31st Annual International Conference of the IEEE EMBS
Minneapolis, Minnesota, USA, September 2-6, 2009

Seizure Prediction Using Cost-Sensitive Support Vector Machine

Theoden Netoff, Yun Park, and Keshab Parhi

***Abstract*— Approximately 300,000 Americans suffer from epilepsy but no treatment currently exists. A device that could predict a seizure and notify the patient of the impending event or trigger an antiepileptic device would dramatically increase the quality of life for those patients. A patient-specific classification algorithm is proposed to distinguish between preictal and interictal features extracted from EEG recordings. It demonstrates that the classifier based on a Cost-Sensitive Support Vector Machine (CSVM) can distinguish preictal from interictal with a high degree of sensitivity and specificity, when applied to linear features of power spectrum in 9 different frequency bands. The proposed algorithm was applied to EEG recordings of 9 patients in the Freiburg EEG database, totaling 45 seizures and 219-hour-long interictal, and it produced sensitivity of 77.8% (35 of 45 seizures) and the zero false positive rate using 5-minute-long window of preictal via double-cross validation. This approach is advantageous, for it can help an implantable device for seizure prediction consume less power by real-time analysis based on extraction of linear features and by offline optimization, which may be computationally intensive and by real-time analysis.**

I. INTRODUCTION

Epilepsy is one of the most common neurological diseases. Approximately 5% of the population experience a seizure within their lifetime, and 1% suffer from multiple seizures, classifying them as epileptic. This disease affects nearly 3 million Americans with an estimated annual cost of $15.5 billion in direct and indirect costs per year [1]. A difficult aspect of epilepsy is the unpredictable nature of seizures. Many epileptics live in constant worry that a seizure could strike at an inopportune time resulting in humiliation, social stigma, and/or injury. Therefore, an implantable device that could predict a seizure by even a few seconds could dramatically change the lives of these patients by alerting them to the impending seizure or triggering a device to abate or suppress the seizure. As yet, however, there is no device or algorithm that provides sufficient power of prediction.

It has long been observed that there are some signals that indicate a seizure is approaching. Approximately 40% of temporal-lobe epileptics demonstrate some form of aura, a sensory perception that indicates a seizure is looming [2]. Epileptologists have also been able to see a change in EEG prior to the onset of a seizure and have labeled it the preictal quiescence or the electrodecremental period [3].

T. Netoff is with Faculty of Biomedical Engineering, University of Minnesota, 312 Church St SE, Minneapolis, MN 55455, USA tnetoff@umn.edu

Y. Park is with the Department of Electrical Engineering, University of Minnesota, Minneapolis, MN 55455, USA parkx296@umn.edu

K. Parhi is with Faculty of Electrical Engineering, University of Minnesota, Minneapolis, MN 55455, USA parhi@umn.edu

However, creating a reliable algorithm for seizure prediction has been elusive. Seizure prediction algorithms using power spectrum [4][5] and cross-correlation between EEG recordings [6][7] have seemed to perform successfully on short windows of EEG data taken immediately before a seizure. But, when compared to EEG recorded for long periods of time, capturing all the changes in the above features measured over the day as a patient undergoes in and out of different conscious states, those algorithms produced a significant number of false positives.

The reason that seizure prediction with high sensitivity and specificity has been difficult to achieve may be the approach itself; thus, we choose an alternative classification approach. Most of the approaches to seizure prediction have been hypothesis-based, where scientists select a certain feature that they believe changes prior to a seizure. However, no feature has been found to be unique to the seizure onset, so these approaches have resulted in many false positives. The alternative approach we used is based on machine learning methodology, where we supply an algorithm with enough data in which EEG is identified as preictal (immediately preceding a seizure) or interictal (ordinary between seizures), as shown in Fig. 1, and let a computer optimize the algorithm to classify those two data sets. This approach is powerful, for the combination of certain sets of features can be examined and complex relationships among the features for finding the seizure onset, which probably could not be found by a human, can be investigated.

Of the available classifiers, we have chosen the support vector machine (SVM) [8], which is currently the most powerful. Due to its robustness for estimating predictive models from noisy, sparse and high-dimensional data, SVM methodology has been successful in many applications rang-

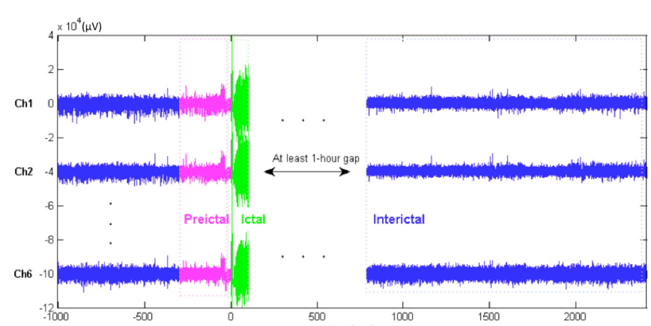


Fig. 1. Seizure recorded with EEG. Ictal (seizure) event shown in green, is immediately preceded by a window we define as "preictal." Windows prior to or an hour following a seizure are considered "interictal."

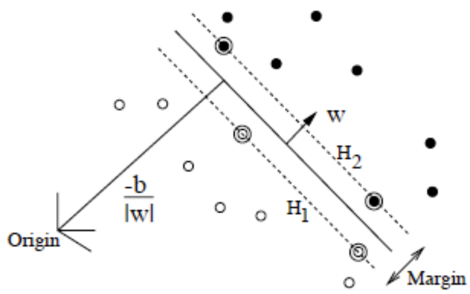


Fig. 2. Support vector machine. Points from two classes are divided by a plane (indicated as line here) in a high dimensional space. Vectors parallel to the dividing line, the support vectors, are used to minimize the error for optimal division of space. Figure from http://research.microsoft.com/en-us/um/people/cburges/papers/svmtutorial.pdf

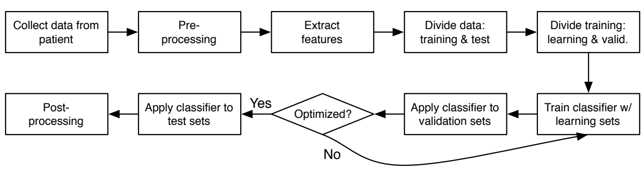


Fig. 3. Flow chart of seizure prediction algorithm

ing from genomics to financial data analysis and signal processing, in addition to seizure detection [9]. SVM finds a mapping of the data into a new space where a linear hyperplane is used to separate the two classes. This is graphically illustrated in Fig. 2.

The cost-sensitive SVM (CSVM) is more suitable for seizure prediction than conventional SVMs, for CSVM can weigh important classes more. It is critical in seizure prediction to treat signals from preictal more significantly than those from interictal, for misclassification of the former is much more penalized than that of the latter and moreover the former is much sparser. By putting more weight on preictal, CSVM can outperform conventional SVMs as a seizure predictor.

II. METHODS

A. *Outline and Patient Database*

The seizure prediction algorithm consists of preprocessing, feature extraction, preparation of data for classification, classification, and postprocessing, as outlined in Fig. 3. Each step will be discussed in its sub-section in detail.

We have tested our algorithm on the Freiburg dataset (https://epilepsy.uni-freiburg.de/freiburg-seizure-prediction-project/eeg-database). This database contains ECoG (intracranial EEG) recordings from 21 patients that suffer from medically intractable focal epilepsy. We selected 9 out of 21 patients, choosing those that had five seizures recorded at least, for performing double-cross validation smoothly. For each patient, six EEG electrodes, 3 at the seizure focus and 3 distal to the focus, were provided in the database, sampled at 256 samples per second. Seizure onsets were identified by an epileptologist, and onset times are provided in the database. Preictal EEG signals are presumed to be recorded in five minutes immediately before each of seizure occurrences. Interictal data for each patient totaled 24 to 26 hours, excluding EEG recordings at least one hour before or after a seizure. A total of 45 seizures and 219-hour-long interictal data from 9 patients in the Freiburg database were examined.

B. *Preprocessing*

EEG data is subject to many artifacts, such as line noise, electrical noise and movement artifacts. Many of these artifacts may produce signals that can distort normal EEG data, therefore causing outliers in much of the analysis and finally leading to faulty analysis. For preliminary analysis, we discarded windows which contain artifacts identified by visual inspection. We also removed line noise with notch filters and detrended each window before analysis.

C. *Feature Extraction*

Power in the following 9 different spectral bands from each of the 6 electrodes was extracted in each preictal or interictal window, which is 20-second-long and half overlapped with the prior: delta (0.5-4Hz), theta (4-8Hz), alpha(8-13Hz), beta(13-30Hz), four gamma(30-50Hz, 50-70Hz, 70-90Hz, 90Hz-), and total power, as shown in Fig. 4. This builds 54 dimensions (6 electrodes and power in 9 bands per electrode) in the input space of each window.

D. *Preparation of Data for Classification*

Double cross-validation was used: a portion of data are used for training the classifier and the other portions are left out for testing the classification algorithm. In double cross-validation, the data is divided into two groups, a training set and a test set, and the training set is then subdivided into a learning set and a validation set. The validation set is used for testing how well the classifier performs after being trained on the learning set, including testing for over-fitting of the algorithm. Once the classifier is fully optimized with the learning and validation sets via v-fold cross-validation, it is applied to the test set to assess final performance.

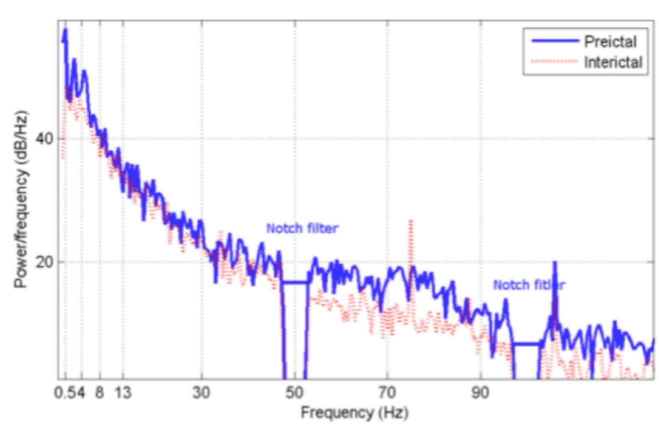
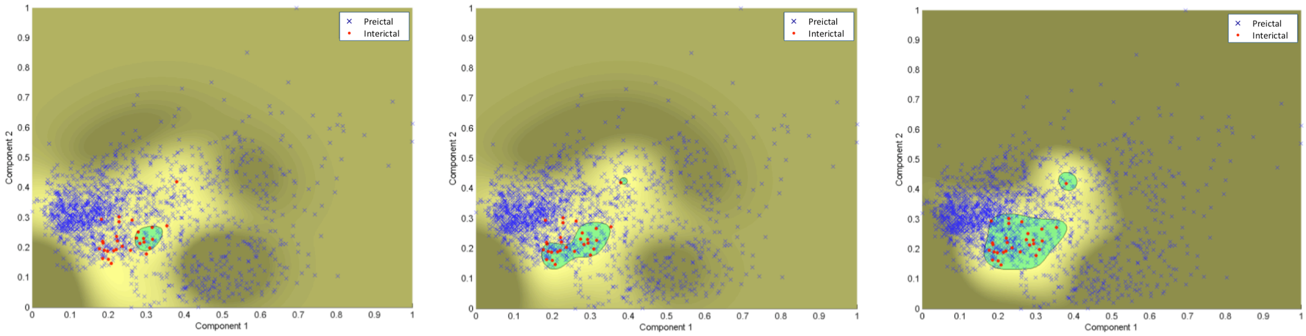


Fig. 4. Power spectrum of preictal and interictal windows broken into 9 frequency bands for comparison



(a) Weighting = 5 (b) Weighting = 10 (c) Weighting = 50

Fig. 5. CSVM classification as a function of weighting: (a) Top two principal components of the 54 features from EEG are plotted for each window of data. Blue X's represent interictal windows; red dots do preictal windows. Are that is classified above a certain threshold is colored by green. At 5 times relative weighting of preictal over interictal, a small region of positives are selected. Most preictal windows are missed and only a few false positives are detected. (b) As relative weighting increases to 10, the green region above threshold expands capturing more preictal points but less interictal ones, leading to higher sensitivity but lower specificity. (c) As weighting increases even greater, most of the preictal are classified but the problem of over-fitting may occur.

E. Classification

CSVM in the package of LIBSVM [10] was used. The SVM was optimized for each patient using the following parameters:

1) The misclassification cost
2) The relative weights of interictal to preictal windows. Effects of weighting are illustrated in Fig. 5.
3) Polynomial kernel order 2, 3, and 4

Five-fold cross validation was performed with the training set. Each classification model is built with the learning set to minimize the following cost function:

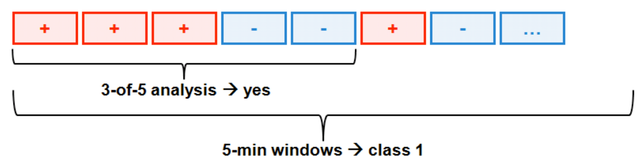
$$\frac{1}{2}\|\omega\|^2 + C^+ \sum_{i\in+class} \xi_i + C^- \sum_{j\in-class} \xi_j$$

where $\frac{1}{\|\omega\|^2}$ is the margin, distance between the support vectors and the decision boundary, ξ_i and ξ_j are slack variables, and C^+ and C^- are misclassification costs for false negatives and false positives, respectively. Then, once optimized through the above process, the classifier was applied with the test set, generating (predicting) the label for the unknown dataset.

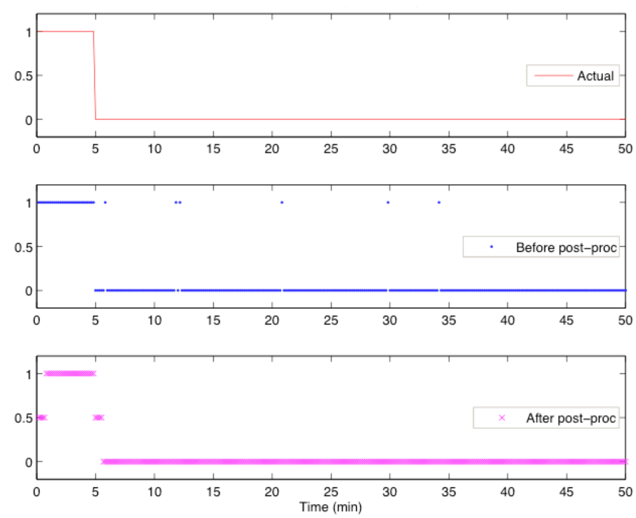
Points may be misclassified for several reasons. One cause may be that the results are noisy and therefore the classification will be wrong for some points. Another reason may be that the EEG is truly preictal but the seizure was somehow avoided. Our operational definition of preictal is that it immediately precedes the seizure; this neglects other windows of similar behavior that may be preictal. Therefore, it is required to separate windows that may have been spuriously misclassified from windows that show true preictal behaviors.

F. Postprocessing

For postprocessing, k-of-n analysis was performed as follows: each window is predicted as positive or negative, with respect to predicting an impending seizure. If there are equal to or more than k positives out of n consecutive windows, then the prediction horizon following the event is considered preictal; otherwise, it is labeled as interictal (see Fig. 6). As postprocessing in the proposed algorithm, 3-of-5 analysis was used to identify the prediction horizon of five minutes prior to a seizure. The reason to choose a five-minute



(a) 3-of-5 analysis



(b) Effect of 3-of-5 analysis

Fig. 6. Postprocessed data: (a) Each window is classified as positive or negative. A 3-of-5 analysis requires at least 3 windows of 5 to be positive to classify the entire next 5 minutes of data as positive. (b) 3-of-5 analysis applied to data results. Five minutes of preictal windows are plotted fooled by approximately 45 minutes of interictal. A few interictal datasets are misclassified, however they are sparse and do not reach 3-of-5 significance cutoff. After post-processing, no false positives are detected and preictal windows are correctly labeled due to the much higher density of positives.

prediction horizon is that it is long enough for patients to prepare for impending seizures but is still specific enough in time to be useful for drug administration or depth-electrode stimulation.

III. RESULTS

Sensitivity ($\frac{TP}{TP+FN}$) and the rate of false positives per hour have been measured to estimate how successfully the proposed algorithm works. Sensitivity measures how many of all the seizure occurrences a prediction model can recognize as such, and false positive rate indicates how many false alarms occur in an hour. From the EEG dataset of 9 patients with 45 seizures and 219-hour-long interictal signals, a conventional SVM classified 56% of the preictal correctly, using the features of spectral power in 9 bands. By optimizing the weights by increasing the preictal data's importance to 32 to 64 times that of interictal and by postprocessing of 3-of-5 analysis, the proposed algorithm predicted 35 of 45 seizures, resulting in sensitivity rate of 77.8% (ranging from 2 of 5 seizures in 1 patient and 4 of 5 or better in 6 patients) with no false positives, as shown in Table I.

IV. DISCUSSION

High sensitivity of nearly 80% and no false positive suggest that spectral power in several bands, which are linear features, may be one of the correct features for seizure prediction and that CSVM and *k*-of-*n* analysis, which work in the non-linear fashion, may perform as an effective classifier for seizures. From the perspective of building an implantable device, the above fact is attractive, for it demonstrates that prediction can be performed with linear features by non-linear classifiers. Once the part of the non-linear classifier in the implantable device is optimized, whose process may be computationally intensive but can be achieved offline, the implantable device can be designed to consume less power, for it operates based on linear features.

Several other studies for developing seizure prediction methods have been tested on these exact same Freiburg EEG datasets [11], but our results demonstrate a significant improvement over previous reports for the following reasons. First, we have achieved significant sensitivity and specificity with just a 5-minute-long prediction horizon. In general, algorithms that use short prediction horizons usually produce low sensitivity less than 60% [12] or high false-positive rates around 1 per hour [13]. Other algorithms that yield somehow comparable sensitivity use windows of 30 minutes or longer [14][15][16], but they require significantly more computation time and result in much longer, and less temporally specific, prediction windows. Second, we used the method of double cross-validation in order to test our algorithm which is optimized with the training set only. Most of the previous studies with high sensitivity and specificity were not evaluated via double cross-validation [11]: they were simply trained and tested their data on the exact same dataset. Some of the previous works might represent better results than ours, but should be treated differently from ours, for prediction requires processing data that has not been recorded yet.

TABLE I
OUTCOMES FROM SEIZURE PREDICTION ANALYSIS

Patient No.	Sensitivity (%)	FP/Hour	Poly deg.
3	80.0	0.0	2,3,4
4	80.0	0.0	2,3,4
5	60.0	0.0	3,4
10	100.0	0.0	2,3,4
16	60.0	0.0	2,3,4
17	100.0	0.0	2,4
18	100.0	0.0	2,3,4
20	40.0	0.0	2,3
21	80.0	0.0	2,3,4
Average	77.8	0.0	

V. CONCLUSION

A patient-specific algorithm for seizure prediction based on EEG recordings is proposed. This algorithm extracts spectral power in nine bands and classifies preictal and interictal using CSVM classification and *k*-of-*n* analysis. Applied to 9 patients' recordings from the Freiburg EEG database via double-cross validation, the proposed algorithm resulted in the average sensitivity of approximately 80% with zero false positive.

REFERENCES

[1] http://www.epilepsyfoundation.org/about/statistics.cfm

[2] E. Wyllie, A. Gupta, and D.K. Lachhwani, *The treatment of epilepsy : principles & practice*, Lippincott Williams & Wilkins, PA; 2006.

[3] H.H. Jasper, Some Physiological Mechanisms Involved in Epileptic Automatisms, *Epilepsia*, vol. 5, 1964, pp 1-20.

[4] G. Alarcon, C.D. Binnie, R.D. Elwes, and C.E. Polkey, Power spectrum and intracranial EEG patterns at seizure onset in partial epilepsy. *Electroencephalogr. Clin. Neurophysiol.*, vol. 94, 1995, pp 326-337.

[5] I. Osorio, M.G. Frei, and S.B. Wilkinson, Real-time automated detection and quantitative analysis of seizures and short-term prediction of clinical onset. *Epilepsia*, vol. 39, pp 1998, pp 615-627.

[6] F. Mormann, et al., Epileptic seizures are preceded by a decrease in synchronization. *Epilepsy Res.*, vol. 53, 2003, pp. 173-185.

[7] K. Schindler, H. Leung, C.E. Elger, and K. Lehnertz, Assessing seizure dynamics by analysing the correlation structure of multichannel intracranial EEG. *Brain*, vol. 130, 2007, pp 65-77.

[8] V.N. Vapnik, *The nature of statistical learning theory*, Springer, New York; 1995.

[9] A. Shoeb, et al., Patient-specific seizure onset detection. *Epilepsy and Behavior*, vol. 5, 2004, pp 483-498.

[10] http://www.csie.ntu.edu.tw/∼cjlin/libsvm/

[11] F. Mormann, et al., Seizure prediction: the long and winding road. *Brain*, vol.130, 2007, pp 314-333.

[12] M. D'Alessandro, et al., Epileptic seizure prediction using hybrid feature selection over multiple intracranial EEG electrode contacts: a report of four patients. *IEEE Trans. Biomed. Eng.*, vol. 50, 2003, pp 603-615.

[13] M. D'Alessandro, et al., A multi-feature and multi-channel univariate selection process for seizure prediction. *Clin. Neurophysiol.*, vol. 116, 2005, pp 506-516.

[14] L.D. Iasemidis, Epileptic seizure prediction and control. *IEEE Trans. Biomed. Eng.*, vol. 50, 2003, pp 549-558.

[15] L.D. Iasemidis, et al., Long-term prospective on-line real-time seizure prediction. *Clin. Neurophysiol.*, vol. 116, 2005, pp 532-544.

[16] W. Chaovalitwongse, et al., Performance of a seizure warning algorithm based on the dynamics of intracranial EEG. *Epilepsy Res.*, vol. 64, 2005, pp 93-113.