An Energy Efficient Seizure Prediction Algorithm

Zhongnan Fang
Electrical Engineering
Stanford University
zhongnan@stanford.edu

Yuan Yuan
Statistics
Stanford University
yuan125@stanford.edu

Andrew Weitz
Bioengineering
Stanford University
aweitz@stanford.edu

***Abstract*—In this project, we sought to develop a learning algorithm that identifies EEG time series as pre-ictal (the time just before a seizure occurs) or inter-ictal (the time between seizures) using undersampled data to reduce the energy consumption and bandwidth usage. By training the model with segmented 1 second intervals and predicting seizures with a decision window, we achieved high seizure prediction accuracy at a downsampling rate as high as 40. These techniques can be broadly applied to reduce energy demands for a variety of wearable medical and health device applications.**

***Index Terms*—Seizure prediction, Logistic regression, SVM, Energy efficiency.**

I. Introduction

Epileptic seizures afflict over 1% of the world's population. If seizures could be predicted before they occur, fast-acting therapies could be delivered to prevent the attack and restore a normal quality of life to patients. Over the last two decades, several studies have explored the use of EEG signals to predict seizures using principles from machine learning [1]–[3]. It is thought that such an algorithm could be implemented in real-time with a wireless, implanted EEG sensor. However, there are two main constraints for such a portable system. First, due to limited battery life, energy consumption must be minimal. Second, due to limited bandwidth, the data transmitted between the sensor and the central processing device (such as mobile phone, tablet, personal computer, etc.) should be small. To address these issues, we sought to develop a robust learning algorithm that identifies EEG time series as pre-ictal (the time just before a seizure occurs) or inter-ictal (the time between seizures) using downsampled data. This could ultimately reduce both power consumption and bandwidth usage for wearable seizure prediction devices (Fig. 1).

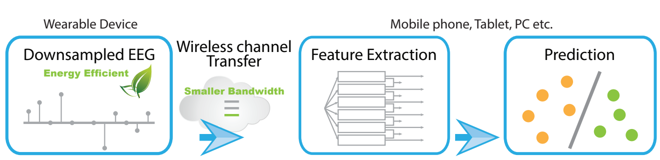

Fig. 1. Undersampling can be used to reduce energy consumption and data transferring bandwidth on personal medical and health devices. However, the effect of undersampling on the performance of machine learning algorithms is still an open question.

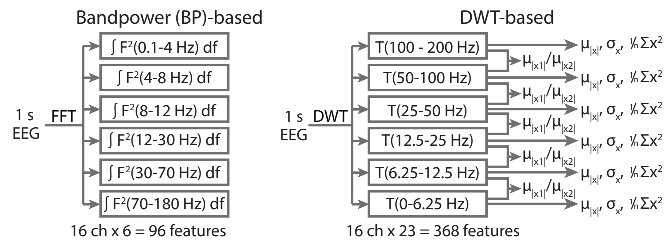

Fig. 2. We selected two feature sets for seizure prediction: bandpower-based (left) and DWT-based (right).

II. Methods

A. Data description

Data was provided by the American Epilepsy Society for its Seizure Prediction Challenge on www.kaggle.com. EEG recordings were collected in a single epileptic canine with 24 pre-ictal (y=1) and 24 inter-ictal (y=0) example files. Each file includes 16 channels of continuous EEG recordings for 10 min, sampled at 399.61 Hz. To increase the amount of data available for training for seizure prediction, we split each 10 min file into 1 s intervals to be individually classified. This resulted in a (400 sample) $\times$ (16 channel) matrix for each of the $m = (48 \text{ files}) \times (600 \text{ sec}) = 28800$ data points.

B. Features

1) Bandpower-based feature: For the first feature set we investigated, we quantified the EEG spectral power in six physiological frequency bands: delta (0.1-4 Hz), theta (4-8 Hz), alpha (8-12 Hz), beta (12-30 Hz), low-gamma (30-70 Hz), and high-gamma (70-180 Hz). This feature set has previously been shown to perform well for seizure prediction using logistic regression and SVM [2], [3]. Spectral power was extracted from each 1 s time series using the `bandpower()` function in Matlab. Because each recording consisted of 16 separate channels, the total number of features per 1 s interval was thus $16 \times 6 = 96$ (Fig. 2 left).

2) Wavelet based feature: For the second feature set we investigated, we used the discrete wavelet transform (DWT) with 5 levels of decomposition. The benefit of the DWT is that it can capture signal changes in both the temporal and frequency domain. This is important since the EEG signal is non-stationary and spectrum of the signal changes over time. To compute the DWT, the time series was decomposed into 6

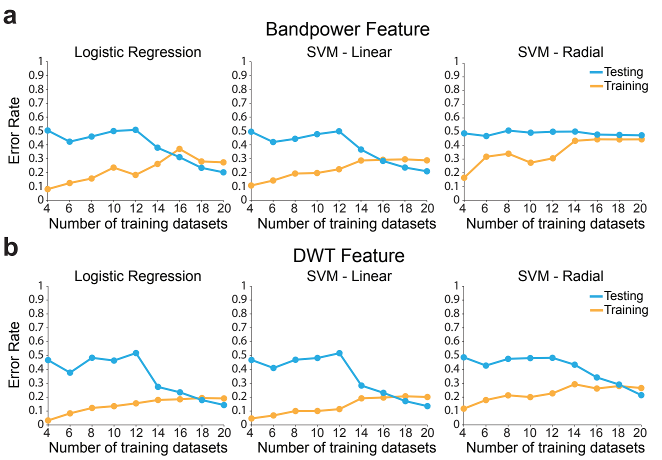

Fig. 3. Learning curves for each combination of learning algorithm and feature set were analyzed for potential over- or under-fitting problems. Each dataset contains six hundred of 1 second intervals.

hierarchical time series with different frequency bands: 100-200 Hz, 50-100 Hz, 25-50 Hz, 12.5-25 Hz, 6.25-12.5 Hz, and 0-6.25 Hz (Fig. 2 right). The following statistical features were calculated to represent the time-frequency distribution of the EEG signals, as originally proposed in [4]:

1) Mean of the absolute values of the wavelet coefficients in each sub-band $\mu_{|x|}$.
2) Average power (i.e. sum of squares) of the wavelet coefficients in each sub-band ${}^{1}/_{n}\Sigma x^2$.
3) Standard deviation of the coefficients in each sub-band σ_x.
4) Ratio of the absolute mean values of adjacent sub-bands $\mu_{|x_1|}/\mu_{|x_2|}$.

With 6 different sub-bands this gives a total of a $6 + 6 + 6 + 5 = 23$ features. Because each recording consists of 16 separate channels, the total number of features per 1 s interval was thus $16 \times 23 = 368$.

C. Training methods

1) Logistic Regression: We implemented logistic regression using the `glmfit` function in Matlab, with a logit link function.

2) Linear SVM: We implemented a linear support vector machine using the Liblinear package in Matlab. Feature vectors were scaled to the range -0.5 to 0.5, as suggested by the Liblinear authors.

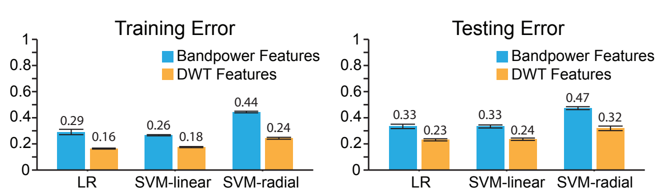

Fig. 4. 66%/33% hold-out cross-validation showed that logistic regression and the linear SVM using DWT-based features result in the lowest error. Error bars give standard error across 10 repeated trials.

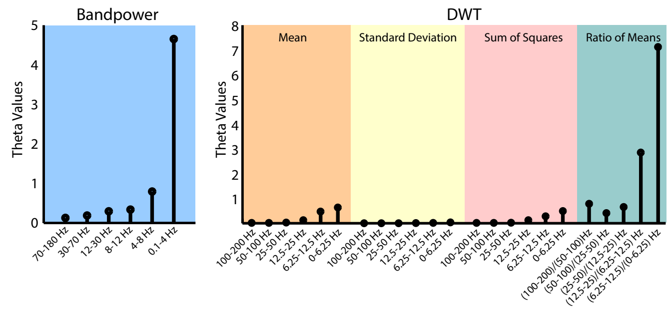

Fig. 5. Logistic θ values were compared for both bandpower-based and DWT-based features to identify those most important feature for seizure prediction. For bandpower, the most highly weighted features corresponded to low frequency bands. For DWT-based features, the the most highly weighted statistic was the ratio between the means of adjacent frequency bands. Lower frequencies were also more highly weighted for 3 of the 4 statistics (mean, sum of squares, and ratio of means). Note that θ values represent the average over 16 channels.

3) Radial (Gaussian) SVM: We implemented the SVM radial kernel using the LibSVM package in Matlab. Feature vectors were also scaled to the range -0.5 to 0.5, as with the linear kernel.

D. Learning curve and cross-validation

To confirm that our learning algorithm could accurately classify pre-ictal and inter-ictal states using the defined features without under- or over-fitting, we first generated learning curves as a function of dataset size (i.e. the number of 10 min EEG files). The testing set was kept at a constant size of 8 EEG files, with 4 positive examples and 4 negative examples. The training set was evaluated for sizes ranging from 8 to 40 EEG files, always with an equal number of positive and negative examples.

The generalization error of each learning algorithm was evaluated using 10 repeated trials of 66%/33% hold-out cross-validation. We split the original 48 EEG files into randomly assigned groups of 32 training and 16 testing datasets.

E. Seizure classification

As noted above, the original data were acquired as single files of 10 min EEG recordings (labelled as pre-ictal or inter-ictal), which we then divided into six hundred 1 s intervals. Although our learning algorithms treated the features from each of these 1 s intervals as a single data point, we ultimately sought a way to predict seizures at a larger temporal scale, such as every 30 s or 1 min. To do so, we implemented a decision window, in which the recording was classified as pre-ictal if the fraction of 1s intervals classified as pre-ictal exceeded a certain threshold. The resulting false-positive and false-negative rates with different thresholds were then used to generate ROC curves for performance evaluation. To determine how much data our algorithm needed to generate a correct prediction, we also varied the decision window size (i.e. how many 1s intervals were used to make a decision on pre-ictal or inter-ictal). In practice, a smaller window means

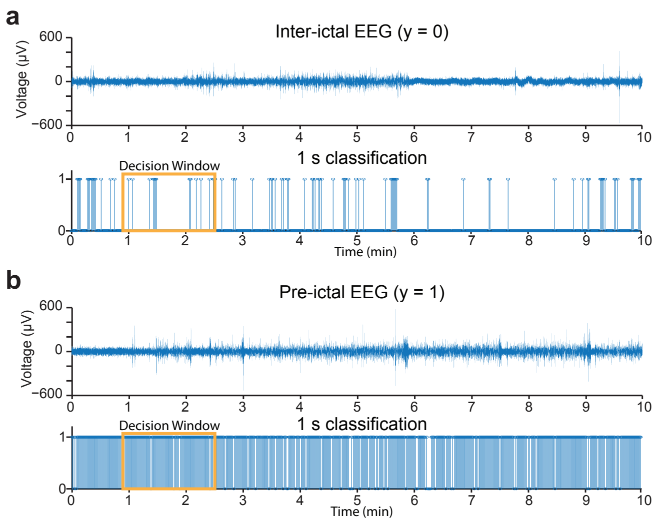

Fig. 6. Representative classification of 1 s intervals for inter-ictal (**a**) and pre-ictal (**b**) EEG recordings are shown. A decision window of varying size is then applied for seizure prediction and generation of ROC curves.

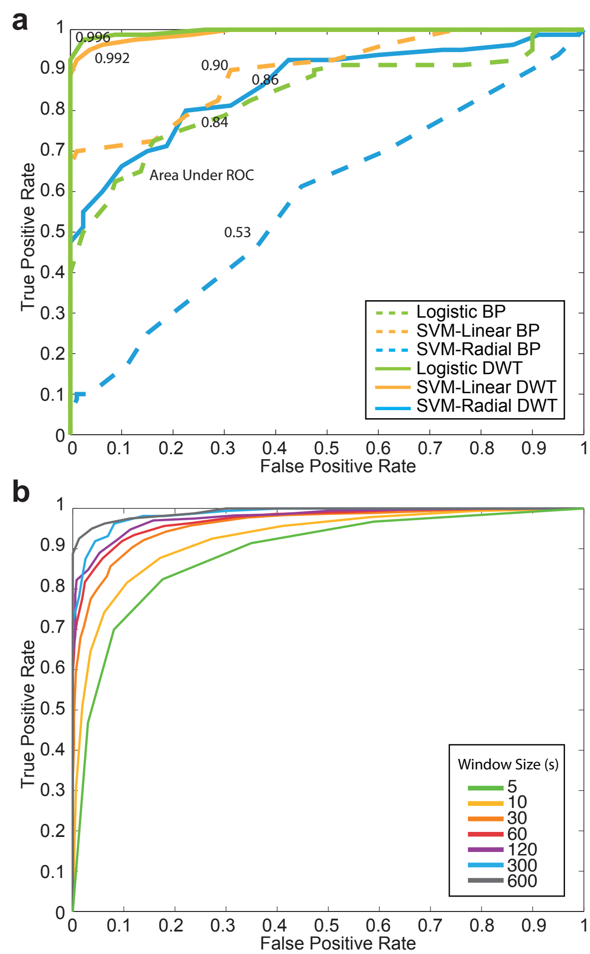

Fig. 7. (**a**) The false positive rate versus true positive rate of different algorithm and feature combinations was compared with a 10 min decision window. Similar to cross-validation, logistic regression and the linear SVM using DWT-based features outperformed all other models. (**b**) As the window size increases, the linear SVM using the DWT-based feature set gives better seizure prediction performance.

that a prediction can be made at a finer temporal scale (Fig. 6).

F. Downsampling pattern

Two different patterns were tested for reducing energy consumption and bandwidth usage: periodic and random downsampling. Periodic downsampling is equivalent to reduce the sampling rate of the original signal. Because of the recent development of compressed sensing theory, random downsampling has also become popular for signal recovery after downsampling [5]. Thus the random downsampling pattern was also tested in this project.

III. Results

A. Identification of the optimal seizure prediction algorithm

1) Validation of features and learning algorithms: We first sought to confirm that the chosen learning algorithms could classify EEG signals as pre-ictal or inter-ictal above a 50% chance level and diagnose if learning was subject to under- or over-fitting. Fig. 3 provides the learning curves for each learning algorithm using the bandpower- and DWT-based features. Logistic regression and the linear SVM gave comparable performance for either feature set, while the radial SVM failed to learn using the bandpower-based features and resulted in relatively higher testing error rate using DWT-based features. With the exception of the radial SVM using bandpower-based features, the general shapes of the learning curves were consistent with expectations. As the training set size increased, the training error went up, while the testing error went down. Thus,the logistic regression and linear SVM models successfully used the extracted features to predict pre-ictal versus inter-ictal states.

2) Selection of the optimal feature and learning algorithm: We next sought to identify the optimal feature and learning algorithm with 66%/33% hold-out cross-validation. As shown in Fig. 4, training with logistic regression or a linear SVM using the DWT-based feature gave the best performance, achieving less than 0.2 training error and less than 0.25 testing error. The radial SVM exhibited poor performance, resulting in almost 0.5 training and testing error using the bandpower-based feature. Thus, we conclude that the logistic regression and radial SVM algorithms using the DWT-based feature are optimal for seizure prediction.

3) Investigating the most important features: We next analyzed the θ values of the logistic regression model to determine which elements are the most important. As shown in Fig. 5a, we found the highest bandpower θ value is at the frequency band 0.1-4Hz, which indicates that the most important differences between the pre-ictal and inter-ictal states are stored in this band. Similarly, we found the lower

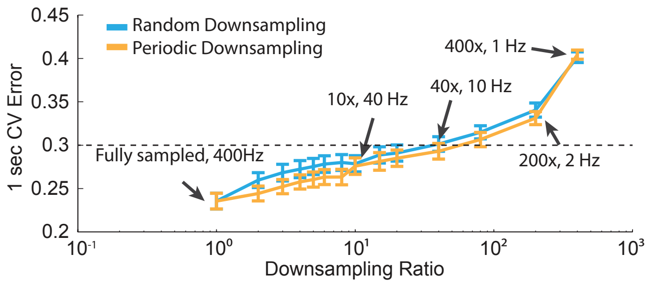

Fig. 8. 1 s interval cross-validation error of linear SVM using the DWT-based features as a function of downsampling rate. Two different downsampling patterns were investigated (periodic and random). For both downsampling patterns, the optimal algorithm achieved less than 0.3 error when the downsampling ratio was less than 40. When the downsampling ratio approaches 400 (i.e. only one sample is acquired per channel each second), the error rate increases to 0.42.

frequency coefficients are also weighted higher in the DWT-based features. Importantly, most significant features that were used for recognition are the ratios between absolute signal means (μ_{x_1}/μ_{x_2}) for the DWT-based features.

4) Seizure classification using decision window: The results reported above refer to the classification of individual 1 s intervals. Because we were ultimately interested in classifying the EEG signal as pre-ictal or inter-ictal at different time scales, we applied simple thresholding on intermediate 1 s decisions within a decision window (Fig. 6). By varying the seizure prediction threshold for a 10 min decision window, an ROC curve that quantifies the true positive and false positive rate at different threshold is obtained (Fig. 7a). It can be concluded that the DWT-feature with logistic regression or a linear SVM gave superior performance (more than 99% area under ROC curve). The bandpower-based feature had worse performance than DWT for each learning algorithm. The radial SVM still performed worse than both logistic regression and the linear SVM for either feature. The size of the decision window was also investigated using one of the best learning algorithms, i.e. DWT-feature with a linear SVM. As shown in Fig. 7b, we found that the seizure prediction performance increases with window size. Surprisingly, reasonable performance (90% area under ROC curve) could still be obtained with only a 5 s window.

B. Robustness of the optimal seizure prediction algorithm under downsampling

1) Cross validation error vs downsampling ratio: We investigated the robustness to downsampling of one of the optimal algorithms (linear SVM with DWT-based features). As shown in Fig. 8, the cross validation error using the 1 s intervals slowly increased from 0.24 to 0.3 as the downsampling ratio increased to 40. When the downsampling ratio reached 400x, the algorithm failed to correctly classify the 1 s intervals and the error dramatically increased to about 0.45. Another finding from this test is that the periodic downsampling performs slightly better than the random downsampling, which could

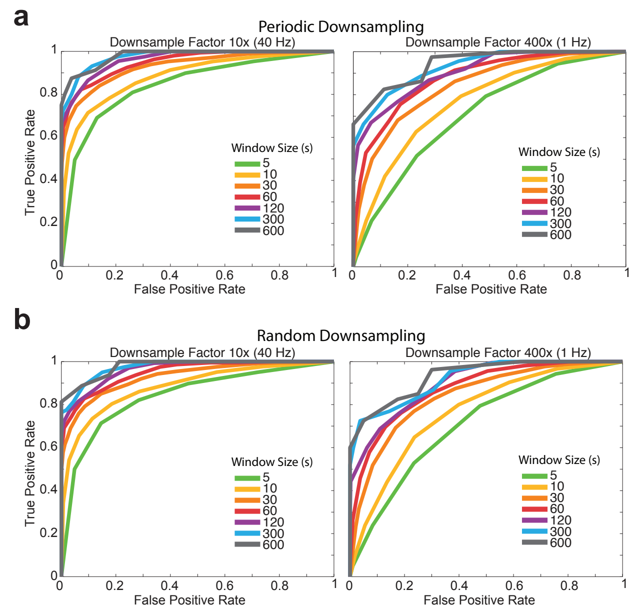

Fig. 9. Although downsampling increased the error rate for 1 sec classification, the seizure prediction performance could be improved by using a larger decision window. For example, when the 1 s interval was downsampled by 10 times, our system could still achieve greater than 0.8 true positive rate and less than 0.2 false positive rate using a decision window of at least 30s.

because the aliasing pattern is more coherent using the periodic undersampling and thus the learning algorithm could pick up the aliased features more easily.

2) Improving performance with larger decision window sizes: Although we showed that the 1 s cross validation error increased as the downsampling ratio increased, the seizure prediction performance could be preserved by increasing the decision window size. As shown in Fig. 9, we investigated the ROC curve with 10x and 400x downsampling using seven different window sizes. The ROC curves are approximately equal between the periodic and random downsampling methods. With a window size larger than 30s, we found that the optimal algorithm could still achieve a higher than 0.8 true positive rate and less than 0.2 false positive rate when the downsampling ratio was as high as 10. At 400 times downsampling, we found that decision windows larger than 300s could still achieve similarly high prediction performance (i.e. larger than 0.8 true positive and less than 0.2 false positive rate).

IV. Conclusions, Discussions

In this project, we developed an energy efficient algorithm for seizure prediction. We investigated two different types of features (powerband-based and DWT-based), and three different learning algorithms (logistic regression, linear SVM and radial SVM). After cross validation, we found that logistic regression and linear SVM using the DWT-based features outperforms all other algorithms. We also found that low frequency signals contain the most information on the differences between pre-ictal and inter-ictal brain states. To improve the robustness of our classification, we designed a decision window method for seizure prediction. With window sizes as large as 600s, the algorithm could achieve greater than 0.95 true positive rate and less than 0.05 false positive rate. Surprisingly, performance was still far better than chance with decision windows as small as 5 s. We next investigated the robustness of one of the optimal algorithms (linear SVM with DWT-based features) to downsampling. We found the cross validation error increased with larger downsampling ratios, although the error rate was still less than 0.3 for downsampling ratios less than 40. We also showed that large decision windows could be used to improve the prediction accuracy using downsampled data. For example, with a window size larger than 30s, we achieved a greater than 0.8 true positive rate and less than 0.2 false positive rate using data sampled at only 10 Hz.

One interesting finding from this project is that the learning algorithm still performs well when the EEG signal is downsampled with aliasing artifacts. This could be because most of the important features were stored in the low frequency domain. Another possible reason is that even when the signal is aliased, the learning algorithm could still use features at specific aliased locations for classification.

With the proposed robust seizure prediction algorithm, energy consumption and bandwidth usage on wearable EEG devices can be largely reduced. This method can also be implemented in other personal wearable devices, such as a heart rate monitor, for higher energy efficiency and longer battery life.

V. Future Work

Since the analysis was conducted on a single animal, more experiments should be conducted to justify the method. The energy savings gained from downsampling should also be investigated so that an optimal decision window size can be identified for embedded applications.

Acknowledgment

The authors would like to thank Professor Andrew Ng for his inspirational lectures and sharing his enthusiasm for the machine learning. We would also like to thank all TAs for their thoughtful suggestions for this project.

References

[1] F. Mormann, R. G. Andrzejak, C. E. Elger, and K. Lehnertz, "Seizure prediction: the long and winding road," *Brain*, vol. 130, no. 2, pp. 314–333, 2007.

[2] Y. Park, L. Luo, K. K. Parhi, and T. Netoff, "Seizure prediction with spectral power of eeg using cost-sensitive support vector machines," *Epilepsia*, vol. 52, no. 10, pp. 1761–1770, 2011.

[3] J. J. Howbert, E. E. Patterson, S. M. Stead, B. Brinkmann, V. Vasoli, D. Crepeau, C. H. Vite, B. Sturges, V. Ruedebusch, J. Mavoori *et al.*, "Forecasting seizures in dogs with naturally occurring epilepsy," *PloS one*, vol. 9, no. 1, p. e81920, 2014.

[4] A. Subasi, "Eeg signal classification using wavelet feature extraction and a mixture of expert model," *Expert Systems with Applications*, vol. 32, no. 4, pp. 1084–1093, 2007.

[5] Z. Zhang, T.-P. Jung, S. Makeig, and B. D. Rao, "Compressed sensing of eeg for wireless telemonitoring with low energy consumption and inexpensive hardware," *Biomedical Engineering, IEEE Transactions on*, vol. 60, no. 1, pp. 221–224, 2013.