Epilepsy Research (2007) **73**, 213—217

journal homepage: www.elsevier.com/locate/epilepsyres

SHORT COMMUNICATION

Seizure prediction: The impact of long prediction horizons

Björn Schelter[a,b,*], **Matthias Winterhalder**[a,b], **Hinnerk Feldwisch genannt Drentrup**[a,b], **Johannes Wohlmuth**[a,c], **Jakob Nawrath**[c], **Armin Brandt**[c], **Andreas Schulze-Bonhage**[b,c], **Jens Timmer**[a,b]

[a] *FDM, Freiburg Center for Data Analysis and Modeling, University of Freiburg, Eckerstr. 1, 79104 Freiburg, Germany*
[b] *Bernstein Center for Computational Neuroscience Freiburg, University of Freiburg, Hansastr. 9a, 79104 Freiburg, Germany*
[c] *Epilepsy Center, University Hospital of Freiburg, Breisacher Str. 64, 79106 Freiburg, Germany*

Received 31 May 2006; received in revised form 25 August 2006; accepted 3 October 2006
Available online 7 November 2006

KEYWORDS
Seizure prediction;
Seizure anticipation;
Statistical assessment;
Random prediction;
Intervention time;
Prediction horizon

Summary Several procedures have been proposed to be capable of predicting the occurrence of epileptic seizures. Up to now, all proposed algorithms are far from being sufficient for a clinical application. This is, however, often not obvious when results of seizure prediction performance are reported. Here, we discuss impacts of long prediction horizons with respect to clinical needs and the strain on patients by analyzing long-term continuous intracranial electroencephalography data.

1. Introduction

The essential symptom of epilepsy is the apparently unforeseeable occurrence of abnormally synchronized discharges in the cerebral cortex which clinically manifest themselves as seizures. The ability to predict upcoming seizures would augment the therapeutic options considerably. A seizure warning device would already increase the quality of life of those epilepsy patients who cannot be treated successfully by either medications or epilepsy surgery. Additionally, various intervention systems could be applied to suppress seizures in advance of their clinical manifestation, e.g., by delivering short-acting antiepileptic drugs or by applying electric stimulation.

Analyses of invasive and scalp electroencephalography (EEG) recordings using linear and nonlinear time series analysis techniques have provided growing evidence that changes in the EEG dynamics may be detectable prior to seizure onset (cf. Litt and Echauz, 2002, for a review). Such changes have been observed minutes up to several hours in advance of seizure onset (Schindler et al., 2002; Iasemidis et al., 2003; Chaovalitwongse et al., 2005; Le van Quyen et al., 2005; Mormann et al., 2005).

* Corresponding author at: FDM, Freiburg Center for Data Analysis and Modeling, University of Freiburg, Eckerstr. 1, 79104 Freiburg, Germany. Tel.: +49 761 203 7710; fax: +49 761 203 7700.
E-mail address: schelter@fdm.uni-freiburg.de (B. Schelter).

0920-1211/$ — see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.eplepsyres.2006.10.002

Fluctuations in the duration of prediction horizons between a few minutes and several hours as reported so far are of limited use and may affect clinical applicability of a prediction device considerably. A high variance in the prediction horizon might lead to substantial everyday restrictions, since, for instance, a simple warning leaves the patient awaiting in vain an upcoming seizure for hours. A triggered intervention would have to be effective against the upcoming seizure almost immediately and the effect of the intervention would have to last for several hours. Both are rather restrictive constraints not only for warning devices but also for automatic intervention systems. Therefore, it would be advantageous to predict upcoming seizures with a precise temporal resolution, at best a long time interval in advance of the seizure.

To assess the temporal resolution of prediction algorithms, the actual prediction horizon has to be divided into an intervention period, i.e., a period between the detection of seizure precursors and the earliest possible occurrence of a subsequent seizure, and an occurrence period, i.e., a period during which the seizure is predicted to occur. Different lengths of both time intervals should be evaluated with respect to sensitivity and specificity. The only limitation is that the intervention would have to become effective within the length of the intervention time and would have to last for the duration of the seizure occurrence period. If the intervention needs less time to become effective, the application of the intervention could be shifted in time.

Moreover, a prediction algorithm does usually raise not only correct but also false alarms. If the fraction of false alarms is comparable to the number of true alarms, long prediction horizons are particularly unfavorable (Mormann et al., 2006). For instance, for three false alarms per day and prediction horizons up to 4 h, the patient will be awaiting seizures that will never occur half of the day. This problem becomes even worse if only a fraction of the seizures were predicted correctly. This strengthens the necessity of short seizure occurrence periods, which does restrict the length of the intervention times by no means.

In this short communication, we present the assessment of a prediction method based on synchronization theory. We investigate the trade-off between intervention times and occurrence periods on the basis of long-term continuous intracranial EEG data sets of four patients extending over several days. Having in mind a therapy based on a prediction algorithm, we estimate the fraction of false alarms as well as the ''false warning time'', i.e., the ratio of time the patient is awaiting a seizure that will never occur, additionally. To control the performance of the prediction method, we apply a statistical test in order to check whether the achieved performance is indeed better than that of a random predictor (Schelter et al., 2006).

2. Materials and methods

In this study, we evaluated the mean phase coherence (Mormann et al., 2003) on long-term continuous intracranial EEG data. The analyzed data base consists of recordings of four patients suffering from medically intractable focal epilepsy. The EEG data were acquired using a Neurofile NT digital video EEG system with 128 channels, 256 Hz or 512 Hz sampling rate, and a 16 bit analogue-to-digital converter. To eliminate possible line noise and low frequency components, the EEG data sets were preprocessed by a 50 Hz notch filter, a high pass filter at 0.5 Hz, and an anti-aliasing filter. A subset of electrode contacts was selected prior to the analysis by visual inspection by an experienced electroencephalographer. Three focal electrode contacts, i.e., three recording sites initially involved in ictal activity based on the available electrode coverage of the brain, and three extra-focal electrode contacts, i.e., recording sites not involved at all or — in most cases — latest during spread of ictal activity, were selected for analysis. Altogether 105 seizures and 699 h interictal data were assessable. Details about the patients and the EEG data sets analyzed in this study are given in Table 1.

The mean phase coherence was applied to the data using a moving window with duration of 32 s. A causal median smoother of 4 min duration was applied afterwards. The prediction performance was assessed using the seizure prediction characteristic $S(FPR_{max}, IT, SOP)$ which is defined as the functional relationship between sensitivity S and the maximum false prediction rate FPR_{max}, the intervention time IT — in Winterhalder et al. (2003) originally referred to as seizure prediction horizon SPH — and the seizure occurrence period SOP (Winterhalder et al., 2003). Furthermore, we estimated the false warning time, i.e., the ratio of time the

Table 1 EEG data characteristics

Patient	Age	Sex	Focus localization	Preictal			Interictal	
				Number of seizures available	Minimum number of seizures analyzed	Maximum number of seizures analyzed	Minimum interictal time (h)	Maximum interictal time (h)
pat01	28	m	NC	28	7	25	150	185
pat02	50	m	HC	34	8	23	185	220
pat03	31	f	NC	15	4	14	141	165
pat04	18	m	NC	28	5	24	94	129
Total				105	24	86	570	699
Mean				26	6	22	143	175

Different numbers of seizures as well as different interictal reference periods are assessed depending on the length of the evaluated intervention time and occurrence period. Focus localization: neocortex (NC), hippocampus (HC).

patient is awaiting seizures that will never occur, as well as the fraction of false alarms with respect to the total number of alarms. To test the statistical significance of the prediction efficacy, we compared sensitivity values obtained by the algorithm to critical sensitivity values for the random predictor calculated on the basis of an analytic approach. The test statistics is derived from a random predictor generating alarms following a Poisson process in time without using any information from the EEG (Schelter et al., 2006). Two critical sensitivity values are utilized. Since the interdependence between the features of the investigated channel combinations is unknown, the first one assumes complete dependence of the channel combinations, thus, the feature is assumed to be one-dimensional. The corresponding critical sensitivity is referred to as lower critical sensitivity. The second critical value is corrected for the multiple tests that have to be performed when choosing the best out of the 15 channel combinations between the focal and extra-focal electrode contacts assuming complete independence, i.e., the upper critical sensitivity. Concerning the tests for multiple values of SOP and IT that have been evaluated, the number of significant results is compared with the expected number of significant results. This in-sample optimization provides evidence whether the prediction method is superior to a random predictor. A 5% significance level has been chosen for all tests.

3. Results

The results are depicted in Fig. 1. For all four patients, the sensitivity of the prediction method is shown in the first row, the lower critical sensitivity of the random predictor in the second row, the upper critical sensitivity of the random pre-

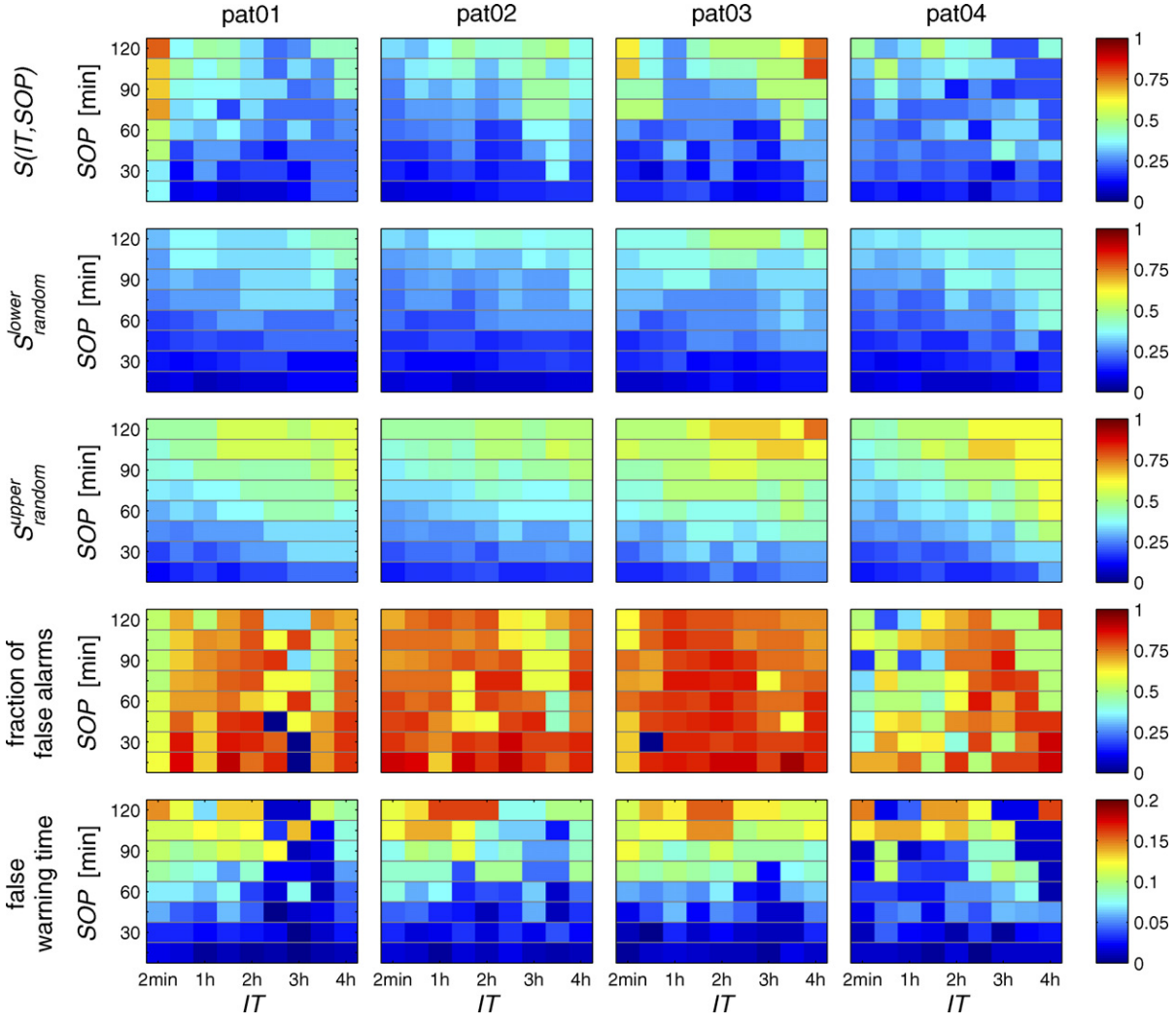

Fig. 1 Sensitivity of the prediction method is shown in the first row, the sensitivity of the random predictor in the second and third row, the fractions of the false predictions with respect to the total number of predictions in the fourth row, and the false warning time in the fifth row for four patients (columns). The seizure occurrence period SOP is varied between 15 min and 120 min and the intervention time IT between 2 min and 4 h and 2 min. The maximum false prediction rate was set to two false predictions per day. The performance of the random predictor is subdivided in a lower sensitivity, i.e., the sensitivity that can be achieved assuming complete dependence between the 15 channel combinations, and an upper sensitivity, i.e., the sensitivity that can be obtained assuming complete independence of the 15 channel combinations.

dictor in the third row, the fractions of the false predictions with respect to the total number of predictions in the fourth row, and the false warning time in the fifth row. The seizure occurrence period SOP is varied between 15 min and 120 min and the intervention time IT between 2 min and 4 h and 2 min in order to comprise the broad range of prediction horizons suggested in the literature. The maximum false prediction rate was restricted to two false predictions per day in order to ensure that only part of the day is covered by intervention times and occurrence periods. A statistically significant prediction performance is obtained for several parameter values of IT and SOP, for all four patients compared to the lower and upper sensitivity of the random predictor. For patient 1 the fraction of significant sensitivities is 42% for the lower sensitivity and 13% for the upper sensitivity, i.e., the number of parameter combinations for which a significant sensitivity was obtained divided by the 72 possible parameter combinations. The significance level of the test based on the random predictor was 5%, thus, we expect a fraction of 5% due to pure chance. Fractions higher than 5% are assumed to reveal results of a prediction technique that can be considered to be superior to a random prediction, especially, if this holds compared to the upper sensitivity of the random predictor. The fractions for the remaining patients are: patient two 46% (6%), patient three 35% (7%), and patient four 40% (4%) for the comparison with the lower (upper) sensitivity of the random predictor.

However, the combination of IT and SOP corresponding to the best performance nevertheless varies considerably between patients. The fraction of false alarms can achieve values close to 100% which corresponds to the fact that almost all alarms are false ones. Averaging all results for all patients more than 50% of all alarms are false ones. The analysis of the false warning time shows that patients would be up to 15% of the day awaiting seizures that will never occur. In general, there are considerable differences in all quantities depending on SOP and IT. For instance, for patient 1, choosing an intervention time of 2 min but an occurrence period of 120 min, a sensitivity of almost 80% can be achieved at the cost of a fraction of the false warning time of 15% and a fraction of false alarms of 50%. Decreasing the occurrence period by one half leads to a decrease in sensitivity of 30%, but the fraction of the false warning time is also decreased by one half, even though the fraction of false alarms is slightly increased. It might be expected in the first place that decreasing the occurrence period by a factor of two leads to halving the fraction of the false warning time. But increasing the intervention time by 30 min leads to a similar fraction of false warning time for this patient, even though there is no direct influence of the intervention time onto the false warning time.

4. Conclusions

By means of an extensive continuous long-term EEG data base we demonstrated that in addition to the sensitivity and specificity of a prediction method and its comparison with a random predictor also additional parameters, i.e., the fraction of false predictions and the ratio of the false warning time, are of particular interest to quantify the strain on patients. Evaluating all four proposed quantities allows a detailed characterization of the seizure prediction method with respect to its implications for the day-to-day life of patients. To this aim, it is essential to estimate all parameters for a certain spectrum of intervention times and occurrence periods, especially if clinical needs are not restrictive. For instance, if an intervention became effective exactly in 5 min and lasted for a few seconds, the two time intervals would be completely fixed. Since this is not expected in applications, ranges of the time intervals should be evaluated. The ranges of time intervals reflect different possible settings for prediction-based interventions. Single numbers might yield intolerable conclusions about the prediction performance.

The achieved sensitivity values depending on IT and SOP are higher than the performance of a random predictor for several combinations of IT and SOP. The achieved number of significant sensitivities for three out of four patients substantiates that the mean phase coherence can be considered to be superior to a random prediction. For patient number 4, superiority to a random predictor cannot finally be verified, especially if taking into account that an in-sample test had to be performed. In such cases, future out-of-sample evaluations have to clarify superiority compared to a random prediction.

Sensitivity values of up to 70% that are achieved have to be put in perspective, when the two additional characteristic quantities are inspected. Fractions of false alarms higher than 50% on average will hamper a broad acceptance among patients. In this study, we have chosen a maximum false prediction rate of two false predictions per day, which is lower than that chosen in several published articles, which are in the order of three to four false predictions per day (cf. Litt et al., 2001; Navarro et al., 2002; Winterhalder et al., 2003; Aschenbrenner-Scheibe et al., 2003; Maiwald et al., 2004; Chaovalitwongse et al., 2005; Iasemidis et al., 2005). The fraction of false alarms increases when the maximum false prediction rate is increased. Strategies to avoid false alarms are, thus, highly desirable. Fist approaches to this aim are discussed in Schelter et al. (in press).

When allowing at most two false predictions per day, the maximum fraction of the false warning time is $(2 \times 2\,\mathrm{h})/24\,\mathrm{h} = 16.7\%$ for a seizure occurrence period of 2 h assuming a seizure free day. This maximum has been observed only for few parameter combinations for all four patients. The fraction of 16.7% a day during which a patient is falsely awaiting a seizure is much better than 100% uncertainty. However, while this result seems to be very promising in the first place, we want to stress that the patient would be falsely awaiting seizures for almost 4 h every single day. Under the assumption that the patient undergoes only a few seizures every month, where a certain fraction of seizures is not predicted correctly since sensitivity is considerably lower than 100%, the false warning time of 4 h a day has to be put in perspective. Thus, if the sensitivity is considerably lower than 100% a device build upon the corresponding prediction algorithm might not be accepted. In contrast, if the patient underwent seizures frequently and sensitivity of the prediction method was high, a false warning time of 4 h a day could be acceptable.

Nevertheless, the actual performance cannot be revealed by discussion of the upper limits of certain parameters. Therefore, we evaluated these parameters on an extensive

EEG data base and thereby showed that results are usually better than expected in the theoretical worst-case scenario. Fractions of false warning times considerably lower than 10% have indeed been observed which provides a promising result for epilepsy patients.

Additionally, values in the order of 16% or lower are quite uncommon. For instance, Mormann et al. (2006) estimated the fraction of the false warning time to be 63% for an exemplary study by Iasemidis et al. (2005). The superior results in our study are mainly based on the fact that we divided the prediction horizon into an intervention time and an occurrence period. The advantage is that the former interval does not contribute to the false warning time. Collapsing both time intervals into one single prediction horizon would result in a fraction of maximum false warning time of $(2 \times (4\,h + 2\,h))/24\,h = 50\%$ for a seizure occurrence period of 2 h and an intervention time of 4 h assuming a seizure free day. A further increase of the maximum false prediction rate does imply that the patient will be confronted with intervention times and seizure occurrence periods during an unacceptable part of the day since during monitoring the days are usually not seizure free. In other words, the particular influence of IT and SOP on sensitivity, the random predictor, the fraction of false alarms, and the false warning time can be disentangled by considering the two time intervals separately; instead of comprising both, IT and SOP, into one single ''prediction horizon''.

We mention though that especially with respect to the false warning time, the coarse view of one prediction horizon leads to an artificial decrease in performance. With respect to sensitivity the performance is artificially increased by utilizing just one prediction horizon. Thus, distinguishing IT and SOP is essential when assessing seizure prediction efficacy. A minimization of the fraction of the false warning time by differentiating IT and SOP enhances the acceptance of seizure prediction devices. The assessment methodology utilized here in combination with a statistical test procedure and the proposed extensions, i.e., the fraction of false alarms with respect to the total number of alarms and the false warning time, enables a patient-individual, statistically verifiable, and clinically motivated selection of optimal prediction parameters.

Acknowledgments

We would like to thank Carolin Gierschner and the team of EEG technicians at the Epilepsy Center Freiburg for their technical support, especially concerning the extensive data acquisition and handling.

This work was supported by the German Federal Ministry of Education and Research (BMBF grant 01GQ0420) and the German Science Foundation (Ti 315/2-2; Sonderforschungsbereich-TR3).

EEG data availability: EEG data sets from epilepsy patients are available on request: Web page: http://www.fdm.uni-freiburg.de/EpilepsyData; E-mail: epilepsydatabase@fdm.uni-freiburg.de.

References

Aschenbrenner-Scheibe, R., Maiwald, T., Winterhalder, M., Voss, H.U., Timmer, J., Schulze-Bonhage, A., 2003. How well can epileptic seizures be predicted? An evaluation of a nonlinear method. Brain 126, 2616—2626.

Chaovalitwongse, W., Iasemidis, L.D., Pardalos, P.M., Carney, P.R., Shiau, D.S., Sackellares, J.C., 2005. Performance of a seizure warning algorithm based on the dynamics of intracranial EEG. Epilepsy Res. 64, 93—113.

Iasemidis, L.D., Shiau, D.S., Chaovalitwongse, W., Sackellares, J.C., Pardalos, P.M., Principe, J.C., Carney, P.R., Prasad, A., Veeramani, B., Tsakalis, K., 2003. Adaptive epileptic seizure prediction system. IEEE Trans. Biomed. Eng. 50, 616—627.

Iasemidis, L.D., Shiau, D.S., Pardalos, P.M., Chaovalitwongse, W., Narayanan, K., Prasad, A., Tsakalis, K., Carney, P.R., Sackellares, J.C., 2005. Long-term prospective on-line real-time seizure prediction. Clin. Neurophysiol. 116, 532—544.

Le van Quyen, M., Soss, J., Navarro, V., Robertson, R., Chavez, M., Baulac, M., Martinerie, J., 2005. Preictal state identification by synchronization changes in long-term intracranial EEG recordings. Clin. Neurophysiol. 116, 559—568.

Litt, B., Esteller, R., Echauz, J., D'Alessandro, M., Shor, R., Henry, T., Pennell, P., Epstein, C., Bakay, R., Dichter, M., Vachtsevanos, G., 2001. Epileptic seizures may begin hours in advance of clinical onset: a report of five patients. Neuron 30, 51—64.

Litt, B., Echauz, J., 2002. Prediction of epileptic seizures. Lancet Neurol. 1, 22—30.

Maiwald, T., Winterhalder, M., Aschenbrenner-Scheibe, R., Voss, H.U., Schulze-Bonhage, A., Timmer, J., 2004. Comparison of three nonlinear seizure prediction methods by means of the seizure prediction characteristic. Physica D 194, 357—368.

Mormann, F., Kreuz, T., Andrzejak, R.G., David, P., Lehnertz, K., Elger, C.E., 2003. Epileptic seizures are preceded by a decrease in synchronization. Epilepsy Res. 53, 173—185.

Mormann, F., Kreuz, T., Rieke, C., Andrzejak, R.G., Kraskov, A., David, P., Elger, C.E., Lehnertz, K., 2005. On the predictability of epileptic seizures. Clin. Neurophysiol. 116, 569—587.

Mormann, F., Elger, C.E., Lehnertz, K., 2006. Seizure anticipation: from algorithms to clinical practice. Curr. Opin. Neurol. 19, 187—193.

Navarro, V., Martinerie, J., Le van Quyen, M., Clemenceau, S., Adam, C., Baulac, M., Varela, F., 2002. Seizure anticipation in human neocortical partial epilepsy. Brain 125, 640—655.

Schelter, B., Winterhalder, M., Maiwald, T., Brandt, A., Schad, A., Schulze-Bonhage, A., Timmer, J., 2006. Testing statistical significance of multivariate time series analysis techniques for epileptic seizure prediction. Chaos 16, 013108.

Schelter, B., Winterhalder, M., Maiwald, T., Brandt, A., Schad, A., Timmer, J., Schulze-Bonhage, A. Do false predictions of seizures depend on the state of vigilance? A report from two seizure prediction methods and proposed remedies. Epilepsia, in press.

Schindler, K., Wiest, R., Kollar, M., Donati, F., 2002. EEG analysis with simulated neuronal cell models helps to detect pre-seizure changes. Clin. Neurophysiol. 113, 604—614.

Winterhalder, M., Maiwald, T., Voss, H.U., Aschenbrenner-Scheibe, R., Timmer, J., Schulze-Bonhage, A., 2003. The seizure prediction characteristic: a general framework to assess and compare seizure prediction methods. Epilepsy Behav. 4, 318—325.