

A multi-feature and multi-channel univariate selection process for seizure prediction

Maryann D'Alessandro^{a,*}, George Vachtsevanos^b, Rosana Esteller^c, Javier Echaz^c, Stephen Cranstoun^a, Greg Worrell^d, Landi Parish^a, Brian Litt^{a,e}

^aDepartment of Bioengineering, University of Pennsylvania, 747 West Madison Circle, Pittsburgh, Philadelphia, PA 15229, USA

^bGeorgia Institute of Technology, School of ECE, Atlanta, GA, USA

^cNeuropace Inc., 10933 Crabapple Rd., Roswell, GA, USA

^dMayo Clinic, 200 First Street SW, Rochester, MN, USA

^eDepartment of Neurology, Hospital of the University of Pennsylvania, Philadelphia, PA, USA

Accepted 3 November 2004

Abstract

Objective: To develop a prospective method for optimizing seizure prediction, given an array of implanted electrodes and a set of candidate quantitative features computed at each contact location.

Methods: The method employs a genetic-based selection process, and then tunes a probabilistic neural network classifier to predict seizures within a 10 min prediction horizon. Initial seizure and interictal data were used for training, and the remaining IIEEG data were used for testing. The method continues to train and learn over time.

Results: Validation of these results over two workshop patients demonstrated a sensitivity of 100%, and 1.1 false positives per hour for Patient E, using a 2.4 s block predictor, and a failure of the method on Patient B.

Conclusions: This study demonstrates a prospective, exploratory implementation of a seizure prediction method designed to adapt to individual patients with a wide variety of pre-ictal patterns, implanted electrodes and seizure types. Its current performance is limited likely by the small number of input channels and quantitative features employed in this study, and segmentation of the data set into training and testing sets rather than using all continuous data available.

Significance: This technique theoretically has the potential to address the challenge presented by the heterogeneity of EEG patterns seen in medication-resistant epilepsy. A more comprehensive implementation utilizing all electrode sites, a broader feature library, and automated multi-feature fusion will be required to fully judge the method's potential for predicting seizures.

© 2004 Published by Elsevier Ireland Ltd on behalf of International Federation of Clinical Neurophysiology.

Keywords: Multiple channels; Multiple features; Feature extraction; Seizure prediction; Classification

1. Introduction

Since the early 1970s, researchers have searched for independent features and reproducible patterns that may herald seizure onset. These features are not chosen randomly. Rather, their selection is hypothesis driven, based upon what has been observed in the interictal IIEEG, what changes in cellular function might be occurring in

the ictal onset zone prior to seizure onset, and our knowledge that some aspects of brain activity can be modeled as a chaotic, non-linear dynamic system (Iasemidis et al., 1988; Lehnertz et al., 2001; Lehnertz and Elger, 1998; Litt and Echaz, 2002). Researchers now agree that the EEG signal of an epileptic patient represents not only 'seizure' and 'interictal' states, but a more complex behavior that includes a 'preictal' stage. We have hypothesized a cascade of events that may wax and wane over a variety of prediction horizons that together could provide a probabilistic model regarding the potential for seizure occurrence (Litt et al., 2001; Litt and Echaz, 2002). The methodology

* Corresponding author. Tel.: +1 412 366 2580; fax: +1 412 766 8027.
E-mail address: mdalessandro@cdc.gov (M. D'Alessandro).

presented here evaluates a short-term prediction horizon, focusing on events that occur up to ten minutes before electrographic seizure onset.

We present a method using computational techniques to select an optimized ‘seizure predictor’ from a candidate set of intracranial electrode contacts and quantitative features found to be useful for this purpose. Several investigators have evaluated preictal entrainment among multiple electrode contacts as a means for predicting seizures (Iasemidis et al., 1996; Iasemidis and Sackellares, 2001), but historically the investigations have been limited to evaluation contacts around the epileptic focus (i.e. ictal onset zone) (Litt and Echauz, 2002; Petrosian et al., 2000; Petrosian and Homan, 1994). The approach we implemented for the Bonn Workshop selects features and electrode contacts within or outside of the ictal onset zone, based on an abbreviation of the methodology reported in (D'Alessandro et al., 2003). The original methodology evaluates every monitored electrode contact using a set of quantitative features derived from the intracranial EEG to determine the best combination of electrode sites and features for seizure prediction. To meet time constraints imposed upon the workshop participants for data analysis, fewer features and a smaller data training set were used for this analysis. Specifically, we selected the features that demonstrated the most predictive power in the original methodology (D'Alessandro et al., 2003).

2. Methods

The algorithm employed was trained on preictal data leading to two and four seizures for patients B and E, respectively, 4 h of baseline EEG data for each patient, followed by selection of a set of features and electrodes using a genetic algorithm and probabilistic neural network (PNN) to adapt the method to individual patients. The trained system was then run on the remaining body of intracranial EEG (IEEG) data for each patient, and performance was assessed. This study evaluated the predictive value of eight bipolar intracranial EEG channels and a small set of candidate features found to be useful for detecting seizure precursor events in past experiments. A systematic approach to feature selection, classification, and validation to predict seizures was applied to both data sets. The chronologically first (leading) seizure in each cluster, and 4 h of baseline data selected early in the hospital stay, were used for training. The remaining data were used for validation. Performance was evaluated for all IEEG data and seizures not used for training. Seizure clusters were included in the validation set due to the limited number of leading seizures available. Performance was assessed by evaluating sensitivity and the number of false positives per hour (FPh).

2.1. Data and preprocessing

Intracranial EEG data provided for analysis at the First Seizure Prediction Workshop from the University of Pennsylvania (Patient E) and the University of Bonn (Patient B) as described in the accompanying workshop summary paper (Lehnertz and Litt, 2005) were evaluated for this study. Ten-minute data epochs before onset of each leading seizure in the training set, from eight IEEG channels were selected for training to address the short-term prediction horizon. A *leading seizure* was defined as the first in a cluster of seizures with no seizures occurring 3 h prior, or a seizure whose onset is at least 3 h removed from any other seizure activity. Only leading seizures were used for training, since clustered seizures have been demonstrated to have different preictal characteristics than leading seizures (Litt et al., 2001). The 3 h criteria were based upon recent results indicating that at least this temporal separation, and preferably 4 h, is required to best observe seizure precursors (Litt et al., 2001). Given the close clustering of seizures in the data shared in the workshop, a separation of 3 h was deemed acceptable. In total, there eight leading seizures for Patient E and four leading seizures for Patient B. Half of the leading seizures for each patient were used for training, while the remaining preictal segments were reserved for validation and testing.

For training purposes, data epochs were considered ‘baselines’ if they were at least 3 h removed from the onset or ending of an electrographic seizure. Our original methodology described in (D'Alessandro et al., 2003), used baselines from both the awake and asleep state during training. Four hours of baseline data were used during training in the method described here, with no regard to the state of consciousness, as sleep-wake markings were not available with data provided in the workshop.

The final validation set for Patient E included 11 preictal segments (2 leading seizures and nine clustered seizures) and approximately 55 h of baseline data; and eight preictal segments (2 leading seizures and six clustered seizures) and approximately 85 h of baseline data for Patient B. Patient E's last two leading seizures and the CDs containing them were not evaluated since the CDs on which these seizures were recorded had inconsistent amplifier gains.

2.2. Electrode contact selection

Due to the limited time available to evaluate all monitored electrode contacts, we empirically selected a subset of electrode contacts for evaluation and pruned our original feature subset to include a smaller number of features identified in previous studies as computationally efficient and potentially useful preictal indicators.

Fig. 1 shows the electrode contacts (analyzed as bipolar signals) evaluated for the study patients. In a four-patient study reported in (D'Alessandro, 2003), the electrode contact most closely associated with the majority of seizures

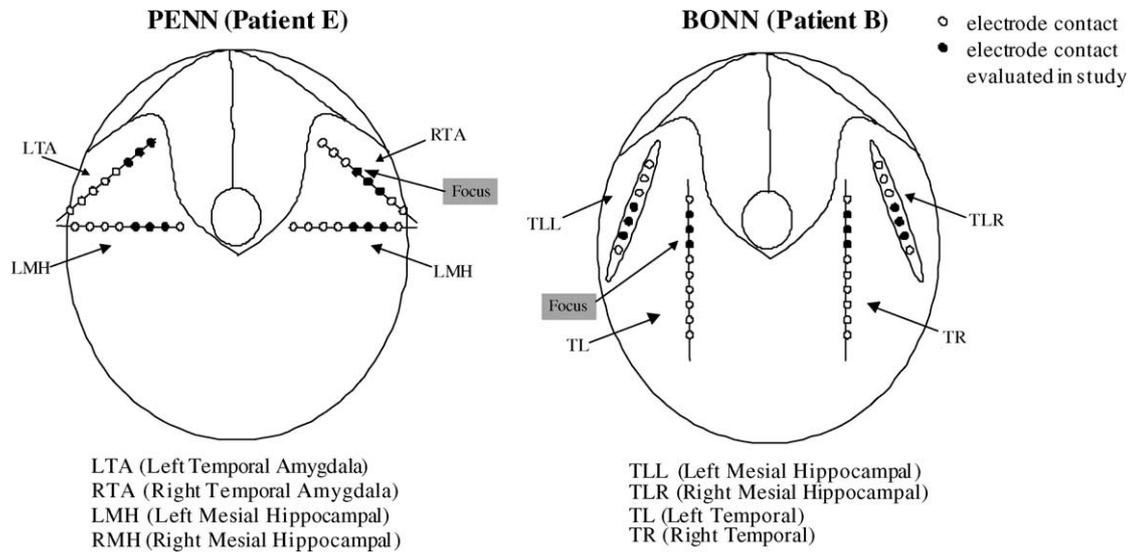


Fig. 1. Selected brain regions evaluated for each patient. Electrode contact locations are approximate and nomenclature is based on the nomenclature used for each monitoring facility. Electrode contact locations (processed as bipolar signals) shown in black were evaluated in this experiment. Arrows point to particular regions and to the focus channel for each patient.

was never selected as the ‘best’ overall channel. In three of the four patients, the ‘best’ channel identified for seizure prediction for this relatively short prediction horizon was found contralateral to the focus channel. Furthermore, several channels exhibited low amplitudes minutes prior to onset. Consequently, the results reported in (D’Alessandro, 2003) were used as a guideline for selecting the channels analyzed for the study reported here. Specifically, contacts both within the seizure onset zone and contralateral to it were selected for processing. Channels near these regions that exhibited lower amplitude but normal electrode impedances were also selected.

The following electrode contacts were used to select the bipolar signals analyzed: (1) focus electrode contact (site of earliest onset read by two boarded electroencephalographers (BL, GW)), (2) electrode adjacent to the focus electrode, (3) electrode contralateral to focus electrode, (4) electrode adjacent to selection 3, (5) low amplitude electrode contralateral to focus electrode, (6) electrode adjacent to selection 5, (7) electrode contralateral to selection 5, (8) electrode adjacent to selection 7.

2.3. Pre-feature selection and processing

In (D’Alessandro, 2003), we conducted a search to select the best among 25,872 features computed over all monitored electrode contacts. The curve length and energy-based features provided the most distinguishability between preictal and baseline signals for the four patients evaluated, all of whom had partial onset temporal lobe epilepsy. Therefore, in the present study we limited the features to the curve length, energy, and nonlinear energy. The reduced feature set was examined to determine if the results obtained in (D’Alessandro, 2003) can be generalized

to any patient studied. The reduced feature set resulted in a search space of 800 features.

Processing was performed using a three step approach, beginning with first-level feature extraction from the raw data. Second-level features were extracted from first-level features (e.g. the mean of the energy), and finally third-level features extracted from second-level features (e.g. the minimum of the mean of the energy). As more levels of processing were completed, the prediction time horizon decreased. This study used window length (L) = 2000 points and a displacement (d) = 500 points for first-level feature extraction, and $L=24$ and $d=1$ for second- and third-level feature extraction. Additional levels of feature extraction required longer feature initialization periods. The objective was to address the 10-min prediction horizon; however, after three levels of feature extraction, the prediction horizon was reduced to approximately eight minutes, due to the increased data requirement for processing. The mathematical expressions and derivations of individual features are discussed in Appendix A.

2.4. Genetic and classifier based performance evaluation

After preprocessing (60 Hz filtering, bipolar montaging) and processing (feature computation for each electrode site), a genetic algorithm selected reasonable features off-line from the preselected group of features to serve as inputs to the classifier based feature selection process (see Appendix A). The genetic algorithm search flowchart is shown in Fig. 2. The number of chromosomes generated at each iteration was equal to three times the length of the chromosome. This process went through a number of iterations (also equal to three times the length of the chromosome, see below). The objective was to obtain

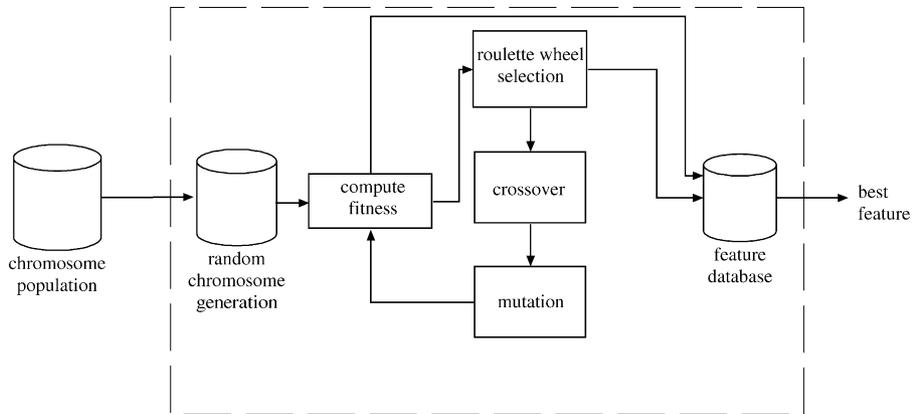


Fig. 2. A selected number of random chromosomes are generated from the available population. Each chromosome’s fitness is computed by comparing the class separability between the feature value generated from class 1 (preictal) chromosome signals to the value generated from class 2 (no pre seizure).

a variety of different curve length and energy based features for analysis.

Genetic algorithms are smart search processes inspired by biological evolution (Chang et al., 1990). Precisely, each possible solution is represented by a coded string of bits called a ‘chromosome’. The 11-bit genetic algorithm chromosome used in this work is shown in Fig. 3. Here, each chromosome contains four ‘genes’ (1) electrode contacts or channels, (2) first-level features, (3) second-level features (these are derived from the first-level features by taking overlapping windowed values such as the mean, minimum, maximum, etc., of the windowed data segments, see Appendix A), (4) third-level features (derived from the 2nd-level features in the same way (e.g. the variance of the mean of the first level feature, see Appendix A). A selected number of chromosomes are generated from the population to identify reasonable features to serve as inputs to the classifier-based selection process.

The degree of ‘fitness’ is computed by comparing the class separability between the feature generated from the class 1 (preictal) chromosome signals to the feature generated from the class 2 (no pre seizure) chromosome signals. The effectiveness or ‘fitness’ of each individual

feature is measured using a probabilistic neural network (PNN) to compare the ‘pre seizure’ and the ‘no-pre seizure’ classes. The best performing feature and the corresponding channel were then evaluated using the method described in (D’Alessandro et al., 2003) to determine the prediction block length used to validate the results.

The preictal data were segmented causally into ‘training’ and ‘testing’ data. Ideally, the baseline data also would be segmented causally; however, enough baseline data segments were not available in each patient. Therefore, four 1-h baselines located chronologically early in the dataset were used for training, while the entire dataset was evaluated for testing purposes.

The figure of merit (FOM) used to evaluate the performance of each feature was a weighted representation of standard performance metrics based on the classifier outputs of the training data (see below). This figure could be adjusted for different seizure types and severities, for example the cost of false negatives could be made extremely high for patients with generalized convulsions, or frequent injuries due to their seizures. The weighting used in this study is based on a general perspective regarding the importance of the preictal versus baseline classifications. The FOM is calculated

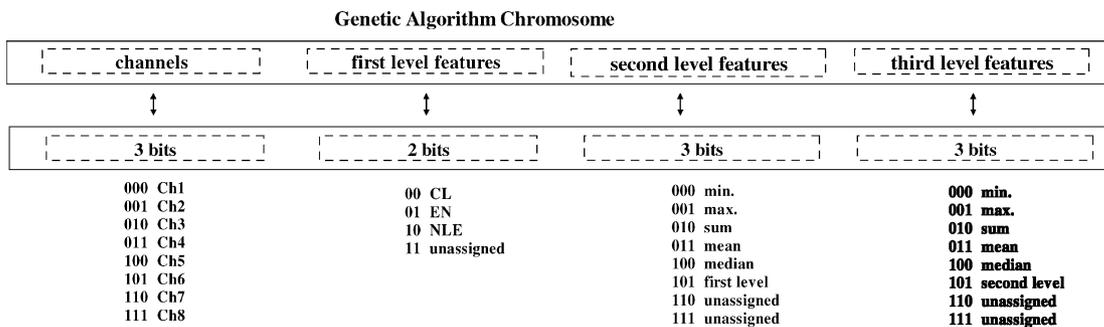


Fig. 3. Genetic algorithm chromosome representing the electrode contacts (channels) and features potentially useful for seizure prediction. Each 11-bit chromosome represents a unique electrode contact and processed signal available for analysis. The chromosome structure is flexible and is easily expanded or compressed to accommodate a particular number of channels and features for investigation. Unassigned values are given a high penalty when selected during the search process.

as follows:

$$\text{FOM} = 0.55P_{\text{TP}} + 0.45P_{\text{TN}} - 0.55P_{\text{FN}} - 0.45P_{\text{FP}} \quad (1)$$

where:

P_{TN} = probability of true negative or correct no pre-seizure classification

P_{TP} = probability of true positive or correct pre-seizure classification

P_{FN} = probability of false negative or incorrect pre-seizure classification

P_{FP} = probability of false positive or incorrect no pre-seizure classification

This equation can be reduced to

$$\text{FOM} = 1 - 1.1P_{\text{FN}} - 0.9P_{\text{FP}} \quad (2)$$

and is easily used to rank the genetically selected features. The FOM is designed to select the best feature combinations, compared to performance of the training set. The number of different combinations of features and electrodes tested was limited to $3^*(\text{length of the chromosome})$ after preliminary experiments showed that this number of trials was sufficient to reach a 'reasonable' answer.

2.5. Classification

At the classifier stage, the probabilistic neural network (PNN) classifier described in [Echauz et al. \(2000\)](#) and [D'Alessandro \(2001\)](#) assigned the output of the feature vector into the class pre-seizure or no-pre-seizure. The advantage of the PNN is that its members can train quickly and it is a nonparametric classifier which suits most of our classification needs. The PNN input space was divided into two classes, and used static training based on a priori baseline preictal data. The PNN outputs from the training data were used to select the block length required for prediction.

2.6. Performance assessment

We chose to declare seizure predictions for blocks of data, for example 1-minute long, rather than for every point. In this way the analysis avoided numerous conflicting predictions over very short periods of time. The block lengths were empirically selected based on the classifier outputs for the training data. False positives (FP) were counted by identifying the number of incorrect classifications found in the data set. One FP was counted for each toggled incorrect classification. True positive (TP) classifications were counted in the minutes an electrographic seizure onset. If an incorrect classification preceded the prediction horizon without toggling off before entering the prediction horizon, an FP was declared, the classification was not identified as preictal unless it was classified as no pre-seizure before entering the prediction horizon.

The block length for Patient E (Penn) was set to one point (2.48 s), since the preictal and baseline training records were correctly classified (yielded no FPs) or false negatives (FNs)). The block length for Patient B (Bonn) was determined by averaging the length of the FP and FN data segments in the training data. The block length for Patient B was set to 5 min. This method provides a clinically sound approach that demonstrates the calculation of data blocks to obtain the sustained prediction requirement.

3. Results

The prediction methodology was applied to the complete records of two patients and the resultant features and electrode contacts selected are shown in [Fig. 4](#). In patient E, the best channel selected was contralateral to the focus channel. The selected feature and electrode contact for patient E was the 'mean of the mean of the curve length for bipolar channel LA1–LA2.'

The feature and classifier outputs for the Patient E training data are shown in [Fig. 5a](#). A PNN output value equal to '0' indicates the no pre-seizure class while a '1'

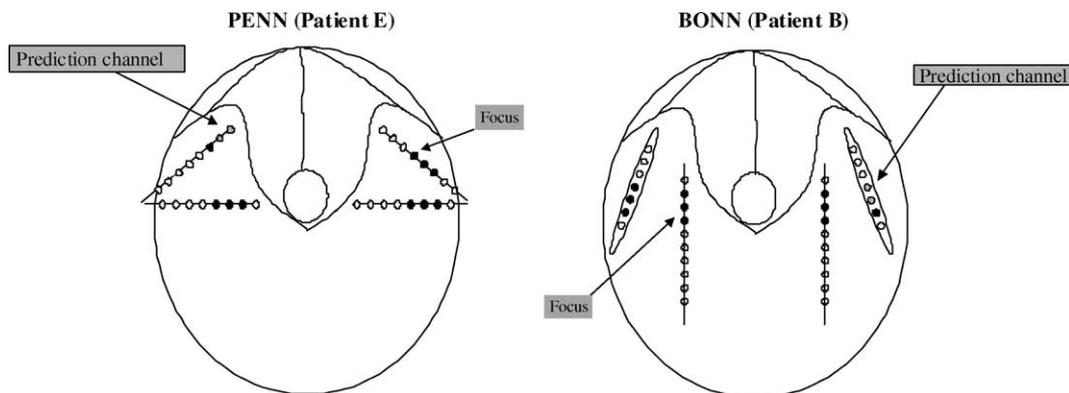


Fig. 4. Channel selections for seizure prediction. Figures represent brain regions evaluated for this study. Specific electrode contacts analyzed in bipolar montage are in solid black.

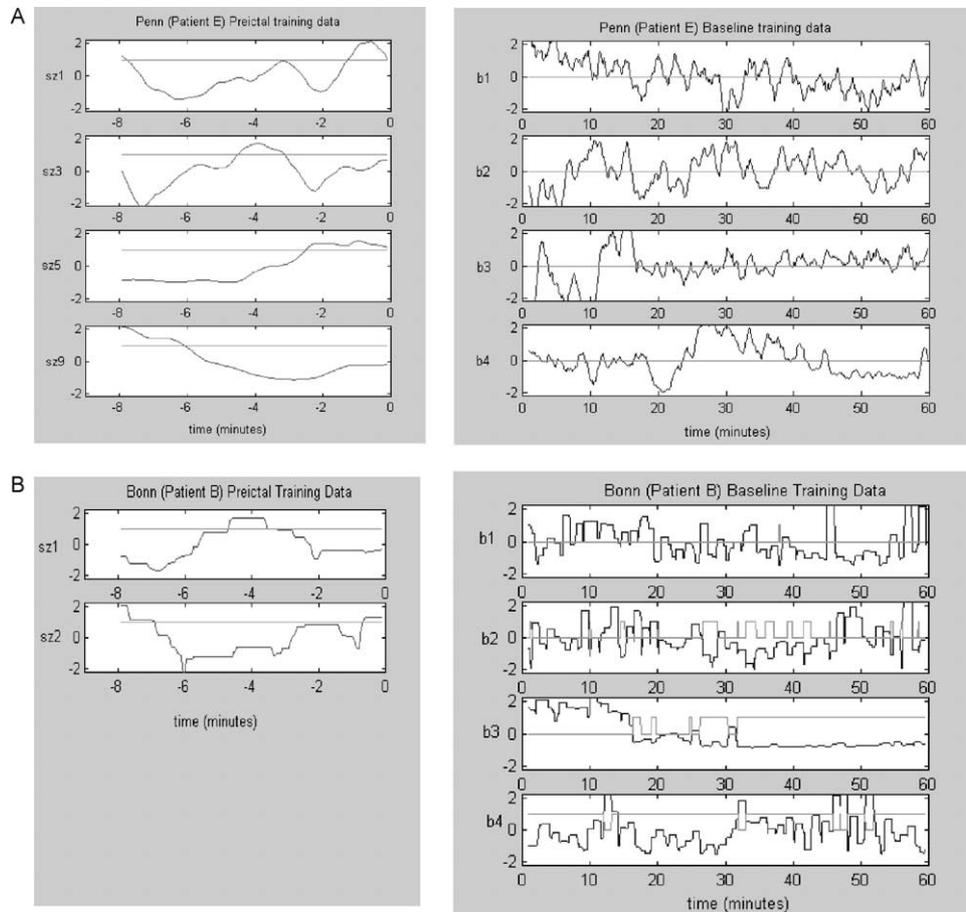


Fig. 5. Training data feature and classifier output plots. The feature outputs for the preictal training data and the baseline training data are shown. The overlaid lines with values precisely '0' or '1' represent the outputs of the classifier. A classifier output equal to '1' indicates a preictal classification, while a classifier output equal to '0' indicates a baseline classification. (a) Normalized feature (mean of the mean of the curve length for bipolar signal LTA1–LTA2) and classifier outputs for training data for the final feature selected for patient E (Penn). (b) Normalized feature (median of the maximum of the energy for bipolar signal TLR4–TLR5) and classifier outputs for training data for the final feature selected for patient B (Bonn).

delineates a preictal classification. The final feature and classifier output for the training data provided optimal performance. There were no false positives or false negatives; therefore, the block length required using the method described above is one data point for the feature, which corresponds to 2.48 s of EEG data. Using a 2.48 s block length and considering the initialization period required by the feature extraction process permitted a maximum prediction horizon of 7.93 min for this patient. Using a 2.48 s block length, the sensitivity over the test data was 100% with an average prediction time of 2.1 min and 1.1 FPh.

The feature and classifier outputs for the Patient B training data are shown in Fig. 5b. The genetically selected feature for this patient was the 'max of the median of the energy for bipolar channel TLR4–TLR5.' The final feature and classifier output for training provided optimal performance for the preictal data segments and suboptimal performance over the baseline training data. Consequently, the average length of the false positives and true positives in

the training data set were evaluated to determine a block length size to be used for testing and validation.

The feature plots and classifier outputs for the entire stay are shown in Figs. 6 and 7 for patient E and B, respectively. The electrographic seizure onsets are marked with a vertical red line. The times identified on the plots specify the prediction time using a block length of 5 min for patient B and 2.48 s for patient E. Fig. 6 provides the feature and classifier outputs for patient E using the 2.48 s block length to declare a false positive.

In Fig. 7, we observe that for Patient B, the performance of the classifier was reasonable for CDs b–f; however, the rate of incorrect classifications was extremely high for the remaining CDs. We explore possible reasons for these inaccuracies in the discussion. The minimal block length required for testing using the method described above is 3.4 min long. The initialization required to implement this block length reduces the prediction horizon from a maximum of 5.3 min to a minimum of 3.4 min. The technique applied here was selected to demonstrate

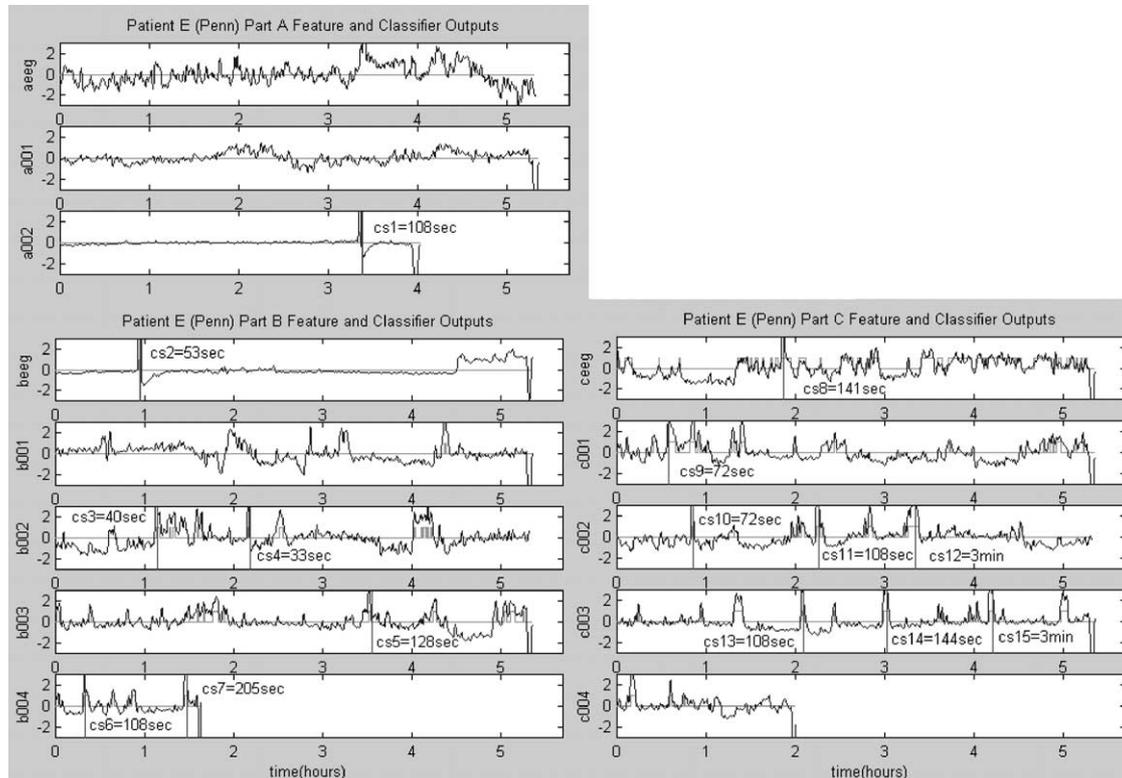


Fig. 6. Feature and classifier outputs for the feature selected for patient E (Penn). Training data included 10 min of data preceding the first four clinical seizures (CS1, CS3, CS5, and CS9) and 4 h of baseline data occurring before clinical seizure nine (CS9). The y-axis is labeled with the respective compact disc number provided for the workshop participants. The feature outputs, classifier outputs, and seizure onsets are shown. The overlaid lines with values precisely '0' or '1' represent the outputs of the classifier. The vertical lines identify the seizure onsets.

the methodology; the block length of the predictor had a profound impact on the false positive rate and the sensitivity, and should have been optimized to demonstrate statistical significance.

Fig. 8 shows the effect of block length on performance. Curves are shown for several values of false positives per hour (FPh) and sensitivity to compare the values selected for evaluation. The results shown in the figure are recorded in Table 1 where the block lengths and the attainable prediction horizon for each value are identified. Each point on the curve corresponds to a different block length and subsequent prediction horizon. The vertical axis represents the sensitivity of prediction based upon the performance of the selected feature and electrode contact and is plotted against false positives per hour (FPh). Points at the bottom left corner of the curve represent performance using block lengths for which there are no false alarms, and none of the seizures can be predicted. Positions at the top right corner represent block lengths for which all seizures are predicted, but there is a high rate of false positives per hour. The most desirable curve is a vertical line at the origin. This means the rate of false positives per hour is minimized and the rate of correct positives is maximized.

The high sensitivity achieved for patient E comes at the expense of a high level of false positives per hour (FPh). Patient B's performance may be partially attributed to

the fact that the baselines selected for training were not representative of the entire data set and may have tuned the classifier to the training set rather than a general representative data sample. The block lengths selected for prediction are shown with the arrows in Fig. 8.

4. Discussion

The objective of this study was to (1) present a possible methodology for selecting an optimal electrode contact and feature among multiple contacts and potential features for seizure prediction, and (2) to confine the prediction horizon to a very useful, well defined period of time, within 10 min of seizure onset, to facilitate incorporation in a therapeutic device. The output of the probabilistic neural network (PNN) classifier for Patient E demonstrates that a system based upon multiple features and electrode sites tailored to individual patients is capable of producing promising results for seizure prediction. That the electrode site selected as best for short-term prediction was contralateral to the focus channel may indicate the importance of brain outside of the ictal onset zone but within the 'epileptic network' in generating clinical seizures. This finding also agrees with previous experiments that demonstrate spread of seizure precursors to the contralateral temporal lobe within 20 min

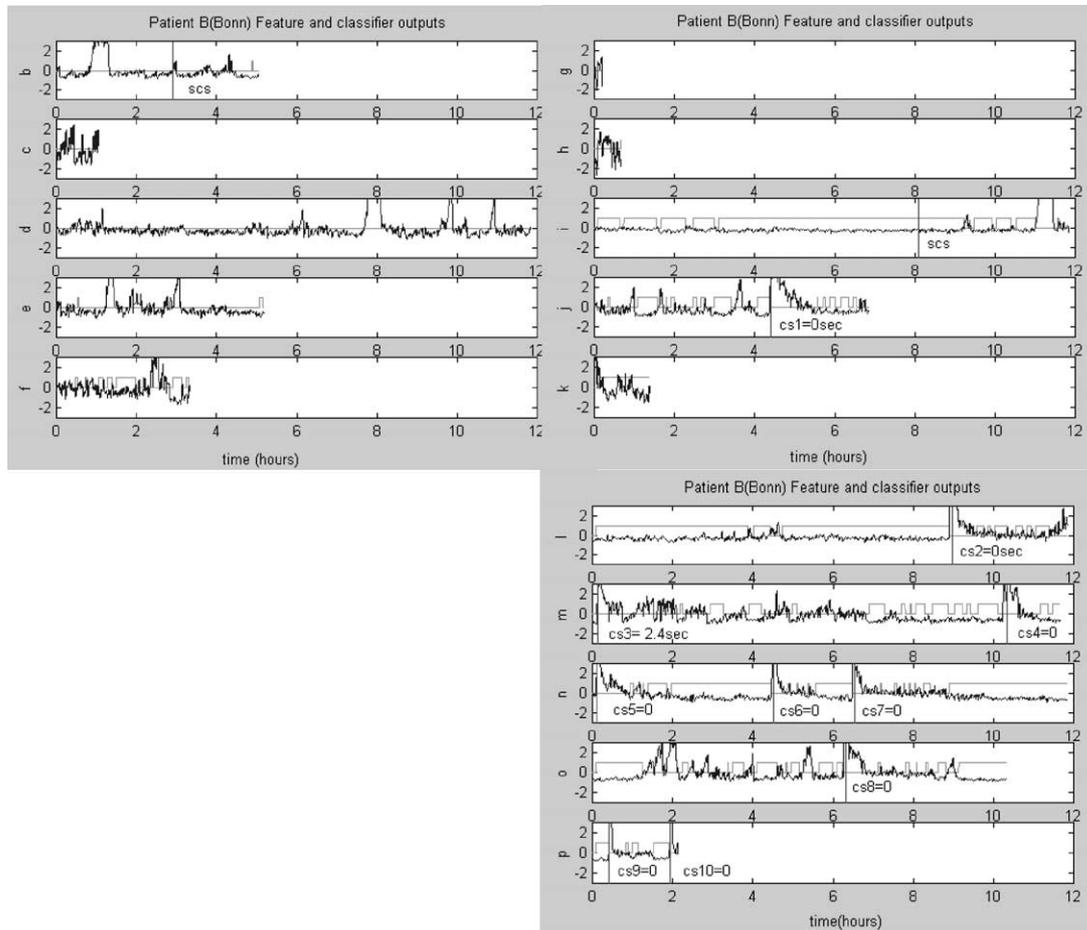


Fig. 7. Feature and classifier outputs for the feature selected for patient B (Bonn). Training data included 10 min of data preceding the first two clinical seizures (CS1 and CS2) and 4 h of baseline data occurring before clinical seizure two (CS2). Note that the plots here indicate that the preictal training segments do not predict. Since the classifier is 'on' prior to the prediction horizon, in both cases the apparent correct classification is considered a false positive. These results are not contradictory to the training outputs since these outputs are considering a 5 min block rather than 2.477 s as in the training data and our imposed requirement to toggle before the prediction horizon restricts a potential prediction.

of electrographic seizure onset. Since this method compares the change from baseline to a pre-ictal state, it is not surprising that this change may be greater in the 'normal' temporal lobe minutes prior to seizure onset than in the 'abnormal' temporal lobe in unilateral temporal lobe epilepsy.

This method requires further refinement and validation, but provides one way of dealing with the heterogeneity of seizure types and individual patterns to be addressed by prediction technology. It may also provide insight into brain mechanisms that underlie seizure generation, for example the role of regions remote from the ictal onset zone but functionally connected to the epileptic network. The performance of the method on the data from Patient E is discouraging. This may be a result of using a very limited set of electrode sites and quantitative features for prediction. It may also be limited by the relatively small number of seizures used for training, and seizure clustering, as previous work has demonstrated that clustered seizures

appear to be different from 'leading' seizures in the way that they are generated (Litt et al., 2001).

In this study, after empirically selecting the validation block length, we manually altered the block length to see its effect on sensitivity and specificity (in the form of FPH). Fig. 7 shows that accurate sensitivity is achieved at the expense of an increased FPH rate. In an implantable device, the electrode site and features might be selected during intracranial monitoring before device implantation, then further refined and retrained over time.

The poor performance on Patient B could be attributed to a number of factors. First, the limited number of leading preictal records used for training may have contributed to suboptimal performance. This problem could be addressed with iterative training over time, after device implantation. Second, since the classifier is 'on' prior to the prediction horizon, in both cases the apparent correct classification is considered a false positive. These results are not contradictory to the training outputs since these outputs consider

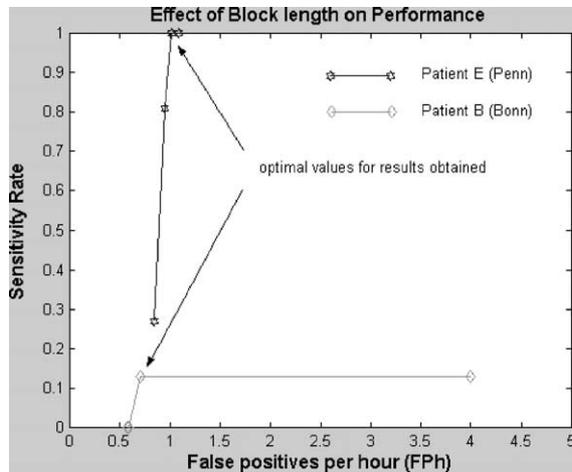


Fig. 8. Performance curves for both patients showing the effect of block length on performance. Optimal performance would be achieved if the curve rapidly approached the upper left hand corner of the graph. This means that the rate of false positives per hour is minimized and the rate of correct positives is maximized. The high sensitivity achieved for patient E, is achieved at the expense of a high level of false positives per hour (FPh). Patient B's performance may be attributed to the small number of lead seizures available for training, and the homogeneity of the baseline training data.

a 5 min block rather than 2.48 s, as in the training data, and our imposed requirement to toggle before the prediction horizon restricts potential predictions. Finally, the classifier appears to have tuned to the training records rather than a representative data sample from the training set. This could have been detected in advance and prevented by comparing baseline records to one another with one of a number of quantitative techniques (e.g. correlation) prior to analysis to ensure a diverse sample of baselines is represented in the training set.

When compared to our original methodology, the importance of sleep staging and including both states of consciousness in the training set may also contribute to the relatively poor performance in patient B. This is a topic of further study in our laboratory. It also is possible that the feature set reduction may have eliminated potentially useful features, or the subset of electrodes chosen for analysis excluded sites important to seizure generation important to

Table 1
Performance results for patients B and E for leading and clustered seizures

	Prediction block length	Sensitivity	FPh	Pred. time horizon (min)
Patient B	2.48 s	0.13	4	7.89
	5.0 min	0.13	0.71	2.89
	5.5 min	0	0.59	2.39
	6.0 min	0	0.59	1.89
	6.5 min	0	0.58	1.39
Patient E	2.48 s	1	1.09	7.89
	15 s	1	1.02	7.68
	30 s	1	0.61	7.43
	45 s	81.82	0.59	7.18
	60 s	0.27	0.58	6.94

seizure generation. Interestingly, Fig. 7 indicates that the preictal training segments do not predict seizures. Since the classifier is 'on' prior to the prediction horizon, in both cases the apparent correct classification is considered a false positive. Requiring this 'toggle' is a more stringent criterion for prediction, included purposefully so as to not bias the performance assessment in favor of positive results. Another option is to use portions of the preictal data segments removed from the prediction horizon as baselines in the training set to minimize the potential conflict demonstrated here.

The next step in this research is to expand the original methodology using all monitored IEEG channels and multiple fused features, to include an optimized method for combining electrode contacts and features to evaluate multivariate feature analysis. Other options include applying approaches such as the 'wrapper' method to feature selection, to introduce a feedback path to the genetic algorithm, thereby enabling the system to learn which features are relevant. In addition, other complementary features such as sleep staging could be incorporated in the process to track seizure generation. Training on continuous data, and eliminating selected epochs, such as baselines, is also an important step toward evolution of a reliable seizure prediction algorithm that can be implemented prospectively. Furthermore, the first step at incorporating classifier based performance metrics used in this study is being refined to incorporate generic specificity and sensitivity values that will provide statistical validation of the protocol.

In our original work, we demonstrated (1) the need for multiple features to accurately anticipate seizures in a short term prediction horizon, (2) a novel method for combining multiple quantitative features derived from the intracranial EEG and selecting the best intracranial electrode sites using a hybrid selection approach, and (3) the unexpected finding that electrodes far from but physiologically connected to the seizure onset zone may be the most useful for predicting seizures in a time horizon in the ten minutes before onset. These findings are important steps in refining our results and optimizing the selection process. The method applied at the Bonn workshop used the features identified as promising in the original work. The results presented here demonstrate that prediction is individual.

The most important lesson learned from this exercise is that a well thought out methodology, if applied appropriately may have great value in predicting seizures. To date, most research has analyzed the focus channel since it appears that the focus channel is the channel from which the seizure originates. Litt et al. suggest that the accumulated energy increases slope minutes prior to seizure onset on EEG, and that it may be useful for seizure prediction when evaluating the focus channel (Litt et al., 2001). However, when evaluating the accumulated energy, asleep and awake records must be differentiated and compared only with those records in the same awareness state. At the Bonn workshop, data analysis was performed without regard to

the awareness state of the patient, which more than likely contributed to the results presented here. This conclusion is consistent with our original methodology where signals representative of each awareness state were included in the training data set. To acquire FDA approval for an implantable device, clinical trials must yield improvement in 25% of patients. The results presented in our original analysis yield an average of 62.5% prediction in all patients analyzed. Since only two patients were evaluated for the Bonn workshop, it is difficult to hypothesize about the predictive power of the abbreviated method used.

Training off-line is not an unreasonable approach, even for implantable device applications, as long as features can be computed online, once the appropriate combination of electrode sites and predicting features is determined. In attempting to expedite the process to meet the workshop timelines, the step between the genetic algorithm selection and the block length determination for the classification step was eliminated and most likely contributed to suboptimal performance of the method. The relevant aspect of the processing time involved in this method is the time the selected feature takes to process the signal. The features selected for this evaluation are capable of being processed in real time, a vital consideration in sensor-equipped implantable devices. (Rise, 2001; Echaz and Esteller, pers. comm. 2001).

Acknowledgements

This research has been funded by The Whitaker Foundation, The Esther and Joseph Klingenstein Foundation, The Dana Foundation, The American Epilepsy Society, The CURE Foundation, The Partnership for Pediatric Epilepsy and through a grant from the National Institutes of Health, Grant #R01NS041811-01.

Appendix A

The mathematical representation of the curve length in its discrete form is:

$$CL[n] = \frac{1}{N} \sum_{i=1}^{n-D} \sqrt{(x[i+1] - x[i])^2} \quad (A1)$$

where $CL[n]$ is the running curve length of the time series $x(n)$, N is the length of the sliding observation window expressed in number of points, n is the discrete time index, and D is the overlap. The curve length is useful for observing amplitude and frequency changes and dimensionality of the signal. The accumulated energy (AE) provided promising results for seizure prediction in all the patients analyzed in (Esteller, 2000) and in (Litt et al., 2001) using the focus channel. However, unless AE is converted to a resetting AE, where it is initialized blind to the seizure

onset, it is not a practical feature for an online pattern recognition system. Consequently, energy is considered as a first-level feature and subsequent feature levels are expected to provide predictive pre-seizure indicators. To calculate the energy, let the sequence $x(n)$ be a preprocessed and fused input signal, then the instantaneous energy of $x(n)$ is given by $x^2(n)$. Considering a nonoverlapping sliding window, the energy of the signal becomes the average power over the window mathematically defined as,

$$E[n] = \frac{1}{N} \sum_{i=n-1}^{n+N-1} x^2(i) \quad (A2)$$

where N is the size of the sliding window expressed in number of points, and $n=1,2,3,\dots$

If an overlap of D points is allowed, then the average energy becomes:

$$E_D[n] = \frac{1}{N} \sum_{i=1+n-D}^{n+N-D} x^2(i) \quad (A3)$$

An overlap of 1500 points was used in this work at the first level of feature extraction while an overlap of 23 points was used at levels 2 and 3.

For the input signal $x(n)$, in its discrete form, the nonlinear (Teager) energy is:

$$NE[n] = x^2(n) - x(n-1)x(n+1) \quad (A4)$$

The nonlinear energy is an instantaneous feature, such that it provides one value for each value of original data. After the NE is obtained, the feature is weighted with a Hanning window; then the mean of the windowed data, $NE_w[n]$, is taken over the desired sliding window. After windowing, the average nonlinear energy is then:

$$ANE = \frac{1}{N} \sum_{n=1+k-1}^{k+N-D} NE_w[n] \quad (A5)$$

where $ANE[k]$ is the average nonlinear energy at time k .

The measure captures both amplitude and frequency changes, and is computationally efficient, and simple to calculate. All first level features were calculated over a 10-s window length with 2 s of displacement. Similarly second- and third-level features were limited to those providing the best results in (D'Alessandro et al., 2003). These included: *minimum*, *maximum*, *sum*, *mean*, and *median*. All were calculated over a one minute window with 2.5 s of displacement.

References

- Chang E. Using genetic algorithms to select and create features for pattern classification. IEEE international joint conference on neural networks 1990.
- D'Alessandro M. The utility of intracranial EEG feature and channel synergy for evaluating the spatial and temporal behavior of seizure

- precursors. PhD Dissertation, Georgia Institute of Technology, Atlanta 2001.
- D'Alessandro M, et al. Epileptic Seizure prediction using hybrid feature selection over multiple intracranial EEG electrode contacts: a report of four patients. *IEEE Trans Biomed Eng* 2003;50:603–15.
- Echauz J, et al. Unified probabilistic framework for predicting and detecting seizure onsets in the brain and multitherapeutic device. US Patent 6,678,548.
- Esteller R. Detection of seizure onset in epileptic patients from intracranial EEG signals. PhD Dissertation, Georgia Institute of Technology, Atlanta 2000.
- Iasemidis L, Sackellares JC. Seizure warning and prediction. US Patent No. 6,304,775. 2001.
- Iasemidis L, et al. Nonlinear dynamics of ECoG data in temporal lobe epilepsy. *Electroencephalogr Clin Neurophysiol* 1988;5:339.
- Iasemidis L, et al. Preictal entrainment of a critical cortical mass is a necessary condition for seizure occurrence. *Epilepsia* 1996;37.
- Lehnertz K, Elger C. Can epileptic seizures be predicted? Evidence from nonlinear time series analysis of brain electrical activity *Phys Rev Lett* 1998;80:5019–22.
- Lehnertz K, Litt B. The First International Collaborative Workshop on Seizure Prediction: summary and data descriptions. *Clin Neurophysiol* 2005;116:493–505.
- Lehnertz K, et al. EEG nonlinear analysis in epilepsy. *J Clin Neurophysiol* 2001;18:209–22.
- Litt B, Echauz J. Prediction of epileptic seizures. *Lancet Neurol* 2002;1:22–30.
- Litt B, et al. Epileptic seizures may begin hours in advance of clinical onset: a report of five patients. *Neuron* 2001;30:51–64.
- Petrosian A, Homan R. The analysis of EEG texture content for seizure prediction. EMBS proceedings of the 16th international conference of the IEEE 1994.
- Petrosian A, et al. Recurrent neural network based prediction of epileptic seizures in intra- and extracranial EEG. *Journal of Neurocomputing* 2000;30:201–18.
- Rise M. Commercialization of seizure prediction technology promises and pitfalls of biosignal analysis: seizure prediction and management (A case study). EMBS proceedings of the 23rd international conference of the IEEE, Istanbul, Turkey 2001.