
A Gradient-Based Adaptive Learning Framework for Online Seizure Prediction

Shouyi Wang*

Department of Industrial and Manufacturing Systems Engineering,
University of Texas at Arlington,
Arlington, TX 76019, USA
E-mail: shouyiw@uta.edu

*Corresponding author

Wanpracha Art Chaovallitwongse

Department of Industrial and Systems Engineering,
Department of Radiology at Medical Center,
University of Washington,
Seattle, WA 98195, USA
E-mail: artchao@u.washington.edu

Stephen Wong

Robert Wood Johnson Medical School,
University of Medicine and Dentistry of New Jersey,
New Brunswick, NJ 08901, USA
E-mail: wongst@umdnj.edu

Abstract: Most of the current epileptic seizure prediction algorithms require much prior knowledge of a patient's pre-seizure electroencephalogram (EEG) patterns. They are impractical to be applied to a wide range of patients due to a very high inter-individual variability of EEG patterns. This paper proposes an adaptive prediction framework, which is capable of accumulating knowledge of pre-seizure EEG patterns by monitoring long-term EEG recordings. The experimental results on five patients indicate that the proposed prediction approach is effective to achieve a personalized seizure prediction for each patient using a gradient-based adaptive learning framework.

Keywords: adaptive seizure prediction; gradient-based learning; time series online monitoring.

Reference to this paper should be made as follows: Wang, S., Chaovallitwongse W., and Wong, S. (2011) 'A Gradient-Based Adaptive Learning Framework for Online Seizure Prediction'.

Biographical notes: Shouyi Wang received a B.S. degree in Control Science and Engineering from Harbin Institute of Technology, Harbin, China, in 2003, and a M.S. degree in Systems and Control

Engineering from Delft University of Technology, Netherlands in 2005. Currently, he is a Ph.D. candidate at the Department of Industrial and Systems Engineering, Rutgers University. And he is also working as a research scientist at the Department of Industrial & Systems Engineering and the Integrated Brain Imaging Center (IBIC), University of Washington. His research interests include time series data mining, pattern discovery, machine learning, intelligent decision-making systems, multivariate time series modeling and forecasting.

Wanpracha Art Chaovalitwongse received a B.S. degree in Telecommunication Engineering from King Mongkut Institute of Technology Ladkrabang, Thailand, in 1999 and M.S. and Ph.D. degrees in Industrial and Systems Engineering from University of Florida in 2000 and 2003, respectively. In 2004, he worked at the Corporate Strategic Research, ExxonMobil Research & Engineering, where he managed research in developing efficient mathematical models and novel statistical data analysis for upstream and downstream business operations. He was an Assistant Professor from 2005 to 2010 and an Associate Professor from 2010 to 2011 in the Department of Industrial & Systems Engineering, Rutgers University. Currently, he is an Associate Professor in the department of Industrial & Systems Engineering and the Department of Radiology at Medical Center, University of Washington.

Stephen Wong received a B.S. degree in 1996 at the California Institute of Technology in Pasadena, CA. He obtained his M.D. at Duke University in the year 2000 in Durham, NC, and received subsequent medical training in neurology and epilepsy at the University of Pennsylvania Health System in Philadelphia, PA. He was the recipient of scholarship for a Master's Degree in Translational Research, which he received in 2010, during which he studied signal processing and machine learning methods applied to clinical neurophysiology and electroencephalography. He is currently an assistant professor of neurology at UMDNJ-Robert Wood Johnson Medical School in New Brunswick, NJ, where he provides clinical care to patients with epilepsy, teaches medical students and residents, and performs computational research related to event detection in neurophysiology.

1 Introduction

According to Engel and Pedley (1997), about 50 million people suffer from epilepsy, which is a chronic neurological disorder characterized by recurrent unprovoked seizures. Epileptic seizures generally occur without any warning, the shift between normal brain state and seizure onset is often described as an abrupt phenomenon. The seemingly unpredictability of epileptic seizures is one of the major causes of morbidity and stress in patients with epilepsy. Therefore, being able to identify pre-seizure symptoms could significantly improve the quality of life for these patients and can also open new diagnostic and therapeutic opportunities in epilepsy treatment.

Over the recent years, there has been accumulating evidence indicating that a transitional pre-seizure state does exist prior to seizure onset. As stated in Delamont et al. (1999), pre-seizure symptoms such as irritability or headache

are frequently exhibited minutes, hours or even days prior to seizure onsets. Some other clinical findings also support the existence of a pre-seizure state, such as increases in cerebral blood flow, oxygen availability, and blood oxygen level-dependent signal, and changes in heart rate prior to seizure occurrence. The quantitative studies of the pre-seizure state are mostly based on EEG recordings from patients with epilepsy. For example, Lehnertz and Elger (1998) showed that the correlation dimension decreases prior to seizures. Quyen et al. (2003) reported a reduction in the dynamical similarity index before seizure occurrence. Iasemidis et al. (1997) noted premonitory pre-seizure changes based on the analysis of dynamical entrainment. Mormann et al. (2006) observed a pre-seizure drop in phase synchronization up to hours prior to seizure onset.

In the past decades, many studies have been carried out aiming to predict epileptic seizures. Seizure predictability by EEG recordings has been confirmed by a number of groups (Elger and Lehnertz (1998); Lehnertz and Elger (1998); Quyen et al. (1999); Litt et al. (2001); Jerger et al. (2005)). An extensive survey of EEG-based seizure prediction techniques can be found in Mormann et al. (2007). Most of the current seizure prediction methods mainly have two steps. Firstly, EEG features are extracted from a sliding moving window. Then each window-EEG is classified as either pre-seizure or normal by comparing the extracted EEG features with the predefined threshold levels. Whenever a window-EEG is classified as pre-seizure, a warning alarm is triggered indicating that an impending seizure may occur within a pre-defined prediction horizon. These methods have shown good results for some patients. However, the reliability and repeatability of the results have been questioned when they were tested on other EEG datasets. Many of the earlier optimistic findings cannot be reproduced or achieved poor performance in extended EEG datasets in later studies as reported in Aschenbrenner-Scheibe et al. (2003). This is not surprising since the optimal threshold obtained from a few number of patients may not be appropriate to many others. Therefore, we regard the future perspective of a practical seizure prediction system with ‘optimized’ thresholds as unrealistic. Instead, we would conjecture the most promising approach should be an intelligent method that can be autonomously adaptive to each individual patient with learning ability.

Inspired by the great reinforcement learning ability of human beings, we attempt to construct an adaptive learning system, which could interactively learn from a patient, and is capable of improving prediction performance over time. In addition, many of the current seizure prediction studies are basically to solve a classification problem for normal and pre-seizure EEG data. The resulting methods do not take sufficient account of the online monitoring property in their prediction methods. The proposed adaptive framework successfully combines the reinforcement learning concept, online monitoring and adaptive control theory to achieve the online patient-specific seizure prediction. This work is among the pioneering attempts to tackle the greatly challenging task of online seizure prediction. The prediction framework avoids a sophisticated threshold-tuning process, and largely enhances the adaptability of the current prediction techniques. The autonomous self-adaptation property of the system makes it convenient to use for end users, such as physicians and patients. The outcome of this study would shed some light on the perspective reliable online seizure prediction techniques. It

may also improve the medical diagnosis and prognosis in other brain diseases, such as sleep disorders and cognitive disorders.

This paper is organized as follows. Section 2 presents the experimental methods, including EEG collection, feature extraction, and the proposed adaptive learning framework. The experimental results are provided in Section 3, and we conclude the paper in Section 4.

2 Methods

All the experiments were performed on an Intel Xeon 2.0 GHz 64-bits workstation with 16 gigabytes of memory running on Windows Server 2003. All calculations and algorithms were implemented and run on MATLAB R2009b. Both default Matlab programs and user-designed programs were used in the experiments.

2.1 Data Collection

Intracranial EEG recordings from five epileptic patients with temporal lobe epilepsy were analyzed in this study. The placement of the EEG electrodes is shown in Figure 1, which is a modified image of the inferior transverse view of the brain from Potter (2006). The recorded EEG data is summarized in Table 1. The EEG recordings consist of 26 standard channels, and the durations are ranged from 3 to 13 days. A total of 89 seizures over 43 days were recorded. The starting and the ending time points of seizure onsets were determined by experts.

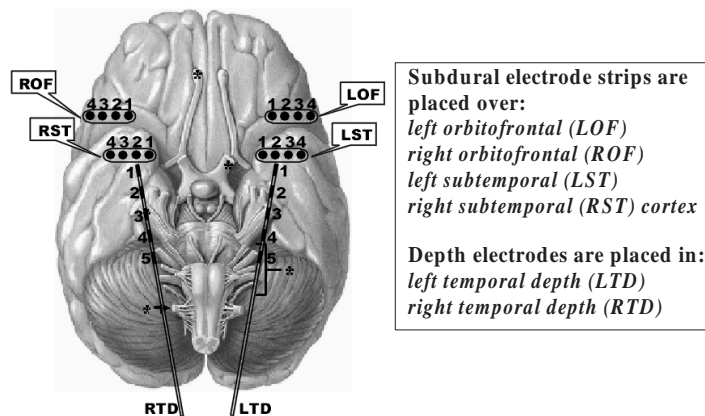


Figure 1 The interior transverse view of the brain and the placement of the 26 EEG electrodes.

2.2 Data Preprocessing & Feature Extraction

Many feature extraction techniques have been developed to analyze EEG signals, such as time-domain analysis, frequency-domain analysis, time-frequency analysis, and spatial-temporal analysis. A comprehensive review of various EEG processing

Table 1 Summary of the Analyzed EEG Data

Patient	Number of Electrodes	Duration of EEG (days)	Number of Seizures	Seizure Rate (per hour)
1	26	3.55	7	0.082
2	26	8.85	22	0.104
3	26	13.13	17	0.054
4	26	6.09	23	0.157
5	26	11.53	20	0.061
Total		43.15	89	

methods can be found in Stam (2005). Since EEG signals are highly nonstationary and seemingly chaotic, there has been an increasing interest in analyzing EEG signals in the context of chaos theory according to Rapp et al. (1989). Chaos theory provides effective quantitative methods to measure EEG dynamics and discover the underlying chaos in the data.

Several chaotic measures are commonly used in recent EEG literature, such as correlation dimension in Silva et al. (1999), largest Lyapunov exponent in Iasemidis et al. (2003), Hurst exponent in Dangel et al. (1999) and entropy in Quiroga et al. (2000). Among such measures, Lyapunov exponent, the average rate of divergence of two neighboring trajectories in phase space, is often considered as the most basic indicator of deterministic chaos (Vastano and Kostelich (1986)). Lyapunov exponents supply a direct measure of the degree of sensitivity to initial conditions for a dynamical system. For a n -dimensional dynamical system, there are n different Lyapunov exponents, λ_i . They measure the exponential rate of divergence of the different trajectories in the phase space. If one of the exponents is positive, it indicates that the two corresponding orbits defined by that exponent diverge exponentially. The magnitude of the exponents indicates the degree of divergence. It has been shown that the chaotic behavior of a dynamical system is usually sufficient to be characterized by the largest Lyapunov exponent instead of all the exponents. The largest Lyapunov exponent has been shown to be reliable and reproducible in Vastano and Kostelich (1986).

In our previous studies, an estimation algorithm called short-term largest Lyapunov exponent (STL_{max}) was developed to quantify EEG dynamics in Iasemidis et al. (2000). Along this line of research, we also employ STL_{max} to characterize raw EEG data in this study. The calculation of STL_{max} can be briefly described as follows. The initial step is to embed each channel of EEG signal in a p -dimensional phase space, and construct p -dimensional vectors $X(t_i) = (x(t_i), x(t_i + \tau), \dots, x(t_i + (p-1)\tau))$, where t_i is the time point, τ is the selected time lag between the components of each vector in the phase space, and p is the selected dimension of the embedding phase space. Define N the number of local STL_{max} s that will be estimated within a duration T data segment, then the

largest Lyapunov exponent is defined as the average of local Lyapunov exponents L_{ij} in the phase space as follows:

$$STL_{max} = \frac{1}{N} \cdot \sum_N L_{ij}, \quad (1)$$

and the local Lyapunov exponents L_{ij} is defined by:

$$L_{ij} = \frac{1}{\Delta t} \cdot \log_2 \frac{|X(t_i + \Delta t) - X(t_j + \Delta t)|}{|X(t_i) - X(t_j)|}, \quad (2)$$

where Δt is the evolution time for the vector difference $\delta X_{i,j}(0) = |X(t_i) - X(t_j)|$ to evolve to the new difference $\delta X_{i,j}(\Delta t) = |X(t_i + \Delta t) - X(t_j + \Delta t)|$. More details of the estimation of STL_{max} can be found in Iasemidis (1991).

2.3 Adaptive Seizure Prediction Framework

The proposed adaptive seizure prediction framework is illustrated in Figure 2. The continuous multichannel EEG data were analyzed by a sliding moving window. The window had a size of 10 min and moved with a 50% overlap each step. Two baselines of normal and pre-seizure states were constructed to classify the window-EEGs using a KNN method. All the baseline samples and window-EEGs were represented in terms of multichannel time profile of STL_{max} s. Based on prediction feedbacks (correct or incorrect), the two baselines were updated according to a reinforcement learning procedure. The adaptive seizure prediction system is discussed in detail in the following.

2.3.1 Baseline Construction & Initialization

To start our prediction system, we need to initialize the pre-seizure and normal baseline samples. The selection of baseline samples highly depends on the presumed time length of pre-seizure period, which is often used as prediction horizon in seizure prediction literature. So far little is known to define pre-seizure duration, which has been reported between a few minutes and several hours prior to seizure onsets. The prediction horizon for epileptic seizures is still an open question in epilepsy research. In this study, we tried three prediction horizons, which are 30 min, 90min, and 150min, respectively. If we denote the prediction horizon at H minutes, then the EEG recordings can be divided into the following three periods:

- Pre-seizure period: 0- H min preceding a seizure onset.
- Post-seizure period: 0-20 min after a seizure onset.
- Normal period: between pre- and post-seizure periods.

The post-seizure period was set at 20 minutes based on the observations of the EEG recordings. The EEG signals generally recovered to normal patterns 20 minutes after the seizure onsets in all the five patients. In addition, the initial

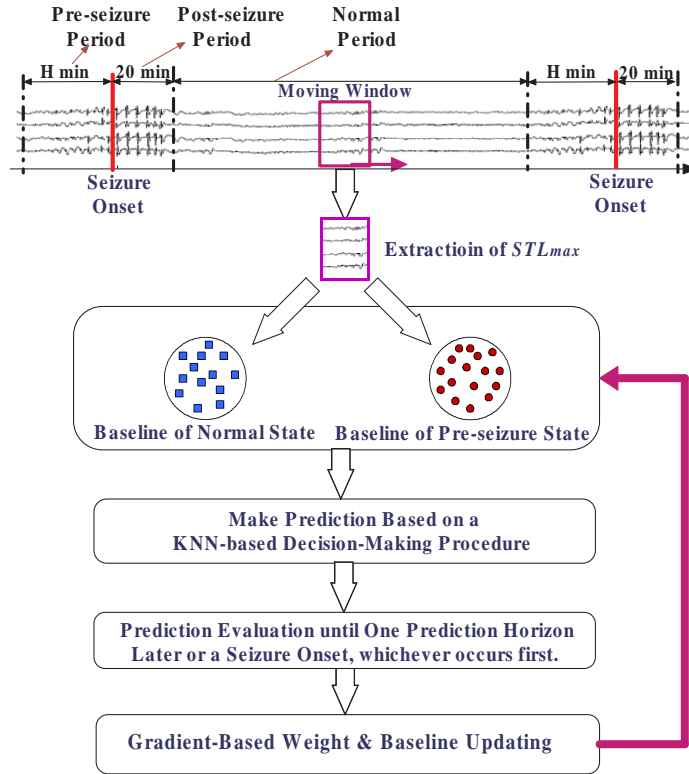


Figure 2 Schematic structure of the adaptive prediction system.

samples of the two baselines were randomly chosen from the normal and pre-seizure period preceding the first seizure onset. The length of the baseline samples is equal to that of the moving window. Since there is no guideline available to determine the number of samples in each baseline, we tentatively stored a fixed number of 50 samples in each baseline.

2.3.2 KNN Similarity Measure

With baselines for normal and pre-seizure states, it is intuitive and practical for physicians to decide the class of a window-EEG based on its degree of similarity between the two baselines. For this purpose, KNN is the best choice because it classifies a new unlabeled sample by comparing it with all the samples of the two baselines. Thus, we employed KNN method to find the K best matching samples in each of the two baselines and compare them to make a decision.

A KNN method has to use similarity measures to quantify the closeness between a moving-window EEG and baseline samples. We employed three frequently used similarity measures for time series data. If we denote two time series of STL_{max} as X and Y with equal length of n , then the three types of distances are briefly described as follows.

- Euclidean distance (EU): $ED_{xy} = \sum_{p=1}^n (x_p - y_p)^2 / n$.

- T-statistical distance (TS): $TS_{xy} = \sum_{p=1}^n |x_p - y_p| / \sqrt{n} \tau_{|X-Y|}$, where $\tau_{|X-Y|}$ is the sample standard deviation of the absolute difference between the time series X and Y .
- Dynamic time warping (DTW): DTW measures similarity based on the best possible alignment or the minimum mapping distance between two time series. A detailed calculation of DTW can be found in Senin (2008).

Once a similarity measure is chosen, the distance between a window-EEG and a baseline sample, denoted as window-sample distance, can be obtained as follows:

$$d_{pre,i} = \sum_{j=1}^M distance(S_{pre,i}^j, S_{mw}^j), \quad (3)$$

$$d_{int,i} = \sum_{j=1}^M distance(S_{int,i}^j, S_{mw}^j), \quad (4)$$

where $M=26$ is the number of EEG channels. $S_{pre,i}^j$ and $S_{int,i}^j$ is the j th channel of the i th pre-seizure and normal baseline sample, respectively; $S_{mw,i}^j$ is the j th channel of the window-EEG epoch. $d_{pre,i}$ and $d_{int,i}$ denote the distance between the window-EEG and the i th sample in the pre-seizure and normal baseline, respectively. The term *distance* in the above formula represents a time series distance measure, which can be EU, TS, or DTW in this paper.

2.3.3 KNN Prediction Procedure

Four choices of K were employed, which were three, seven, half, and all of the baseline samples, respectively. For a specific value of K , the weighted summation of K nearest window-sample distances in a baseline was considered as the distance between the window-EEG and that baseline. We call the two distances as window-normal distance and window-preseizure distance, respectively. For each window-EEG, its distances to the two baselines can be calculated as follows:

$$D_{pre}^K = \sum_{k=1}^K \alpha_k d_{pre,k}, \quad (5)$$

$$D_{int}^K = \sum_{k=1}^K \beta_k d_{int,k}, \quad (6)$$

where D_{pre}^K and D_{int}^K are the window-preseizure distance and window-normal distance, respectively. $d_{pre,k}$ and $d_{int,k}$ are the window-sample distances of the k th sample of the K nearest neighbors in the pre-seizure and normal baseline, respectively. α_k and β_k are the weights of the k th sample in the pre-seizure and normal baseline, respectively; In the beginning, the initial weights of all baseline samples were equal, which are given by:

$$\alpha_i = \beta_i = \frac{1}{n}, \quad i = 1, \dots, n, \quad (7)$$

$n=50$ is the number of samples in each baseline. We assume that different baseline samples may have different power in decision-making. The ‘importance’ of a baseline sample can be represented by a weight associated with it. The weights of the baseline samples are updated through a gradient-based learning rule which will be discussed in a later part of this paper.

Once the two baseline-window distances are obtained, the prediction decision can be made by:

$$predictor = \begin{cases} 1, & \text{if } D_{pre}^K / D_{int}^K \leq h \text{ (trigger a warning),} \\ 0, & \text{otherwise (no warning),} \end{cases}$$

where the threshold $h = 1$ by default. The schematic structure of the KNN prediction rule is illustrated in Figure 3.

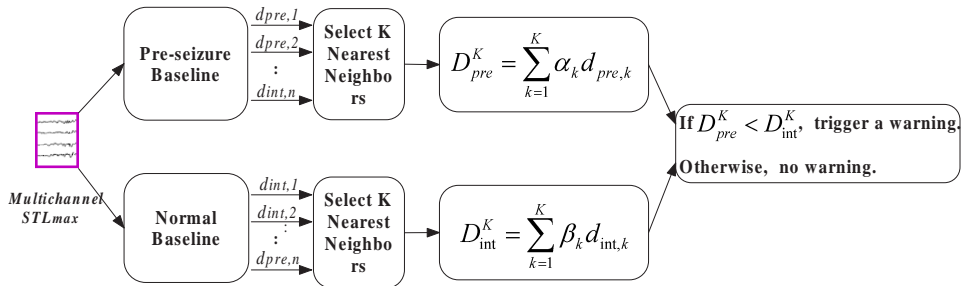


Figure 3 A demonstration of the KNN-based prediction rule.

2.3.4 Evaluation of a Prediction Result

If the prediction horizon is H min, then the feedback of each prediction outcome can be classified into one of the following four categories:

- True positive (TP): if $predictor = 1$ and a seizure occurs within H minutes after the prediction.
- False positive (FP): if $predictor = 1$ and no seizure occurs within H minutes after the prediction.
- True negative (TN): if $predictor = 0$ and no seizure occurs within H minutes after the prediction.
- False negative (FN): if $predictor = 0$ and a seizure occurs within H minutes after the prediction.

The concept of the prediction evaluation is also illustrated in Figure 4 for a better explanation. Based on this definition, we can evaluate a prediction result by identifying its corresponding category. Accordingly, the baselines will be updated based on the categorical prediction evaluations. In the following, the adaptive learning mechanism will be presented.

		Prediction Outcome	
		<i>pre-seizure</i>	<i>normal</i>
Actual	<i>pre-seizure</i>	TP	FN
	<i>normal</i>	FP	TN

Figure 4 The categorization of a prediction outcome. Each prediction outcome can be categorized into one of the four subsets.

2.3.5 Gradient-Based Weight Update Rule

The flowchart of the baseline update framework from delayed prediction feedback is illustrated in Figure 5. Let $r \in (0, 1)$ denote the learning rate to control the update size for the weights, then the gradient-based weight update rules are represented as follows:

- For cases of TP & FN, the weight update rule is:

$$\alpha_i = \alpha_i \left(1 - \frac{d_{pre,i} - \bar{d}_{pre}}{\bar{d}_{pre}}\right) \times r, \quad (8)$$

$$\beta_i = \beta_i \left(1 + \frac{d_{int,i} - \bar{d}_{int}}{\bar{d}_{int}}\right) \times r. \quad (9)$$

- For cases of FP & TN, the weight update rule is:

$$\alpha_i = \alpha_i \left(1 + \frac{d_{pre,i} - \bar{d}_{pre}}{\bar{d}_{pre}}\right) \times r, \quad (10)$$

$$\beta_i = \beta_i \left(1 - \frac{d_{int,i} - \bar{d}_{int}}{\bar{d}_{int}}\right) \times r, \quad (11)$$

where $\forall i = 1, 2, \dots, n$, $\bar{d}_{pre} = \sum_{i=1}^n d_{pre,i}/n$, and $\bar{d}_{int} = \sum_{i=1}^n d_{int,i}/n$.

Since the above update rule cannot guarantee that the summation of the sample weights in one baseline equals to 1, we normalized the obtained new weight vectors after each update as follows:

$$\alpha = \alpha / \sum_{i=1}^n \alpha_i, \quad (12)$$

$$\beta = \beta / \sum_{i=1}^n \beta_i. \quad (13)$$

2.4 Baseline Sample Update Rule

After each weight update, the weights of ‘good’ samples are increased while the weights of ‘bad’ samples are decreased by the gradient-based weight update method. The weights indicate the ‘importance’ of the corresponding baseline

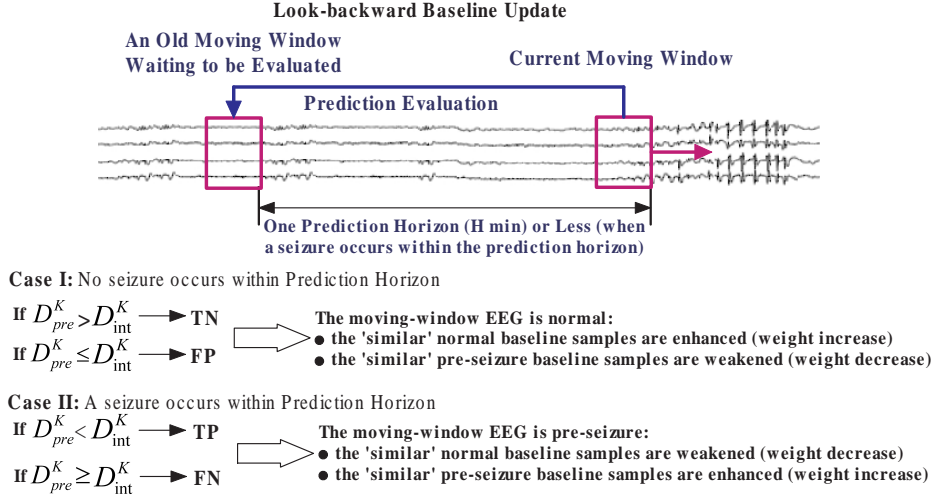


Figure 5 An Illustration of the gradient-based updating framework.

samples. If the weight of a sample is large, it means that the sample is important in prediction. While if the weight of a sample is small, it may disclose that the sample mainly plays a negative role in the past predictions and become less important. It is a natural method to update the baselines by replacing the low-weighted samples with moving-window EEGs. In this study, the update rule of baseline samples is based on the categorical prediction evaluations as follows:

- TP (correctly predict a pre-seizure window-EEG as pre-seizure): If the minimum weight in the pre-seizure baseline falls below a critical minimum value w_{min} , defined as $1/10$ of the initial weight value ($1/n$) in this study, then replace the corresponding pre-seizure baseline sample with the window-EEG.
- TN (correctly predict a normal window-EEG as normal): If the minimum weight in the normal baseline falls below the critical minimum value w_{min} , then replace the corresponding normal baseline sample with the window-EEG.
- FP (incorrectly predict a normal window-EEG as pre-seizure): Use the window-EEG to replace the normal baseline sample which has the largest sample-window distance.
- FN (incorrectly predict a pre-seizure window-EEG as normal): Use the window-EEG to replace the pre-seizure baseline sample which has the largest sample-window distance.

Once a replacement is made, an initial weight has to be assigned to the new baseline sample. For example, the baseline sample i is replaced, then the initial weight α_i of the new baseline sample is given by:

$$\alpha_i \leftarrow \alpha_i + \frac{(1 - \alpha_i)n}{n - 1}. \quad (14)$$

To make the summation of all the weights equal to 1, a normalization step is also performed subsequently by

$$\alpha \leftarrow \frac{\alpha}{1 + \frac{(1-\alpha)n}{n-1}}. \quad (15)$$

After normalization, the weight of the new baseline sample is scaled to $1/n$, while the relative relationship of the weights for all other baseline samples keeps unchanged.

2.5 Evaluation of Prediction Performance

Time Block-Based Sensitivity and Specificity: We label the continuous EEG by a series of time blocks. The length of each block is equal to the length of the prediction horizon (H min). In particular, the pre-seizure periods were defined as pre-seizure time blocks. The normal period between two seizures were divided into a number of equal-sized normal time blocks with a length of H min. The prediction outcome of each time block can be categorized into one of the following four subsets:

- TP-block: a pre-seizure time block with at least one warning within the block.
- TN-block: a normal time block with no warning.
- FP-block: a normal time block with at least one warning within the block.
- FN-block: a pre-seizure time block with no warning.

Then the time block-based sensitivity and specificity are defined as follows:

$$sen_{blk} = \frac{\text{number of TP-blocks}}{\text{number of pre-seizure time blocks}}, \quad (16)$$

$$spe_{blk} = \frac{\text{number of TN-blocks}}{\text{number of normal time blocks}}. \quad (17)$$

The time block-based sensitivity and specificity are more suitable to evaluate prediction performance than the traditional definition of sensitivity and specificity, since they consider the effects of prediction horizon for online seizure prediction.

Receiver Operating Characteristic (ROC) Analysis: Based on sensitivity and specificity, a common method in comparing the performance of a predictive model is to use the ROC curve. The ROC curve is a plot of sensitivity versus false alarm rate (1-specificity) as the discrimination setting of a classifier is varied. The ROC curve for a perfect prediction model is the line connecting $[0, 0]$ to $[0, 1]$ and $[0, 1]$ to $[1, 1]$. And the diagonal line connecting $[0, 0]$ to $[1, 1]$ is the ROC curve corresponding to a random model. Generally, a ROC curve lies between these two extreme lines. The *area under the ROC curve* (AUC) is often used as an important metric to evaluate a prediction model. The AUC is an overall summary of prediction accuracy across the spectrum of its decision-making values. AUC values are usually between 0.5 and 1. The AUC of a perfect predictor is 1 while a purely random chance model has an AUC of 0.5 on average. The higher the AUC value is to one, the better prediction power it indicates.

3 Results

3.1 Prediction Performance based on sen_{blk} and spe_{blk}

We tested the proposed adaptive prediction framework on the EEG recordings from five patients with epilepsy. Table 2 summarizes the sen_{blk} and spe_{blk} of the adaptive prediction system with the best parameter settings. The results showed that the sen_{blk} was ranged from 57.89% (Patient 5, 30 min) to 100.00% (Patient 1 at 90 & 150 min, Patient 3 at 90 min), and the spe_{blk} was ranged from 37.04% (Patient 3 at 90 min) to 78.22% (Patient 2 at 30 min). The best averaged accuracy was achieved at 70.60% using the prediction horizon of 150 min. The averaged prediction accuracies over the three prediction horizons were 69.99%, 67.50%, and 70.60%, respectively. In contrast with the adaptive framework, a Poisson random predictor was also tested. The Poisson predictor randomly triggered a warning alarm with a mean interval of λ minutes. We tested different values of λ and obtained similar prediction performances. As an example, the performance of the Poisson predictor with $\lambda=60$ min is shown in Table 3.

After comparing the prediction results, one can see that the proposed adaptive scheme achieved better prediction performance than the Poisson random predictor. The adaptive scheme achieved an overall accuracy of over 65%, while the random predictor only had an overall prediction accuracy around 50%. Moreover, we also notice that the adaptive prediction scheme generated consistent good results in terms of sensitivity and specificity for different prediction horizons. In each case, both of the sensitivity and specificity are larger than 50%. However, the random predictor generated very poor and unbalanced sensitivity and specificity, especially when using long prediction horizons.

Table 2 The summary of sen_{blk} and spe_{blk} obtained by the adaptive learning predictor with the best parameter settings. The KNN classification method employed four choices of K : 3, 7, half, and all; and three distance measures: EU, TS, and DTW.

Patient	30min			90min			150min		
	Sen.	Spe.	Setting	Sen.	Spe.	Setting	Sen.	Spe.	Setting
1	83.33%	64.75%	3-EU	100.00%	52.94%	half-DTW	100.00%	72.22%	half-EU
2	66.67%	78.22%	all-EU	66.67%	46.38%	all-DTW	66.67%	48.72%	all-TS
3	62.50%	67.65%	all-EU	100.00%	37.04%	all-EU	87.50%	75.00%	half-EU
4	76.47%	66.53%	3-TS	82.35%	60.29%	7-EU	88.24%	55.26%	all-EU
5	57.89%	75.90%	3-TS	73.68%	55.65%	half-DTW	63.16%	49.25%	all-DTW
Ave.	69.37%	70.61%	-	84.54%	50.46%	-	81.11%	60.09%	-
Accuracy	69.99%			67.50%			70.60%		

3.2 Receiver Operating Characteristic Analysis

We employed ROC analysis to further investigate whether the proposed gradient-based updating scheme did improve the prediction power of the system. Instead of a random predictor, we compared the adaptive system with the non-updating system which kept the initial baselines unchanged. Figure 6 plots the averaged AUC values over the five patients for 36 parameter settings (4 choices of $K \times 3$ distance measures $\times 3$ prediction horizons). One can observe that the AUC

Table 3 The summary of sen_{blk} and spe_{blk} obtained from a Poisson random predictor with a mean interval of 1 hour.

Prediction Horizon	30 min		90 min		150 min	
Patient	sen.	spe.	sen.	spe.	sen.	spe.
1	52.00%	39.00%	100.00%	2.10%	100.00%	0.44%
2	58.67%	39.33%	98.67%	0.53%	97.33%	0.55%
3	51.00%	40.63%	99.00%	1.38%	100.00%	0.89%
4	48.47%	40.39%	92.47%	1.47%	94.12%	0.56%
5	54.11%	40.24%	99.79%	0.99%	100.00%	0.30%
Ave.	52.85%	39.92%	97.99%	1.29%	98.29%	0.55%
Accuracy	46.38%		49.64%		49.42%	

values of the adaptive prediction scheme are generally larger than those of the non-updating system. Compared with the non-updating scheme, 34 (out of 36) parameter settings of the adaptive schemes increased the AUC values of the prediction system. In other word, the ROC analysis indicated that the gradient-based learning framework did improve the prediction power of the system. For the adaptive system, the averaged AUC value across 36 settings is 0.60, which is 15% larger than that of the non-updating system.

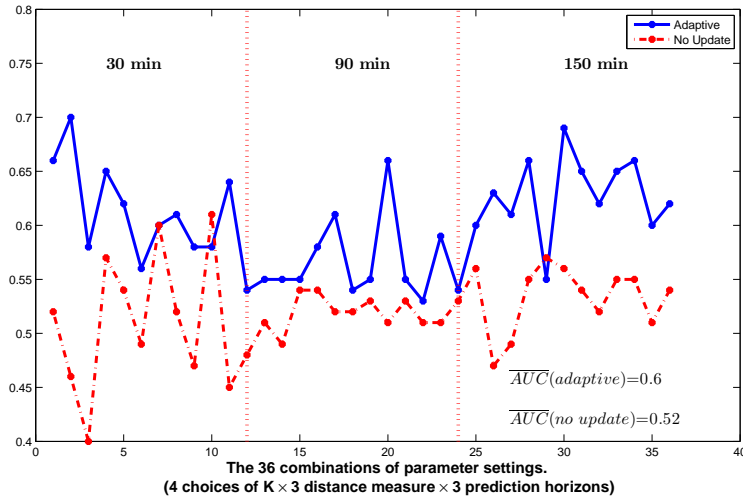


Figure 6 Comparison of the gradient-based self-updating system and the non-updating system in terms of the AUC values, which were averaged over the five patients for each parameter setting. The averaged AUC values across 36 parameter settings were 0.60 and 0.52 respectively for the adaptive and non-updating system.

4 Conclusions

In this paper, we studied the open problem of online seizure prediction based on multichannel EEG data. A gradient-based adaptive seizure prediction framework

was proposed. The adaptive seizure prediction method combined the reinforcement learning concept, online monitoring and adaptive control to autonomously capture the most prominent normal and pre-seizure EEG samples for each individual patient with epilepsy. Starting from the initial normal and pre-seizure baselines, the adaptive updating scheme allows the system to incorporate more and more representative samples and abandon the 'bad' samples in the baselines as the system learns more about a patient. Our experimental results showed that the adaptive prediction scheme could achieve a consistent better prediction performance than a chance model and the non-updating system. The proposed adaptive updating scheme did improve the prediction power of the system. It is noticed that there are only 7 to 23 seizures to train the learning system based on the available EEG recordings. We anticipate that the performance of the proposed reinforcement learning prediction system could be further improved when longer EEG data are available.

The outcome of this study confirmed that the concept of using adaptive learning algorithms to improve the adaptability of seizure prediction is conceivable. A perspective clinical valid seizure prediction system definitely requires both high sensitivity and specificity. If such a seizure warning device is ever to become a reality, adaptive learning techniques will definitely play an important role in handling the great variety of the brain-wave patterns. The long-term goal of this research is to design intelligent machine-learning interfaces that could adaptively predict abnormal mental states for patients with brain diseases. Hypothetically, this prediction system could eventually take the form of an implanted 'brain pacemaker', stimulating the brain to prevent the seizure from happening in its very early stage. We are confident that the adaptive prediction schemes will have a great potential to function well for a wide range of patients.

References

- R. Aschenbrenner-Scheibe, T. Maiwald, M. Winterhalder, H. Voss, J. Timmer, and A. Schulze-Bonhage. (2003) 'How well can epileptic seizures be predicted? An evaluation of a nonlinear method', *Brain*, Vol. 126, pp.2616–2626.
- S. Dangel, P. Meier, H. Moser, S. Plibersek, and Y. Shen. (1999) 'Time series analysis of sleep EEG', *Computer assisted Physics*, pp. 93–95.
- R. Delamont, P. Julu, and G. Jamal. (1999) 'Changes in a measure of cardiac vagal activity before and after epileptic seizures', *Epilepsy Research*, Vol. 35, pp:87–94, 1999.
- C. Elger and K. Lehnertz. (1998) 'Seizure prediction by non-linear time series analysis of brain electrical activity', *European Journal of Neuroscience*, Vol. 10, pp:786–789.
- J. Engel and T. Pedley. (1997) 'Epilepsy: A Comprehensive Textbook. *Lippincott Williams & Wilkins*, Philadelphia, PA.
- L. Iasemidis. (1991) 'On the dynamics of the human brain in temporal lobe epilepsy', *PhD thesis*, University of Michigan, Ann Arbor.

- L. Iasemidis, P. Pardalos, J. Sackellares, and D. Shiau. (2000) 'Measurement and quantification of spatiotemporal dynamics of human epileptic seizures', *Nonlinear Signal Processing*, volume. II, pp.294–318.
- L. Iasemidis, J. Principe, J. Czaplewski, R. Gilman, S. Roper, and J. Sackellares. (1997) 'Spatiotemporal transition to epileptic seizures: A nonlinear dynamical analysis of scalp and intracranial EEG recordings', *Spatiotemporal Models in Biological and Artificial Systems*, pp.81–88. Amsterdam: IOS Press.
- L. Iasemidis, D. Shiau, W. Chaovaitwongse, J. Sackellares, P. Pardalos, J. Principe, P. Carney, A. Prasad, B. Veeramani, and K. Tsakalis. (2003) 'Adaptive epileptic seizure prediction system', *IEEE Transactions on Biomedical Engineering*, Vol. 50, No. 5, pp:616C627.
- K. Jerger, S. Weinstein, T. Sauer, and S. Schiff. (2005) 'Multivariate linear discrimination of seizures', *Journal of Clinical Neurophysiology*, Vol. 116, No. 3, pp.545–551.
- K. Lehnertz and C. Elger. (1998) 'Can epileptic seizures be predicted? evidence from nonlinear time series analysis of brain electrical activity', *Physics Review Letters*, Vol. 80, pp.5019–5022.
- K. Lehnertz and B. Litt. (2005) 'The first international collaborative workshop on seizure prediction: summary and data description', *Clinical Neurophysiology*, Vol. 116, No. 3, pp.493–505.
- B. Litt, R. Esteller, J. Echaz, M. D'Alessandro, R. Shor, T. Henry, P. Pennell, C. Epstein, R. Bakay, and M. Dichter. (2001) 'Epileptic seizures may begin hours in advance of clinical onset: a report of five patients', *Neuron*, Vol. 30, pp.51–64.
- F. Mormann, R. Andrzejak, C. Elger, and K. Lehnertz. (2007) Seizure prediction: The long and winding road. *Brain*, Vol. 130, No. 2, pp.314–333.
- F. Mormann, T. Kreuz, C. Rieke, R. Andrzejak, A. Kraskov, P. David, C. Elger, and K. Lehnertz. (2006) 'On the predictability of epileptic seizures', *Journal of Clinical Neurophysiology*, Vol. 116, No. 3, pp.569–587.
- H. Potter. Anatomy of the brain. <http://faculty.ucc.edu/biology-potter/TheBrain/>, 2006.
- R. Quiroga, J. Arnhold, K. Lehnertz, and P. Grassberger. (2000) 'Kulback-leibler and renormalized entropies: applications to electroencephalograms of epilepsy patients', *Physical review. E, Statistical physics, plasmas, fluids, and related interdisciplinary topics*, Vol. 62, pp.8380–8386.
- M. Quyen, J. Martinerie, M. Baulac, and F. Varela. (1999) 'Anticipating epileptic seizures in real time by non-linear analysis of similarity between EEG recordings', *NeuroReport*, Vol. 10, pp.2149C2155.
- M.L.V. Quyen, V. Navarro, M. Baulac, B. Renault, and J. Martinerie. (2003) 'Anticipation of epileptic seizures from standard EEG recordings', *The Lancet*, Vol. 361, No. 9361, pp.970–971.

- P. Rapp, T. Bashore, J. Martinerie, A. Albano, I. Zimmerman, and A. Mess. (1989) 'Dynamics of brain electrical activity', *Brain Topography*, Vol. 2, pp.99–118, 1989.
- P. Senin. (2008) Dynamic time warping algorithm review. *Technical report*, Information and Computer Science Department University of Hawaii, Honolulu, 2008.
- C. Silva, I. Pimentel, A. Andrade, J. Foreid, and E. Ducla-Soares. (1999) 'Correlation dimension maps of EEG from epileptic absences', *Brain Topography*, Vol. 11, pp.201–209.
- C. Stam. (2005) 'Nonlinear dynamical analysis of EEG and MEG: review of an emerging field', *Clinical Neurophysiology*, Vol. 116, No. 10, pp.2266–2301.
- J. Vastano and E. Kostelich. (1986) Comparison of algorithms for determining Lyapunov exponents from experimental data. *International conference on dimensions and entropies in chaotic systems*, pp.100–107, Pecos River, NM, USA, 1986.