.

2.4 Classification based on C-ELM

In this section, the Curious Extreme Learning Machine (C-ELM) algorithm is briefly described. A detailed explanation can be found in [28]. Descriptions of Support Vector

Table 2.1: The 22 features extracted from the preprocessed data

Feature type	# of features	Description
Power Ratio (3-sec/10-sec)	7	Power ratio between 3 seconds starting from the current epoch and prior 10 seconds of the 0-50 Hz band and six individual sleep bands (alpha, beta, etc.)
Power Ratio (1-sec/10-sec)	7	Power ratio between the current epoch and prior 10 seconds of the 0-50 Hz band and six individual sleep bands (alpha, beta, etc.)
Sleep spindle	1	Power ratio between sigma and alpha plus beta using 3-second/10-second windows
Mean frequency	1	The signal's mean frequency of each 1-second epoch Eq. (2.1)
Power	1	The power (0-50 Hz) of each epoch
Max Power Frequency	1	The frequency corresponding to the maximum amplitude in FFT amplitude spectrum
Time Domain Features	3	The zero-crossing frequency, mean value and corresponding standard deviation of the each epoch
sleep stages	1	Annotations of sleep stages

Machine (SVM) appear in [29, 30, 31, 32].

Extreme Learning Machine (ELM) is a fast, easy-to-implement machine learning algorithm based on a single hidden layer feedforward neural network (SLFN) without parameter tuning. It has been reported to have good performance and generalization ability [33, 34, 35]. Details about ELM and related algorithms based on ELM can be found in [33, 36, 37, 38, 39, 40]. Curious Extreme Learning Machine (C-ELM) is a psychological curiosity driven algorithm based on ELM. It follows psychological theory of curiosity and performs curiosity appraisal towards each input data. The algorithm has four variables (novelty $\mathcal{N}(\mathbf{x}^t)$, uncertainty $\mathcal{U}(\mathbf{x}^t)$, conflict $\mathcal{C}(\mathbf{x}^t)$ and surprise $\mathcal{S}(\mathbf{x}^t)$) and three learning strategies (neuron addition, neuron deletion and parameter update). The four variables are computed for each input vector $\mathbf{x}^t$ and compared with initialized thresholds. According to the comparison result, one corresponding learning strategy is utilized to adjust the structure or update the parameters of the neural network automatically.

The conditions for the three learning strategies [28] are briefly summarized below.

- Neuron Addition Strategy: Given an input $\mathbf{x}^t$, the neuron addition condition is:

$$\mathcal{N}(\mathbf{x}^t) > \theta_{\mathcal{N}^{add}} \ \text{ AND } \ \mathcal{U}(\mathbf{x}^t) > \theta_{\mathcal{U}} \ \text{ AND } \ \mathcal{S}(\mathbf{x}^t) > \theta_{\mathcal{S}}, \tag{2.2}$$

where $\mathbf{x}^t = [x_1^t, \cdots, x_M^t]^T \in \Re^M$ is the tth M-dimensional input vector (or feature vector) of the training data $\{(\mathbf{x}^1, c_1), \cdots, (\mathbf{x}^t, c_t), \cdots\}$ ($c_t \in [1, 2, \cdots, N]$ is the class label of $\mathbf{x}^t$, N represents the total number of distinct classes), $\theta_{\mathcal{N}^{add}}$, $\theta_{\mathcal{U}}$ and

$\theta_{\mathcal{S}}$ are initialized neuron addition thresholds in the range of $[0.1, 0.5]$, $[0.1, 0.3]$, $[0.2, 0.9]$ for novelty, uncertainty and surprise, respectively.

- Neuron Deletion Strategy: Given an input $\mathbf{x}^t$, the neuron deletion condition is:

$$\mathcal{S}(\mathbf{x}^t) > \theta_{\mathcal{S}} \ \text{ AND } \ \mathcal{C}(\mathbf{x}^t) > \theta_{\mathcal{C}} \ \text{ AND } \ \mathcal{N}(\mathbf{x}^t) < \theta_{\mathcal{N}^{del}}, \tag{2.3}$$

where $\mathbf{x}^t$ is the same as in Eq. (2.2), $\theta_{\mathcal{N}^{del}}$, $\theta_{\mathcal{C}}$ and $\theta_{\mathcal{S}}$ are initialized neuron deletion thresholds in the range of $[0.1, 0.8]$, $[0.1, 0.3]$, $[0.2, 0.9]$ for novelty, conflict and surprise, respectively.

- Parameter Update Strategy: When both neuron addition and deletion conditions not satisfied, it indicates the new input vector is a 'familiar' data. The number of hidden neurons will not be changed and the output weights are updated.

A pseudocode description of C-ELM is given in Algorithm1.

Algorithm 1 Pseudocode for Curious Extreme Learning Machine.

1: **Step 1**: Present an input vector $(\mathbf{x}^t, c_t)$.

2: **Step 2**: Compute four variables (novelty $\mathcal{N}(\mathbf{x}^t)$, uncertainty $\mathcal{U}(\mathbf{x}^t)$, conflict $\mathcal{C}(\mathbf{x}^t)$ and surprise $\mathcal{S}(\mathbf{x}^t)$) according to the input vector.

3: **Step 3**: Select one learning strategy out of three (Neuron Addition, Neuron Deletion, Parameters Update) based on the four variables and corresponding thresholds.

4: **Step 4**: Increment t to t+1, repeat Step 1 to Step 3.

2.5 Performance Evaluation

In this section, we first apply C-ELM and SVM to the feature vectors of the dataset. Since our data are limited, a 10-fold cross validation is utilized to gain insight into how our model will generalize to an independent dataset (i.e., how accurately this model will perform in practice). Then the Area Under the Curve (AUC) and Accuracy (ACC) are computed and used as the criteria for our performance evaluation. In addition, the training speeds of our C-ELM based model and the SVM based model are discussed and compared.

During a patient's overnight sleep, the number of arousal events can range from tens to hundreds. Each event can last from several seconds to more than 15 seconds (currently no terminal criteria is established according to ASDA scoring rules [6]). However, the total duration of all arousals during one night of sleep is quite small, around 20 or 30 minutes out of 8 hours. That is to say, the data can be quite imbalanced when applied to a classifier. Thus, the accuracy could be overestimated. In this study, there are only 144 epochs among the total of 7680 epochs which are labeled as positive data (arousals) by sleep experts. To solve the imbalance problem, we perform classifications using the following procedure.

- First, 144 negative epochs are selected randomly from a total of 7536 non-arousal epochs. The 144 positive epochs are combined with the selected negative epochs to form a balanced dataset of 288 epochs in total.

- Second, randomize the dataset of 288 epochs obtained in the first step and divide it into 10 folds for cross validation. Then we apply C-ELM and SVM to the randomized dataset. Each one of the 10 folds is used as a test set in turn, with the other 9 folds used as training sets. Thus, for each test fold, decision value of each input epoch (decision value is used to determine the predicting result, such as positive or negative in binary classification) is obtained.

- Third, a Receiver Operating Characteristic (ROC curve) is plotted according to decision values obtained from the second step. For details of ROC curve, please refer to Appendix D. Finally, AUC and ACC are computed from the ROC curve.

- Repeat step 1 through step 3 for 50 times. The 50 AUC and ACC results are discussed later in this section.

In this thesis, the Library of Support Vector Machine (LIBSVM)[41] is used to train and test data in the SVM based model. In the training step, Radial Basis Function (RBF) kernel function is used for the Support Vector Machine because RBF kernel usually has a better performance for classification problems [29]. A grid search is utilized to tune parameters in order to optimize the performance of the SVM based model. The C-ELM based model is trained and tested using the source code from [28]. The parameters used are the ones that provided the best classification performance in previous experiments according to [28]. The learning thresholds are set as follows.

- The low threshold of novelty = 0.1;

Table 2.2: Average AUC comparison of C-ELM based model and SVM based model

Properties	C-ELM based model	SVM-based model
Average AUC of 50 datasets	0.85	0.69
Standard deviation of 50 datasets	0.0163	0.1573

- The high threshold of novelty = 0.4;

- The uncertainty threshold = 0.1;

- The conflict threshold = 0.3;

- The surprise threshold = 0.4;

2.5.1 AUC and ACC Evaluation

The average AUC and ACC results of the 50 datasets for the C-ELM based model and the SVM-based model are listed in Table. 2.2 and Table. 2.3, respectively. The standard deviation of the 50 AUC and ACC results are also listed in Table. 2.2 and Table. 2.3. The best C-ELM based result and its corresponding SVM result are summarized in Table. 2.4 and the ROC curves are plotted in Fig. 2.2. The best SVM based result and its corresponding C-ELM result are summarized in Table. 2.5 and the ROC curves are plotted in Fig. 2.3.

According to the results shown in two tables, an average AUC of 0.8527 and ACC of 0.7903 are achieved by our C-ELM based model while an average AUC of 0.6916 and ACC of 0.6719 are obtained by SVM based model. These results indicates the sleep

Table 2.3: Average ACC comparison of C-ELM based model and SVM based model

Properties	C-ELM based model	SVM-based model
Average ACC of 50 datasets	0.79	0.67
Standard deviation of 50 datasets	0.0179	0.1218

arousal detection model based on C-ELM performs very good on our datasets and the SVM based detection model is relatively poor. This comparison result is consistent with those of other problems reported in [28]. In [28], both C-ELM and SVM are evaluated on the benchmark problems from the UCI machine learning repository which contains three multicategory classification problems and three binary classification problems. It is reported that C-ELM performs better than SVM on all the six problems. The overall accuracy of C-ELM is greater than that of SVM by 0.12 for the Vehicle problem.

The standard deviation of AUC of our C-ELM model is only 0.0163 while the SVM based model reaches 0.1573. We can see that the best AUC achieved by C-ELM based model is around 0.89 from Table. 2.4 and Fig. 2.2. The similarity between the best result and the average AUC 0.8527 and a relative small standard deviation of 0.0163 may indicate the input data of most datasets among the 50 datasets are randomized well and the C-ELM based model is stable on all the 50 datasets. However, the average AUC of the SVM based model is around 0.7. It is much smaller than the best AUC which is around 0.89. And we can also find that the standard deviations of 50 AUCs and ACCS for the SVM based model are much greater than those of C-ELM based model. Because we apply the same dataset to both C-ELM based model and SVM based model

simultaneously, and good results of C-ELM based model have suggest the datasets used are randomized well, The above-mentioned relatively poor results of the SVM based model probably indicate this model is unstable on our datasets. The possible reasons for the relatively poor performance of SVM based model compared to the model proposed by us are discussed below.

- In our study, we used RBF kernel function for SVM based model, then two parameters of the model and the kernel function need to be tuned to optimize the performance and avoid overfitting. For each one of 50 datasets, we do grid searches to choose the best values for the two parameters. When we do cross validation for each dataset, grid search is applied to determine values of the parameters for each 1 of 10 folds. The complexity in tuning parameters for SVM results in big variance of parameter's values which may cause the unstable performance of SVM based model. In regarding to this problem, we might choose another parameter tuning strategy. For example, we apply grid search on each 1 of the 50 datasets and then use a major vote method to determine one optimal value for each of the 2 parameters. For all the cross validation procedures we can use the fixed parameters obtained by major vote.

- Choosing the kernel function is probably the most tricky part of using SVM. The kernel function is important because it creates the kernel matrix which summarizes all the data [42]. RBF kernel function is used in SVM classifier in our study because this kernel function is always a good try in various problems [29, 42]. However, what

could happen is that RBF kernel is not a good choice on our data. For example, if our data is linear distributed, we used RBF kernel instead of linear kernel with poor parameters selected, this could cause over fitting problem which leads to a less effective classifier. Another seperate experiment is done to observe whether SVM based model has an over fitting problem. It is shown that the average training accuracy of 50 datasets is greater than the average testing accuracy by around 0.1 which indicates over fitting problem might have occur in some of the 50 datasets.

- Curious Extreme Learning Machine (C-ELM) is based on Extreme Learning Machine (ELM). Compared to SVM, ELM has some advantages which may lead to a better performance in our study. The hidden node parameters can be generated without the knowledge of the training data and no parameter tuning is needed for ELM [38]. The constraint of the choose of kernel is much smaller on ELM than SVM. That is to say, ELM may generalize better than SVM regardless of kernel choosing and the distribution of the data.

- Curious Extreme Learning Machine (C-ELM) is an enhanced ELM. It is reported to have a better performance than ELM on all the 3 binary classification benchmark problems studied in [28]. It reduces the randomization effect of ELM mainly by providing an optimal number of hidden neurons. The hidden neuron addition or deletion strategy based on curiosity may helps in avoiding over fitting.

Table 2.4: Best performance of C-ELM based model among 50 datasets and the corresponding performance of SVM based model of the same dataset

Properties	Best C-ELM based performance	SVM based performance
AUC of one dataset	0.8843	0.8271

Figure 2.2: ROC curves for C-ELM based and SVM based models for the dataset which gives the highest AUC for C-ELM. The red line is a random classification, the blue line is the curve of C-ELM and the green line is the curve of SVM

Table 2.5: Best performance of SVM based model among 50 datasets and the corresponding performance of C-ELM based model of the same dataset

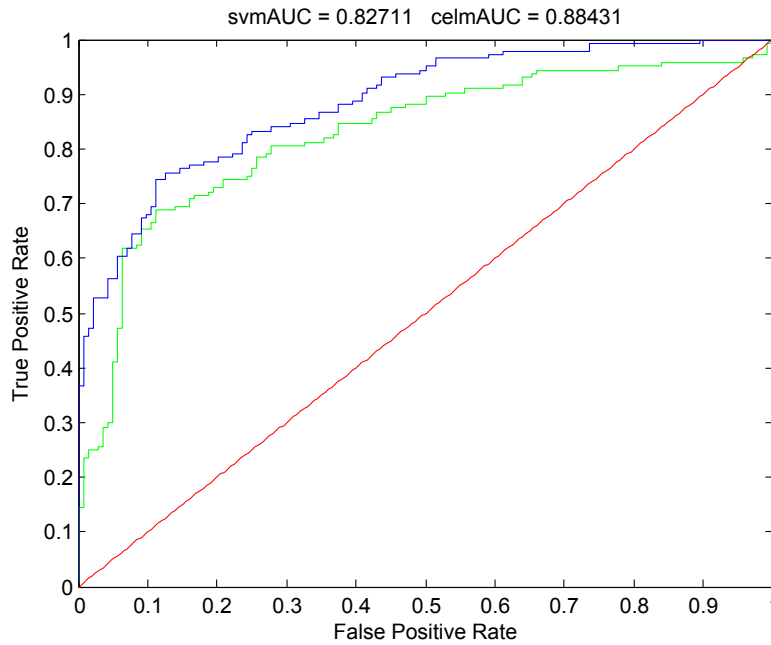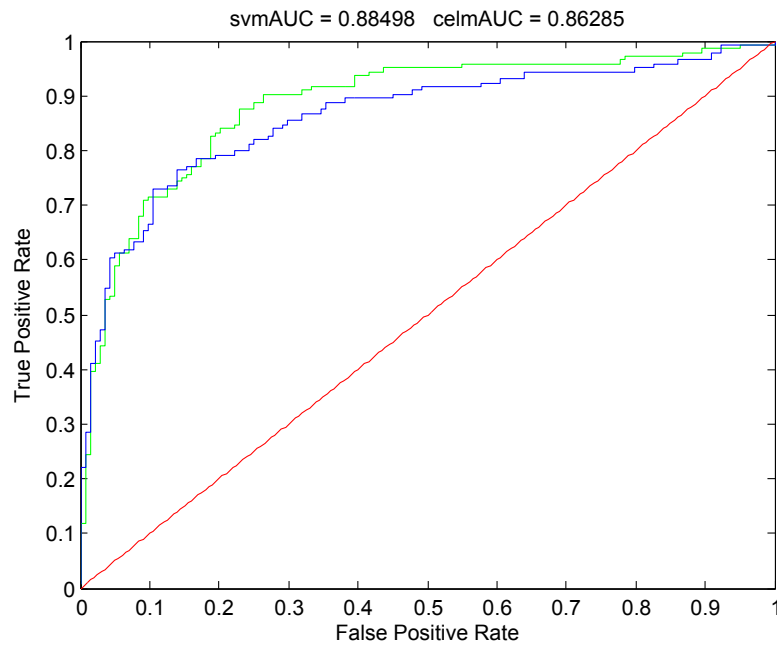| Properties	Best SVM based performance	C-ELM based performance
AUC of one dataset	0.8850	0.8629

Figure 2.3: ROC curves for C-ELM based and SVM based models for the dataset which

gives the highest AUC for SVM. The red line is a random classification, the

blue line is the curve of C-ELM and the green line is the curve of SVM

2.5.2 Speed Evaluation

In order to do a fair comparison between the training times for the C-ELM based model and the SVM based model, we use the built-in SVM training function of MATLAB R2012b to do the classification instead of the function from LIBSVM[41] since the LIBSVM utilizes c/c++ source code (Matlab is slower than C/C++ which would make the comparison unfair). The kernel function of the SVM classifier is RBF. A total of 288 observations, each with 22 features (288*22 matrix), containing 144 positive data and 144 negative data were selected randomly from the 7680 observations. The training times for both models are shown in Table. 2.6. It can be seen that the training speeds for the two models are similar. This result is consistent with those reported in [28]. In [28], the training times for C-ELM and SVM are similar for all the three benchmark binary classification problems. For example, for Brest cancer problem, The training time of SVM is 0.11 while that of C-ELM is 0.09. A total of 300 training data with dimension of 9 are used in the Breast cancer problem. In this thesis, it is just a rough comparison for the specific dataset. The training time depends on the dataset, kernel function used as well as the coding implementation of the algorithm and so on. In addition, we don't tune parameters in our model while a grid search is utilized to optimize the SVM based model. If considering the total executing time of the sleep arousal detection model, the SVM based model is slower than the C-ELM based model and the executing time of SVM based model depends on the complexity of the grid search. A more thorough evaluation of the training times is required but this is not the main aim of this thesis.

Table 2.6: Training times for C-ELM based model and SVM based model for a dataset with dimension of 22

	C-ELM based model	SVM based model
training time (seconds)	0.079	0.080

2.6 Summary

In this Chapter, a new model based on a new set of 22 features and Curious Extreme Learning Machine (C-ELM) for sleep arousal detection has been proposed. Data acquisition, preprocessing and segmentation are first described followed by the feature extraction procedure. Brief descriptions of C-ELM and SVM classification algorithms are also provided. The performance of the new model is presented and compared to that of a SVM-based model. It is found that the proposed model of sleep arousal detection has a good performance even though only one single EEG channel and limited data is used. The proposed model has a higher AUC and ACC which indicates a better ability to correctly classifies a random data as a sleep arousal or a non-arousal while its training speed is similar to that for the SVM based model.

Chapter 3

Automatic Detection based on C-ELM and MRMR

In this Chapter, we present an improved sleep arousal detection algorithm. In this algorithm, the Minimum Redundancy Maximum Relevance (MRMR) feature selection method is added to the previous mentioned C-ELM classifier based algorithm. In Section.3.1, the improved model is illustrated and a brief introduction is provided. In Section.3.2, various feature selection methods are described and discussed including the MRMR approach utilized by our algorithm. In Section.3.3, the improved algorithm is applied to the sleep dataset. The performance is evaluated including providing the average AUC and ACC of 50 datasets. For each dataset, the AUCs and ACCs are computed when different feature subsets are used for C-ELM based algorithm. A brief summary of this chapter is provided in Section.3.4.

3.1 Improved Arousal Detection Model

This improved model is based on the model proposed in Chapter2. Most of the methods proposed for automatic sleep arousal detection do not have a feature selection step. However, it is reported in [43] has reported that the selection of different feature subsets have a significant influence on the sleep arousal detection. All of the studies mentioned in Section 1.2 select features from previous works or add new features from their own perspectives. The influence of different feature subsets on the performance of arousal detection methods was not reported in these studies. In our model, a ranking of all the 22 features is obtained using the MRMR feature selection method [44]. We applied the C-ELM classifier on different feature subsets according to the ranking. It is found that a subset of the 17 highest ranked features has the best performance and a feature set of the 6 highest ranked features achieves a similar performance with the 17 feature set. Dimensionality reduction of the input vectors can reduce the complexity and training time of model while keeping a reasonable performance. The improved arousal detection model is shown in Fig. 3.1.

3.2 Feature Selection

In machine learning problems, the experimental performances can be negatively influenced by data dimensionality [43]. In some real problems, a small number of high-dimensional data may cause over fitting problem. Although the amount of data needed
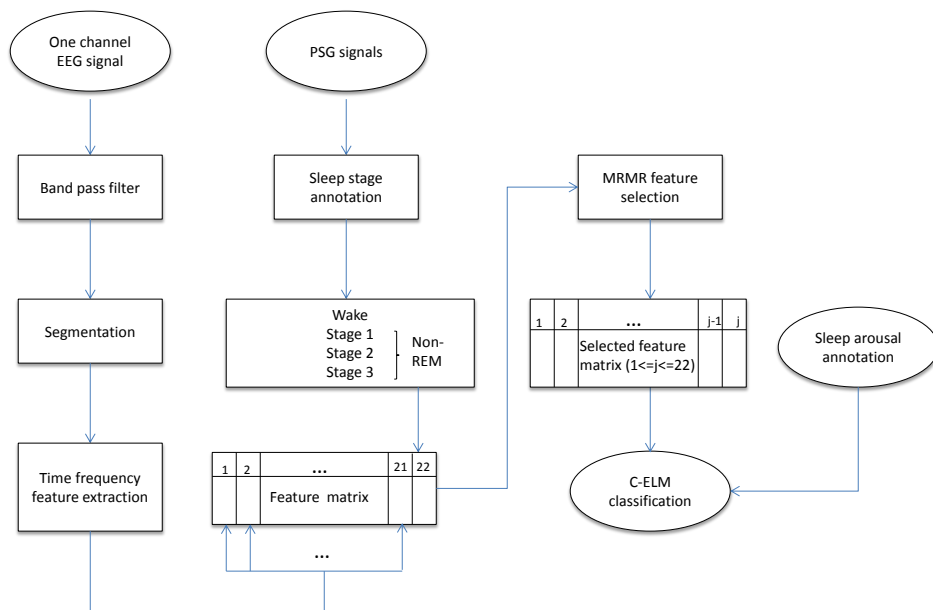
Figure 3.1: Improved sleep detection model based on MRMR feature selection.

to properly train a model may not be obvious, dimensionality reduction of the input data may be of benefit in our study. First, dimensionality reduction of the input data can reduce the complexity of our model and the training time of C-ELM classifier. Second, although we do not have a huge amount of features, EEG signals obtained from patients always have big noise which results in noisy features which may mislead the classification algorithm thereby reduces the accuracy of sleep arousal detection. It is thus important to add a feature selection step in our sleep arousal detection model. In this section, two main categories of feature selection methods are introduced and compared. The MRMR feature selection method which is adopted in our model is briefly described as well.

3.2.1 Filters and Wrappers

Feature selection methods can be roughly grouped into two main categories: filters and wrappers [43]. Filter methods carry out the selection step based on intrinsic characteristics of the training data to determine their relevance or discriminant power with regards to the target classes (the true label for each observation, named positive or negative in binary classifications) [45]. Filter methods are totally independent of classifiers used in the classification step. Wrappers use induction algorithms, e.t. Multi-Layer Perceptron (MLP) and Support Vector Machine (SVM) are used as induction algorithms in [43] to explore each subset of features. During the induction process, wrapper methods are dependent on classifiers.

Filter methods give a ranking of input features or a subset of significant features

based on different measures of the input data and corresponding classes. Filters based on information gain, information entropy, mutual information and statistical tests (such as t-test, F-test) etc. have been developed [43, 45]. Different filter feature selection methods are expected to have significantly different rankings of features since various measures are used. An effective filter method is believed to improve the classification performance while reducing computing time. Another advantage for filter methods is that filters are independent of the learning algorithms.

For wrapper methods, feature selection is wrapped around a learning algorithm such as SVM. The effectiveness of a feature is decided by the estimated accuracy of the learning algorithm. Two well known strategies utilized in wrapper methods are sequential forward selection (SFS) and sequential backward selection (SBS) [43]. SFS starts with an empty feature set and add features one by one while SBS starts with a full feature set and delete features one by one [43]. Wrapper methods can give high accuracy if the learning algorithm used in the classification step is the same one as used in the wrapper. However, wrappers require a long running time when the dataset is big because they need to train an induction algorithm numerous times. Moreover, wrapper methods have lower generalization ability than filters as they depend on the learning algorithm.

In our sleep arousal detection problem, time efficiency is quite important since the dataset obtained from patients is usually quite big. Thus, in our study, a filter feature selection method named minimum redundancy-maximum relevance (MRMR) [45] is adopted instead of wrapper methods.

3.2.2   MRMR Feature Selection

In common filter methods, a simple ranking of features is obtained based on some specific measures (information gain, mutual information etc.). Then we can select m features out of the total n features (m<=n). One deficiency of these approaches is that the selected m features can be correlated among themselves. Thus, redundancy of features can still exist in the feature set although this set has strong predict power in classification. This issue can lead to two main problems [45]. (1) If a feature set contains highly mutual correlated features, then the true unique features are fewer, and some of the features are wasted.(2) The feature set is "narrow" because it can only represent one or a few dominant characteristics of the data which limits the generalization ability of the feature set.

Minimum Redundancy-Maximum Relevance (MRMR) is a filter method which requires features to be maximally dissimilar to each other (minimum redundancy) while keeping the maximum relevance criteria used in other filter methods such as maximizing the mutual information between the features and the target classes. With this approach, a smaller feature set with better representative and generalization properties which reduces complexity of the model may be obtained.

A brief description of MRMR feature selection for discrete variables is provided below. For more details, please refer to [44, 45]. The minimum redundancy condition is

$$\min \boldsymbol{W_I}, \quad \boldsymbol{W_I} = \frac{1}{|\boldsymbol{S}|^2} \sum_{i,j \in \boldsymbol{S}} \boldsymbol{I}(i,j), \qquad (3.1)$$

where $\boldsymbol{I}(i,j)$ is the mutual information between two different features i and j, $\boldsymbol{S}$ is the selected feature set, $|\boldsymbol{S}|$ is the number of features in $\boldsymbol{S}$.

The maximum relevance condition is

$$\max \boldsymbol{V_I}, \quad \boldsymbol{V_I} = \frac{1}{|\boldsymbol{S}|} \sum_{i \in \boldsymbol{S}} \boldsymbol{I}(h,i), \tag{3.2}$$

where $\boldsymbol{I}(h,i)$ is the mutual information between feature i and classes $h = \{h_1, h_2, ..., h_K\}$.

Two schemes - Mutual Information Difference (MID) (Eq. (3.3)) and Mutual Information Quotient (MIQ) (Eq. (3.4)) are utilized to optimize Eq. (3.1) and Eq. (3.2) simultaneously. Optimization of both conditions requires combining them into a single criterion function. Since the two conditions are equally important, two simplest combination criteria (Eq. (3.3) and Eq. (3.4)) are considered [45].

$$\max(\boldsymbol{V_I} - \boldsymbol{W_I}), \tag{3.3}$$

$$\max(\boldsymbol{V_I}/\boldsymbol{W_I}), \tag{3.4}$$

The feature selection step works as follows. The first feature is selected according to Eq. (3.2), i.e. the feature with the highest $\boldsymbol{I}(h,i)$. Earlier features selected remain in the set, a new feature is selected and added according to one of the two criteria (Eq. (3.3) and Eq. (3.4)). That is to say, the last added feature is the one with the lowest rank. Finally, a ranking of all the features can be obtained.

3.3   Performance Evaluation

In this section, we first apply MRMR feature selection on the 7680 input epochs mentioned in Chapter 2. Then a ranking of the 22 features is obtained using the feature selection process. The ranking of the data used in our study is listed in Table. 3.1. Ideally, it is a better way to do the feature selection step using a different dataset from the training and testing datasets. However, in our study, very limit data was used, thus, a sub-optimal way was utilized (we use the 7680 data to do feature selection, and the 50 datasets used to do performance evaluation are also from the 7680 data). For detailed explanation of the features, please refer to Sec.2.3. In our study, both Mutual Information Quotient (MIQ) and Mutual Information Difference (MID) criteria are tried in the feature selection step. It is found that MIQ criterion performs much more effective and thus MIQ criterion is utilized as the MRMR feature selection scheme in our model.

According to the ranking, different feature subsets are used as input feature vectors to train the C-ELM classifier. 50 datasets are randomly chosen using the same steps in Chapter 2 to evaluate the performance (the same logic as Chapter 2). AUC and ACC are used as the criteria for our performance evaluation. The steps for the performance evaluation are described in Algorithm 2.

For each of the 50 datasets, we have 22 AUCs and ACCs computed for 22 feature sets. The first feature set only contains the highest ranked feature; the second feature set contains the top 2 features and so on. The 22nd feature set contains a full set of 22 features. The average performance of the 50 datasets for each feature set is computed.

Table 3.1: MRMR feature ranking using MIQ scheme

Feature rank	Feature name
1	Mean frequency
2	3-sec delta power ratio
3	1-sec alpha power ratio
4	max power frequency
5	1-sec power (0-50 Hz) ratio
6	sleep stage annotation
7	zero-crossing frequency
8	3-sec theta ratio
9	mean value
10	1-sec power
11	3-sec alpha power ratio
12	standard deviation
13	1-sec sigma ratio
14	3-sec power (0-50 Hz) ratio
15	3-sec sigma power ratio
16	3-sec gamma power ratio
17	sleep spindle
18	1-sec beta power ratio
19	3-sec beta power ratio
20	1-sec delta power ratio
21	1-sec theata ratio
22	1-sec gamma ratio

Algorithm 2 The algorithm for the improved model's performance evaluation.

---

for iteration $i := 1$ to 50 **do**

  select 144 negative epochs (each epoch is a 22-dimension vector) randomly from a

  total of 7536 non-arousal epochs. 144 positive epochs are combined with the selected

  negative epochs to form a balanced dataset of 288 epochs in total.

  for iteration $j := 1$ to 22 **do**

    choose the top j features according to the MRMR ranking for each input vector.

    Thus, the current dataset becomes a 288*j input dataset.

    randomize the dataset obtained in the previous step and divide it into 10 folds for

    cross validation. Then C-ELM classifier is applied on the dataset. For each test

    fold, decision values (probabilities of positive or negative) are obtained.

    A ROC curve is plotted according to decision values and AUC, ACC are computed

    from the ROC curve.

  end for

end for

Results are shown in Fig. 3.2 and Fig. 3.3.

From Fig. 3.2, it can be seen that the best average AUC achieved is 0.86 when a set of 17 top-ranking features are selected. We also observe that a set of the 6 top-ranking features can achieve a reasonable good performance with an average AUC of 0.85. Fig. 3.3 shows that the best average ACC (0.80) is obtained with the 6 top ranked features. Using the full set of 22 features, the average AUC is 0.85 while the average ACC is 0.79. Thus, the MRMR feature selection successfully improves the average AUC from 0.85 to 0.86 while reducing the number of features from 22 to 17. Moreover, we can reduce the number of features to 6 and thereby reduce the training time of our model while maintaining a similar performance. In order to find out the training times of dataset with different number of features. Another experiment is done on a training data set of 288 observations (each observation is a m-dimensional input vector, here m is the number of features which is between 1 and 22). It is observed that the training time is 0.043 seconds for the dataset with 6 features while the training time achieves 0.079 secons for the dataset with 22 features. However, the relationship between the training time and the number of features is not a simple linear correlation because the training time depends on the kernel function you used, the convergence time of the method used to find a separating hyperplane (in SVM) and so on. This may be a good topic to work on, however, in our study, it is not the main topic to discuss. The average AUC and ACC of input data with different feature set size are listed in Table. 3.2.

As above-mentioned, we have big noise in EEG signal. Thus, we may have noisy
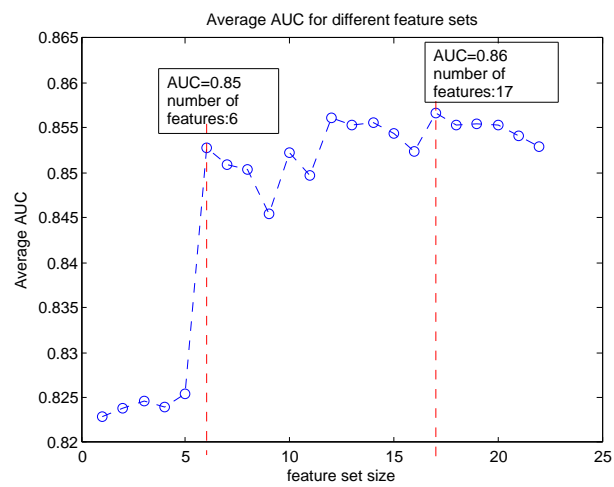
Figure 3.2: Average AUC plot for different feature sets. The x-axis is the feature set size.Feature set of size 1 contains top 1 ranking feature, feature set of size 2 contains 2 top ranked features and so on.

Figure 3.3: Average ACC plot for different feature sets. The x-axis is the feature set size.Feature set of size 1 contains top 1 ranking feature, feature set of size 2 contains 2 top ranked features and so on.
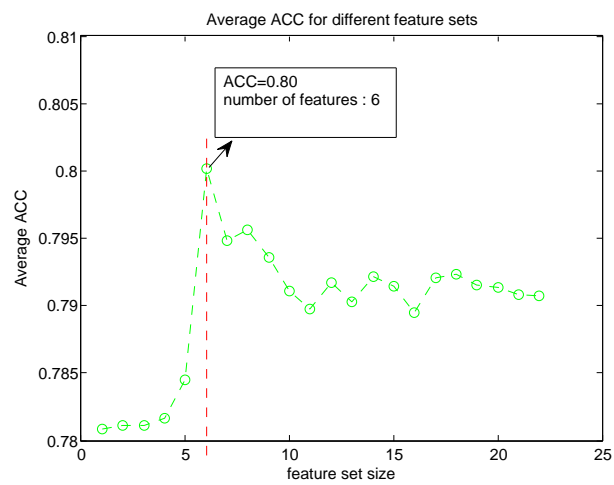
Table 3.2: Average performance of different feature sets using C-ELM

Number of features	Average AUC	Average ACC
1	0.82	0.78
6	0.85	0.80
17	0.86	0.79
22	0.85	0.79

features in the feature set. In our study, we only use one channel of EEG signal, so the predictive power of a feature is important. It is observed from Table. 3.2 that the AUC and ACC of the 6th feature set is higher than those of others which indicates that on top of the 6 top ranked features, the remaining features does not provide as much predict power as the 6 top ranked features. The 6 top features are mean frequency, 3-sec/10-sec delta power ratio, 1-sec/10-sec alpha ratio, max power frequency, 1-sec/10-sec power ratio (0-50 Hz), and sleep stage annotations. From the 6 features we can see that power ratios have made good contributions to a effective feature set which indicates the power changes could represent EEG frequency shift to some extent.

As a supplementary, we also applied MRMR feature selection to the SVM based model. The best average AUC of 0.69 and ACC of 0.67 are achieved by the set of 6 top ranked features (the feature ranking is the same as above-mentioned), while the average AUC and ACC for a full set of features are the same as those of the set with 6 features. part of the AUCs and ACCs for different feature sets are listed in Table. 3.3. This result indicates a good generalization of the 6 top ranked features regardless of the classifier.

Table 3.3: Average performance of different feature sets using SVM

Number of features	Average AUC	Average ACC
1	0.65	0.65
6	0.69	0.67
22	0.69	0.67

3.4 Summary

In this Chapter, an improved non-REM sleep arousal detection model with a Minimum Redundancy- Maximum Relevance (MRMR) feature selection step is presented. Several feature selection methods are briefly introduced, followed by an illustration of the MRMR feature selection. The performance of this proposed model with MRMR is evaluated using the AUC and ACC criteria. It is found that the sleep arousal detection model can provide a similar performance with a reduced feature set size. It should be noted that only one single EEG channel and limited data were used in our simulations. Better performance may be achieved if the model is trained on a larger dataset or more channels of physiological signals are utilized.

Chapter 4

Conclusions and Future Work

In this chapter, we conclude this thesis by summarizing the research results and contributions. Future research topics are suggested as well.

4.1   Conclusions

We studied the problem of automatic detection of sleep arousals. A new detection model was proposed based on a set of 22 features and Curious Extreme Learning Machine (C-ELM) in Chapter 2. This model was found to provide a good performance on the dataset used in our study. A Support Vector Machine (SVM) based model was also evaluated for comparison with our model. In Chapter 3, an improved detection model was presented, in which a Minimum Redundancy Maximum Relevance (MRMR) feature selection step is added to the model proposed in Chapter 2. The improved model allows a reduction of the size of the feature set, and have a decreased training time while maintaining a similar detection performance.

- In Chapter.2, we presented a new model for sleep arousal detection. In this model, data was first preprocessed and segmented. Then, a proposed set of 22 features are

extracted, followed by the use of a Curious Extreme Learning Machine (C-ELM) classifier. An average Area Under the ROC Curve (AUC) of 0.85 (an AUC of 1 corresponds to perfect classification, whereas an AUC of 0.5 corresponds to random classification) and an average accuracy (ACC) of 0.79 was achieved by the proposed model while an average AUC and ACC of 0.69 and 0.67 respectively was achieved for the SVM based model. The results indicates that our system for sleep arousal detection has a high performance on the dataset utilized in this study. In addition, the training speed of C-ELM is similar to that of SVM and the total executing time of our model is less than SVM based model since a grid search was done to optimize the SVM based model which increased the total running time. The detailed running time of the SVM based model varies a lot based on the complexity of grid search.

- In Chapter.3, we proposed an improved model for sleep arousal detection based on the model proposed in Chapter.2 by adding a Minimum Redundancy Maximum Relevance feature selection step to remove redundant features. Using this improved model, it was found that the size of the feature set could be reduced to 6 from 22 without a significant performance change while reducing the training time of the classifier. The average AUC and ACC achieved by a 6-feature model are 0.85 and 0.80 respectively while the average AUC and ACC obtained by a model with the full set of 22 features were 0.85 and 0.79 respectively. The best average AUC achieved was 0.86 with an average ACC of 0.79 when using a set of 17 features. The results of this chapter suggests adding an effective feature selection step (such as MRMR

feature selection) in automatic sleep arousal detection system is significant.

4.2   Future Work

Some possible extensions of the research work on automatic sleep arousal detection are outlined below based on what have been observed and learnt in this project.

- The data used in our study is limited. A better performance may be expected if we have a larger dataset to train the classifier. It would be interesting to apply the proposed model on a big dataset obtained from real patients.

- During one patient's overnight sleep, arousal events happen frequently. However, the events usually only last several seconds and the total time of all arousals during a night is quite short, typically 20 or 30 minutes out of 8 hours. Thus, the imbalance between positive and negative data is a big issue no matter in training a classifier or doing a performance evaluation. In our study, we simply choose datasets with equal number of positive and negative data. Other methods can be explored to solve the imbalance problem such as using over or under sampling strategies [46].

- In our study, we applied a bandpass filter to the dataset for preprocessing in our model. EEG signals can have a lot of noise caused by the movement during sleep. A considerable range of methods have been proposed to remove artifacts if multi-channel EEG recordings are used. However, few methods have been proposed to

remove artifacts of a single channel EEG [47]. Thus, studies on artifact removal for a single channel EEG would be useful for sleep arousal detection.

- In our work, we added a MRMR feature selection step to the model proposed and have observed a good performance. This result indicates the significance of feature selection step. Thus, how to choose an effective feature selection method for sleep arousal detection would be another interesting topic.

Bibliography

[1] A. L. Goldberger, L. A. N. Amaral, *et al.*, "PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, pp. e215–e220, 2000 (June 13), circulation Electronic Pages: http://circ.ahajournals.org/cgi/content/full/101/23/e215 PMID:1085218; doi: 10.1161/01.CIR.101.23.e215. [Online]. Available: http://physionet.org/cgi-bin/atm/ATM

[2] Wikipedia, "Sleep — wikipedia the free encyclopedia," 2015, online; accessed 22-June-2015. [Online]. Available: https://en.wikipedia.org/w/index.php?title=Sleep&oldid=667626425

[3] ——, "Receiver operating characteristic — wikipedia the free encyclopedia," 2015, [Online; accessed 22-June-2015]. [Online]. Available: https://en.wikipedia.org/w/index.php?title=Receiver_operating_characteristic&oldid=667584494

[4] "Report of the committee on methods of clinical examination in electroencephalography: 1957," *Electroencephalography and Clinical Neurophysiology*, vol. 10, no. 2, pp. 370–375, 1958. [Online]. Available: http://www.sciencedirect.com/science/article/pii/0013469458900531

[5] P. Halsz, M. Terzano, L. Parrino, and R. Bdizs, "The nature of arousal in sleep," *Journal of Sleep Research*, vol. 13, no. 1, pp. 1–23, 2004. [Online]. Available: http://dx.doi.org/10.1111/j.1365-2869.2004.00388.x

[6] M. Bonnet, D. Carley, *et al.*, "EEG arousals: Scoring rules and examples: A preliminary report from The Sleep Disorders Atlas Task Force Of The American Sleep Disorders Association," *Sleep*, vol. 15, pp. 173–184, 1992.

[7] A. Rechtschaffen, A. Kales, *et al.*, *A Manual of Standardized Terminology, Techniques and Scoring System for Sleep Stages of Human Subjects*, ser. Publication. Brain Information Service/Brain Research Institute, University of California, 1968. [Online]. Available: https://books.google.ca/books?id=wzdRnQEACAAJ

[8] L. Gennaro, M. Ferrara, and M. Bertini, "EEG arousals in normal sleep: Variations induced by total and selective slow-wave sleep deprivation," *Sleep*, vol. 24, no. 6, 2001.

[9] J. Fleetham, P. West, *et al.*, "Sleep, arousals, and oxygen desaturation in chronic obstructive pulmonary disease. the effect of oxygen therapy," *The American review of respiratory disease*, vol. 126, no. 3, p. 429433, September 1982. [Online]. Available: http://europepmc.org/abstract/MED/7125332

[10] C. Guilleminault, R. Stoohs, *et al.*, "A cause of excessive daytime sleepiness. the upper airway resistance syndrome." *Chest*, vol. 104, no. 3, pp. 781–787, 1993. [Online]. Available: +http://dx.doi.org/10.1378/chest.104.3.781

[11] K. Cheshire, H. Engleman, *et al.*, "Factors impairing daytime performance in patients with sleep apnea/hypopnea syndrome," *Archives of Internal Medicine*, vol. 152, no. 3, pp. 538–541, 1992.

[12] P. Collard, M. Dury, *et al.*, "Movement arousals and sleep-related disordered breathing in adults." *American Journal of Respiratory and Critical Care Medicine*, vol. 154, no. 2, pp. 454–459, 1996.

[13] M. Bonnet, K. Doghramji, *et al.*, "The scoring of arousal in sleep: reliability, validity, and alternatives," *J Clin Sleep Med*, vol. 3, no. 2, pp. 133–145, 2007.

[14] M. Drinnan, A. Murray, *et al.*, "Automated recognition of EEG changes accompanying arousal in respiratory sleep disorders," *Sleep*, vol. 19, no. 4, p. 296303, May 1996. [Online]. Available: http://europepmc.org/abstract/MED/8776787

[15] S.-P. Cho, J. Lee, H. Park, and K. Lee, "Detection of arousals in patients with respiratory sleep disorders using a single channel EEG," in *Engineering in Medicine and Biology Society, 2005. IEEE-EMBS 2005. 27th Annual International Conference of the*, Jan 2005, pp. 2733–2735.

[16] D. J. Buysse *et al.*, "The Pittsburgh sleep quality index: A new instrument for psychiatric practice and research," *Psychiatry Research*, vol. 28, no. 2, pp. 193 – 213, 1989. [Online]. Available: http://www.sciencedirect.com/science/article/pii/0165178189900474

[17] G. Sorensen, J. Kempfner, P. Jennum, and H. Sorensen, "Detection of arousals in parkinson's disease patients," in *Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE*, Aug 2011, pp. 2764–2767.

[18] R. J. Thomas, "Arousals in sleep-disordered breathing: patterns and implications," *Sleep*, vol. 26, no. 8, pp. 1042–1048, 2003.

[19] M. Grigg-Damberger, D. Gozal, *et al.*, "The visual scoring of sleep and arousal in infants and children," *J Clin Sleep Med*, vol. 3, no. 2, pp. 201–240, 2007.

[20] T. Sugi, F. Kawana, and M. Nakamura, "Automatic EEG arousal detection for sleep apnea syndrome," *Biomedical Signal Processing and Control*, vol. 4, no. 4, pp. 329 – 337, 2009, special Issue on Biomedical Systems, Signals and Control Extended Selected papers from the {IFAC} World Congress, Seoul, July 2008. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1746809409000494

[21] R. Agarwal, "Automatic detection of micro-arousals," *IEEE Engineering in Medicine and Biology Society.Annual Conference*, vol. 2, p. 1158, 2005.

[22] F. De Carli, L. Nobili, P. Gelcich, and F. Ferrillo, "A method for the automatic detection of arousals during sleep," *Sleep*, vol. 22, no. 5, p. 561, 1999.

[23] D. Álvarez Estévez and V. Moret-Bonillo, "Model comparison for the detection of EEG arousals in sleep apnea patients," in *Proceedings of the 10th Interna-*

tional Work-Conference on Artificial Neural Networks: Part I: Bio-Inspired Systems: Computational and Ambient Intelligence, ser. IWANN '09, 2009, pp. 997–1004.

[24] O. Shmiel, T. Shmiel, *et al.*, "Data mining techniques for detection of sleep arousals," *Journal of Neuroscience Methods*, vol. 179, no. 2, pp. 331 – 337, 2009. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0165027009000715

[25] G. Pillar, A. Bar, *et al.*, "Autonomic arousal index: an automated detection based on peripheral arterial tonometry." *Sleep*, vol. 25, no. 5, pp. 543–549, 2002.

[26] ——, "An automatic ambulatory device for detection of AASM defined arousals from sleep: the WP100," *Sleep Medicine*, vol. 4, no. 3, pp. 207–212, 2003.

[27] G. H. Klem, H. O. Lüders, H. Jasper, and C. Elger, "The ten-twenty electrode system of the International Federation of Clinical Physiology," *Electroencephalogr Clin Neurophysiol*, vol. 52, no. suppl. 52, p. 3, 1999.

[28] Q. Wu and C. Miao, "C-ELM: A curious extreme learning machine for classification problems," in *Proceedings of ELM-2014 Volume 1*, ser. Proceedings in Adaptation, Learning and Optimization. Springer International Publishing, 2015, vol. 3, pp. 355–366. [Online]. Available: http://dx.doi.org/10.1007/978-3-319-14063-6_30

[29] C.-W. Hsu, C.-C. Chang, *et al.*, "A practical guide to support vector classification," 2003. [Online]. Available: http://www.csie.ntu.edu.tw/~cjlin/papers.html

[30] T. Joachims, *Introduction to support vector machines.* Cambridge University Press, 2002.

[31] N. Cristianini and J. Shawe-Taylor, *An introduction to support vector machines and other kernel-based learning methods.* Cambridge University Press, 2000.

[32] D. Fradkin and I. Muchnik, "Support vector machines for classification," *Discrete methods in epidemiology*, vol. 70, pp. 13–20, 2006.

[33] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "Extreme learning machine: theory and applications," *Neurocomputing*, vol. 70, no. 1, pp. 489–501, 2006.

[34] Q. Yuan, W. Zhou, S. Li, and D. Cai, "Epileptic EEG classification based on extreme learning machine and nonlinear features," *Epilepsy research*, vol. 96, no. 1, pp. 29–38, 2011.

[35] X.-K. Wei, Y.-H. Li, and Y. Feng, "Comparative study of extreme learning machine and support vector machine," in *Advances in Neural Networks-ISNN 2006.* Springer, 2006, pp. 1089–1095.

[36] R. Rajesh and J. S. Prakash, "Extreme learning machines: a review and state-of-the-art," *International Journal of Wisdom Based Computing*, vol. 1, no. 1, pp. 35–49, 2011.

[37] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "Extreme learning machine: a new learning scheme of feedforward neural networks," in *Neural Networks, 2004. Proceedings. 2004 IEEE International Joint Conference on*, vol. 2. IEEE, 2004, pp. 985–990.

[38] G.-B. Huang, "Extreme learning machine: learning without iterative tuning," 2010. [Online]. Available: http://www.extreme-learning-machines.org/

[39] S. Suresh, S. Saraswathi, and N. Sundararajan, "Performance enhancement of extreme learning machine for multi-category sparse data classification problems," *Engineering Applications of Artificial Intelligence*, vol. 23, no. 7, pp. 1149–1157, 2010.

[40] Z. Bai, G.-B. Huang, *et al.*, "Sparse extreme learning machine for classification," *Cybernetics, IEEE Transactions on*, vol. 44, no. 10, pp. 1858–1870, 2014.

[41] C. C. Chang and C. J. Lin, "LIBSVM: A library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, pp. 27:1–27:27, 2011, software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm.

[42] M. Law, "A simple introduction to Support Vector Machines," *Dept. of Computer Science and Eng. Michigan State Univ.* [Online]. Available: http://www.cise.ufl.edu/class/cis4930sp11dtm/notes/intro_svm_new.pdf

[43] D. lvarez Estvez, N. Snchez-Maroo, *et al.*, "Reducing dimensionality in a database of sleep EEG arousals," *Expert Systems with Applications*, vol. 38, no. 6, pp. 7746 – 7754, 2011. [Online]. Available: http://www.sciencedirect.com/science/article/pii/ S0957417410014971

[44] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 27, no. 8, pp. 1226–1238, 2005.

[45] C. Ding and H. Peng, "Minimum redundancy feature selection from microarray gene expression data," *Journal of bioinformatics and computational biology*, vol. 3, no. 02, pp. 185–205, 2005.

[46] N. V. Chawla, "Data mining for imbalanced datasets: An overview," in *Data mining and knowledge discovery handbook*. Springer, 2005, pp. 853–867.

[47] C. H. Xun Chen and H. Peng, "Removal of muscle artifacts from single-channel EEG based on ensemble empirical mode decomposition and multiset canonical correlation analysis," *Journal of Applied Mathematics*, p. 10, 2014.

[48] Wikipedia, "10-20 system (eeg) — wikipedia the free encyclopedia," 2015, [Online; accessed 22-June-2015]. [Online]. Available: https://en.wikipedia.org/w/index.php?title=10-20_system_(EEG)&oldid=664716238

Appendices

Appendix A

Sleep Stages

Usually sleepers pass through five sleep stages: 1, 2, 3, 4 and REM (rapid eye movement) sleep. Sleep stage 3 and 4 are always combined into one sleep stage (we use sleep stage 3 to represent sleep stage 3 and 4). Sleep stage 1 to 3 are called non-REM sleep stages. These stages progress cyclically from 1 through REM stage. An overnight sleep of a healthy individual usually contains 4 to 5 sleep cycles as shown in Fig. A.1. Sleep stage 1 is known as a transitional stage usually occurs between sleep and wakefulness. In this stage, the brain produces high amplitude and low frequency theta waves. Brain waves during Sleep stage 2 are mainly in the theta wave range. This sleep stage is characterized by two phenomena: sleep spindles and K-Complex. Sleep stage 3 is known as slow wave sleep or deep sleep characterized by delta wave along with sleep spindles, although much fewer than sleep stage 2. REM sleep stage is a stage during which EOG shows a rapid eye movement. Dreams often occur in this stage.
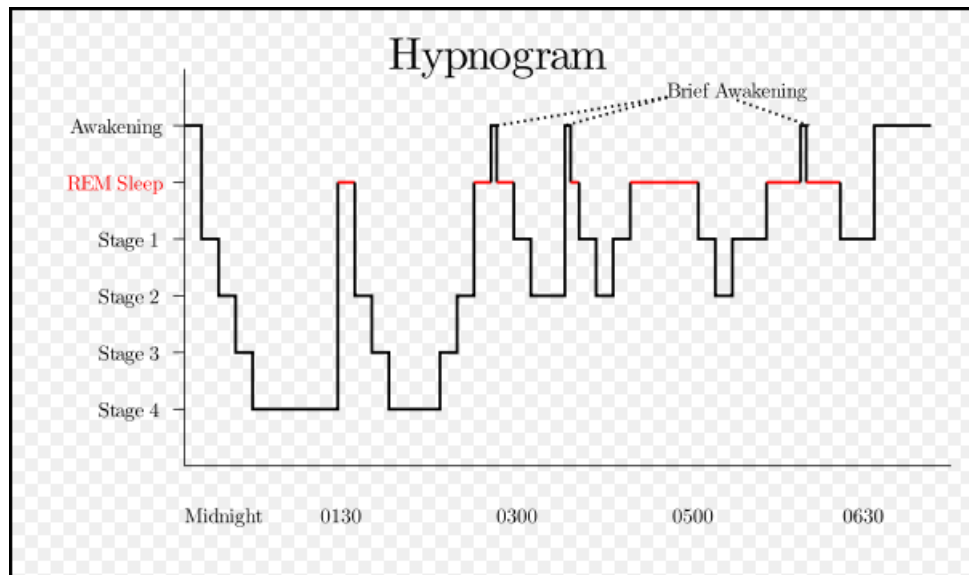
Figure A.1: Sleep stages [2]

Appendix B

Basic Concept (sensitivity, etc.)

In a two-class prediction problem (binary classification), in which the outcomes are labeled either as positive (p) or negative (n). There are four possible outcomes from a binary classifier. If the outcome from a prediction is p and the actual value is also p, then it is called a true positive (TP); however if the actual value is n then it is said to be a false positive (FP). Conversely, a true negative (TN) has occurred when both the prediction outcome and the actual value are n, and false negative (FN) is when the prediction outcome is n while the actual value is p [3]. Please see Fig. B.1.

Please see Table. B.1 for computing sensitivity, specificity, accuracy, selectivity, positive predictive value (PPV) and so on.

		Condition (as determined by "Gold standard")	
Total population		Condition positive	Condition negative
Test outcome	Test outcome positive	True positive	False positive (Type I error)
	Test outcome negative	False negative (Type II error)	True negative

Figure B.1: Binary classification basic concept [3]

Table B.1: Computing accuracy, sensitivity and so on

The name of the measure	Computing formula
Accuracy (ACC)	(TP+TN)/Total population
True positive rate (TPR), Sensitivity	TP/Condition positive
True negative rate (TNR), Specificity	TN/Condition negative
False positive rate (FPR), Fall-out	FP/Condition negative
False negative rate (FNR), Miss rate	FN/Condition positive
Positive predictive value (PPV), Precision	TP/Test outcome positive
Selectivity	TP/(TP+FP)

Appendix C

EEG 10–20 International System

The 10-20 system or International 10-20 system is an internationally recognized method to describe and apply the location of scalp electrodes in the context of an EEG test or experiment. This method was developed to ensure standardized reproducibility so that a subject's studies could be compared over time and subjects could be compared to each other. This system is based on the relationship between the location of an electrode and the underlying area of cerebral cortex. The "10" and "20" refer to the fact that the actual distances between adjacent electrodes are either 10% or 20% of the total front-back or right-left distance of the skull [4, 27, 48]. Please see Fig. C.1.
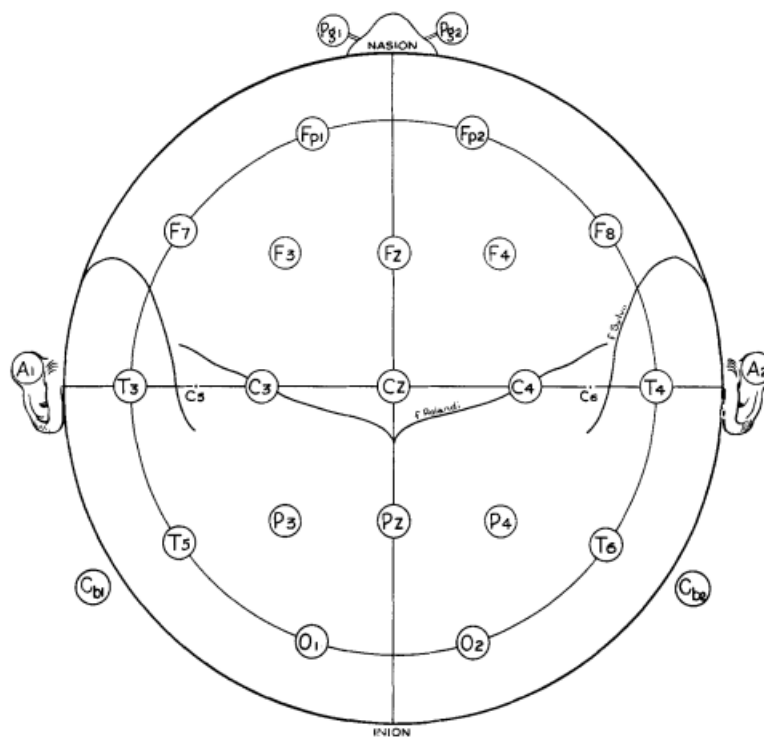
Figure C.1: EEG 10-20 system [4]

Appendix D

ROC Curve

A receiver operating characteristic (ROC), or ROC curve, is a graphical plot that illustrates the performance of a binary classifier system as its discrimination threshold is varied. The curve is created by plotting the True positive rate (Sensitivity) against the False positive rate (1-Specificity) at various threshold settings [3]. The Area Under the ROC Curve indicates the ability of a classifier to discriminate a positive data from a negative data. A value of 1 means a perfect test while 0.5 means random classification.