Automatic Sleep Arousal Detection based on C-ELM

Yuemeng Liang*, Cyril Leung†, Chunyan Miao‡, Qiong Wu§ and Martin J. McKeown¶
*†*Department of Electrical and Computer Engineering*
The University of British Columbia (UBC), Vancouver, Canada
†‡§¶*Joint NTU-UBC Research Centre of Excellence in Active Living for the Elderly*
Nanyang Technological University (NTU), Singapore
¶*Pacific Parkinson's Research Centre and Division of Neurology, UBC*
Email: {yuemengl,cleung}@ece.ubc.ca,{ascymiao,wu.qiong}@ntu.edu.sg,martin.mckeown@ubc.ca

***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 time consuming and cumbersome, and manual scoring results can vary widely among different expert scorers. This paper reports the design and performance evaluation of an effective and efficient method to automatically detect sleep arousals using a single channel EEG. 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 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. The best AUC from among the 50 datasets used was 0.88. In comparison, the average AUC and ACC of a Support Vector Machine (SVM) based model were 0.69 and 0.67 respectively, and the best AUC from among the same 50 datasets used was 0.88. This indicates that the proposed C-ELM based model works well for the sleep arousal detection problem.

***Keywords*-Sleep Arousal Detection, Curious Extreme Learning Machine**

I. INTRODUCTION

Sleep problems are a frequent complaint among many people, especially the elderly, and have a substantial impact on the 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 spontaneous vigilance level oscillations [1]. 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 [1].

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 [2]. There are several disadvantages for this kind of traditional sleep test. Firstly, 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. Secondly, visual inspection is a relatively subjective way to diagnose sleep arousals. There can be large differences between individual sleep experts. Thirdly, it has been suggested that arousals of short durations (less than 3 seconds) may also be significant [2]. However, identification and agreement on events of such short durations are difficult to achieve, if scored manually. Lastly, from a patient's perspective, the cost for a PSG test is high, ranging from $700 to $6000.

Due to the above mentioned disadvantages, a lot of research interests have been shown on fast, accurate computer-aided automatic arousal detection approaches and portable, less obtrusive detection devices [3]. 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. A large number of channels of information collected means more inconvenience for patients since more electrodes need to be placed on them. In order to manufacture portable, less obtrusive devices, the use of fewer electrodes is desirable. However, fewer channels of information lead to lower accuracy. Thus, methods which can achieve relatively high accuracy with less physiological information collected should be studied.

In this paper, we propose an automatic sleep arousal detection algorithm based on a recently proposed fast classifier named Curious Extreme Learning Machine (C-ELM) [4]. The widely used Support Vector Machine (SVM) classifier [5] 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 the C-ELM-based and the SVM-based detection algorithms. 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 the SVM-based model.

The rest of the paper is organized as follows: in Section II we discuss the related works and their drawbacks. After that, in Section III, we introduce the proposed sleep arousal detection model. Section IV introduce the data preprocessing and segmentation methods. Section V discusses the classification method based on C-ELM. In Section VI, we present the experiments and evaluation results. The paper is concluded in Section VII.

II. Related Works

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 reliability and validity have been developed and discussed in [2].

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 [3]. 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 [6], [7]. In [7], 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 [7] that the sensitivity and specificity are 0.80 and 0.79 respectively and area under ROC curve (AUC) is around 0.87. They only reported the results of patients with at least 20 arousals/hours [7].

In [8], 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. An automatic detection method of EEG arousals is described in [9]. 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. In [6], 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. Another automatic detection method based on the idea of segmentation, spectral feature extraction, statistical methods and decisional rules is described in [10]. Two EEG channels of 2 patients' PSG recordings were utilized and three sleep experts scored the sleep arousal events in this study.

To our knowledge, none of the works has reported a comparison between its own result and that of other works. 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. Second, different physiological signals are used in different studies. 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. 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.

III. 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 [11]. 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.1.

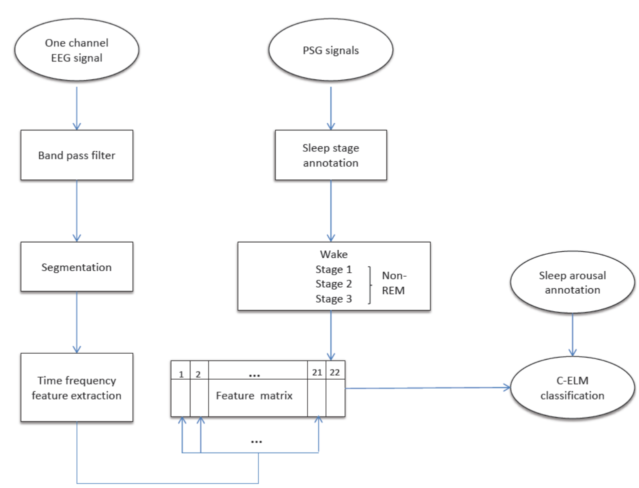

Figure 1: Sleep arousal detection model.

IV. Data Preprocessing and Segmentation

A. 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 paper. The EEG raw data is downloaded from Sleep Heat Health Study (SHHS) PSG DataBase of PhysioBank [11]. The SHHS is a prospective cohort study designed to investigate the relationship between sleep disordered breathing and cardiovascular disease. The age of the patient 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 [12], 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 [12]. Therefore 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.

B. Band Pass Filter and Segmentation

According to [3], 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 [9], 16-40 Hz [10] or >13 Hz [8]. In [12], 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 [3]. 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.

C. 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. 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 [12], 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 [3], we make the "future" window 1 second in length and we also choose another "future" window of 3 seconds according to [13]. 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 [12]. In [12], 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 [12] 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 [3].

Mean Frequency: The signal's mean frequency of each 1-second epoch is extracted as a feature [3]. The mean frequency is computed as follows [9].

$$\bar{f} = \frac{\sum p_i \times f_i}{\sum p_i} \tag{1}$$

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 [12].

Sleep Stages: Although it is widely accepted that the scoring of sleep arousals is independent of Rechtschaffen and Kales criteria [14], 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 paper, the annotations of sleep stages are downloaded from Physionet [11] which is manually scored by sleep experts.

V. Classification based on C-ELM

In this section, the Curious Extreme Learning Machine (C-ELM) algorithm is briefly described.

Extreme Learning Machine (ELM) is a fast, easy-to-implement machine learning algorithm based on a single hidden layer feedforward neural network (SLFN). It has been reported to have good performance and generalization ability [15]. Details about ELM and related algorithms based on ELM can be found in [15]. Curious Extreme Learning Machine (C-ELM) is a curiosity driven algorithm based on ELM, which follows psychological theory of curiosity and performs curiosity appraisal towards each input data. The algorithm has four variables (novelty $\mathcal{N}(x^t)$, uncertainty $\mathcal{U}(x^t)$, conflict $\mathcal{C}(x^t)$ and surprise $\mathcal{S}(x^t)$) and three learning strategies (neuron addition, neuron deletion and parameter update). The four variables are computed for each input vector 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 [4] are briefly summarized below.

Neuron Addition Strategy: Given an input x^t, the neuron addition condition is:

$$\mathcal{N}(\mathbf{x}^t) > \theta_{\mathcal{N}^{add}} \;\; \textbf{AND} \;\; \mathcal{U}(\mathbf{x}^t) > \theta_{\mathcal{U}} \;\; \textbf{AND} \;\; \mathcal{S}(\mathbf{x}^t) > \theta_{\mathcal{S}}, \tag{2}$$

where $\theta_{\mathcal{N}^{add}}$, $\theta_{\mathcal{U}}$, and $\theta_{\mathcal{S}}$ are neuron addition thresholds for novelty, uncertainty, and surprise, respectively. If these parameters are chosen close to 1, then very few input data can trigger the neuron addition strategy and the network cannot approximate the decision function accurately. If these parameters are chosen close to 0, then many input data can trigger the neuron addition strategy, leading to poor generalization ability. In general, $\theta_{\mathcal{N}^{add}}$ is chosen in the range of [0.1,0.5], $\theta_{\mathcal{U}}$ is chosen in the range of [0.1,0.3], and $\theta_{\mathcal{S}}$ is chosen in the range of [0.2,0.9].

Neuron Deletion Strategy: Given an input x^t, the neuron deletion condition is:

$$\mathcal{S}(\mathbf{x}^t) > \theta_{\mathcal{S}} \;\; \textbf{AND} \;\; \mathcal{C}(\mathbf{x}^t) > \theta_{\mathcal{C}} \;\; \textbf{AND} \;\; \mathcal{N}(\mathbf{x}^t) < \theta_{\mathcal{N}^{del}}, \tag{3}$$

where $\theta_{\mathcal{S}}$, $\theta_{\mathcal{C}}$, and $\theta_{\mathcal{N}^{del}}$ are neuron deletion thresholds for surprise, conflict, and novelty, respectively. When $\theta_{\mathcal{S}}$ and $\theta_{\mathcal{C}}$ are chosen close to 1 and $\theta_{\mathcal{N}^{del}}$ is chosen close to 0, then very few input data can trigger neuron deletion, leading to poor generalization ability. When $\theta_{\mathcal{S}}$ and $\theta_{\mathcal{C}}$ are chosen close to 0 and $\theta_{\mathcal{N}^{del}}$ is chosen close to 1, then many input data can trigger neuron deletion and the network cannot approximate the decision function accurately. In general, $\theta_{\mathcal{S}}$ is chosen in the range of [0.2,0.9], $\theta_{\mathcal{C}}$ is chosen in the range of [0.1,0.3], and $\theta_{\mathcal{N}^{del}}$ is chosen in the range of [0.1,0.8].

Parameter Update Strategy: When both neuron addition and deletion conditions are 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 pseudo-code for the C-ELM learning algorithm is given in Algorithm 1.

Step 1. Present a sample $(\mathbf{x}^t, c^t)$.
Step 2. Compute four variables ($\mathcal{N}(\mathbf{x}^t)$, uncertainty $\mathcal{U}(\mathbf{x}^t)$, conflict $\mathcal{C}(\mathbf{x}^t)$, surprise $\mathcal{S}(\mathbf{x}^t)$) according to the input vector.
Step 3. Select one learning strategy out of three (Neuron Addition, Neuron Deletion, Parameter Update) based on the four variables and corresponding thresholds.
Step 4. Increment $t = t + 1$, repeat Step 1 to Step 3.

Algorithm 1: Pseudo-code for C-ELM classifier

VI. Experiment

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 [12]). 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 work, the Library of Support Vector Machine (LIBSVM) [16] 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. 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 [4]. The parameters used are the ones that provided the best classification performance in previous experiments according to [4]. The learning thresholds are set as follows.

- The low threshold of novelty = 0.1;
- The high threshold of novelty = 0.4;
- The uncertainty threshold = 0.1;
- The conflict threshold = 0.3;
- The surprise threshold = 0.4;

A. 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 I and Table II, respectively. The standard deviation of the 50 AUC and ACC results are also listed in Table I and Table II. The best C-ELM based result and its corresponding SVM result are summarized in Table III and the ROC curves are plotted in Fig 2. The best SVM based result and its corresponding C-ELM result are summarized in Table IV and the ROC curves are plotted in Fig. 3.

According to the results shown in two tables, an average AUC of 0.85 and ACC of 0.79 are achieved by our C-ELM based model while an average AUC of 0.69 and ACC of 0.67 are obtained by SVM based model. These results indicates the sleep 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 [4]. In [4], 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 0.02 while the SVM based model reaches 0.16. We can see that the best AUC achieved by C-ELM based model is around 0.89 from Table III and Fig. 2. The similarity between the best result and the average AUC 0.85 and a relative small standard deviation of 0.02 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 0.67. It is much smaller than the best AUC which is around 0.89. We 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. The possible reasons for the relatively poor performance of SVM

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

Properties	C-ELM model	SVM model
Average AUC	0.85	0.69
Standard deviation	0.02	0.16

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

Properties	C-ELM model	SVM model
Average ACC	0.79	0.67
Standard deviation	0.02	0.12

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

Properties	Bset C-ELM Performance	SVM performance
AUC	0.88	0.83

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

Properties	Bset SVM Performance	C-ELM performance
AUC	0.89	0.86

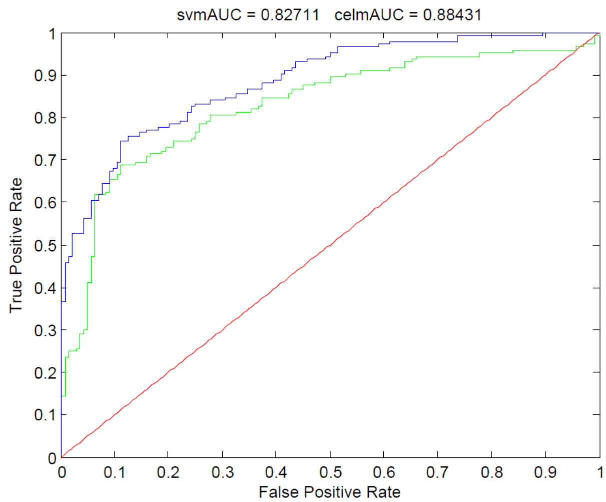

Figure 2: ROC curves for C-ELM based and SVM based models for the dataset which gives the highest AUC for C-ELM. The red line is a random classification, the blue line is the curve of C-ELM and the green line is the curve of SVM.

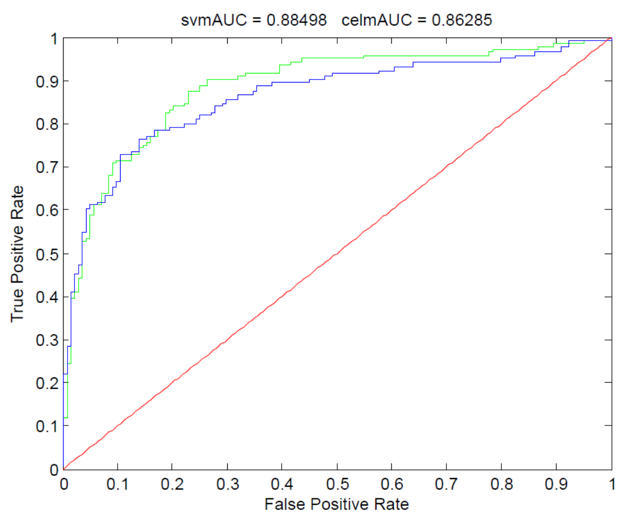

Figure 3: ROC curves for C-ELM based and SVM based models for the dataset which gives the highest AUC for SVM. The red line is a random classification, the blue line is the curve of C-ELM and the green line is the curve of SVM

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.

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 [17]. RBF kernel function is used in SVM classifier in our study because this kernel function is always a good try in various problems [17]. 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.

C-ELM is based on 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 [15]. 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. 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 [4]. 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.

B. *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 [16] 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 [4]. In [4], 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 paper, 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 work.

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

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

VII. Conclusion

In this paper, a new detection model for sleep arousal, based on a new set of 22 features and Curious Extreme Learning Machine (C-ELM), has been proposed. 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 when only one single EEG channel and limited data is available. The proposed model has a higher AUC and ACC which indicates a better ability to correctly classify a random data as a sleep arousal or a non-arousal while its training speed is similar to that of the SVM based model.

VIII. Acknowledgments

This work was supported in part by the Natural Sciences and Engineering Research Council (NSERC) of Canada under Grant RGPIN 1731-2013, by the UBC Faculty of Applied Science, by the National Research Foundation, Prime Minister's Office, Singapore under its IDM Futures Funding Initiative and administered by the Interactive and Digital Media Programme Office.

References

[1] 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.

[2] M. Bonnet, K. Doghramji, T. Roehrs, E. Stepanski, S. Sheldon, A. Walters, M. Wise, and A. Chesson Jr, "The scoring of arousal in sleep: reliability, validity, and alternatives," *J Clin Sleep Med*, vol. 3, no. 2, pp. 133–145, 2007.

[3] 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 Proceedings of the International Conference on Engineering in Medicine and Biology Society*, pp. 2733–2735, 2006.

[4] Q. Wu and C. Miao, "C-elm: A curious extreme learning machine for classification problems," *In Proceedings of the Internationl Conference on Extreme Learning Machines*, pp. 355–366, 2015.

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

[6] 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.

[7] G. Pillar, A. Bar, A. Shlitner, R. Schnall, J. Shefy, and P. Lavie, "Autonomic arousal index: an automated detection based on peripheral arterial tonometry.," *Sleep*, vol. 25, no. 5, pp. 543–549, 2002.

[8] O. Shmiel, T. Shmiel, Y. Dagan, and M. Teicher, "Data mining techniques for detection of sleep arousals," *Journal of neuroscience methods*, vol. 179, no. 2, pp. 331–337, 2009.

[9] 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, pp. 561–572, 1999.

[10] R. Agarwal, "Automatic detection of micro-arousals," *In Proceedings of Annual International Conference of the Engineering in Medicine and Biology Society*, pp. 1158–1161, 2006.

[11] A. L. Goldberger, L. A. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley, "Physiobank, physiotoolkit, and physionet components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, pp. e215–e220, 2000.

[12] M. Bonnet and D. Carley, "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, no. 2, pp. 173–184, 1992.

[13] D. lvarez Estvez and V. Moret-Bonillo, "Model comparison for the detection of eeg arousals in sleep apnea patients," *Springer*, pp. 997–1004, 2009.

[14] A. Rechtschaffen and A. Kales, "A manual of standardized terminology, techniques and scoring system for sleep stages of human subjects," 1968.

[15] 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.

[16] C.-C. Chang and C.-J. Lin, "Libsvm: A library for support vector machines," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 2, no. 3, p. 27, 2011.

[17] M. Law, "A simple introduction to support vector machines," *Lecture for CSE*, vol. 802, 2006.