Low-Complexity Seizure Prediction From iEEG/sEEG Using Spectral Power and Ratios of Spectral Power

Zisheng Zhang, *Student Member, IEEE*, and Keshab K. Parhi, *Fellow, IEEE*

***Abstract*—Prediction of seizures is a difficult problem as the EEG patterns are not wide-sense stationary and change from seizure to seizure, electrode to electrode, and from patient to patient. This paper presents a novel patient-specific algorithm for prediction of seizures in epileptic patients from either one or two single-channel or bipolar channel intra-cranial or scalp electroencephalogram (EEG) recordings with low hardware complexity. Spectral power features are extracted and their ratios are computed. For each channel, a total of 44 features including 8 absolute spectral powers, 8 relative spectral powers and 28 spectral power ratios are extracted every two seconds using a 4-second window with a 50% overlap. These features are then ranked and selected in a patient-specific manner using a *two-step* feature selection. Selected features are further processed by a second-order Kalman filter and then input to a linear support vector machine (SVM) classifier. The algorithm is tested on the intra-cranial EEG (iEEG) from the Freiburg database and scalp EEG (sEEG) from the MIT Physionet database. The Freiburg database contains 80 seizures among 18 patients in 427 hours of recordings. The MIT EEG database contains 78 seizures from 17 children in 647 hours of recordings. It is shown that the proposed algorithm can achieve a sensitivity of 100% and an average false positive rate (FPR) of 0.0324 per hour for the iEEG (Freiburg) database and a sensitivity of 98.68% and an average FPR of 0.0465 per hour for the sEEG (MIT) database. These results are obtained with leave-one-out cross-validation where the seizure being tested is always left out from the training set. The proposed algorithm also has a low complexity as the spectral powers can be computed using FFT. The area and power consumption of the proposed linear SVM are 2 to 3 orders of magnitude less than a radial basis function kernel SVM (RBF-SVM) classifier. Furthermore, the total energy consumption of a system using linear SVM is reduced by 8% to 23% compared to system using RBF-SVM.**

***Index Terms*—Branch and bound, linear separability, low-complexity architecture, power spectral density, ratio of spectral power, seizure prediction, two-step feature selection.**

I. Introduction

A. Epilepsy Background

APPROXIMATELY 1.0% of the world's population suffers from epileptic seizures. About 50 million people worldwide have epilepsy, and nearly 80% of the epileptic patients live in developing countries [1]. Epilepsy is the second most common neurological disorder [2]. A seizure is commonly defined as an abnormal, excessive or hypersynchronous neuronal activity in the brain [3]. Reliable seizure prediction, which refers to anticipating epileptic seizures based on continuous electroencephalogram (EEG) recordings of epileptic patients, is important for improving the lives of epileptic patients by alerting them to the potentially impending seizures. A device that can predict seizures can be used to deliver an anti-epileptic drug (AED) or stimulate the brain before seizures strike. Such a device can also significantly reduce patients' constant worries that a seizure may strike at an improper time resulting in embarrassment, harassment, injury, or even death [4].

Manuscript received May 30, 2015; accepted August 03, 2015. Date of publication October 26, 2015; date of current version March 04, 2016. This paper was recommended by Associate Editor R. Butera.

The authors are with the Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis, MN 55455 USA (e-mail: zhan1116@umn.edu; parhi@umn.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TBCAS.2015.2477264

B. Seizure Prediction

Seizure prediction can be viewed as a binary classification problem where one class consists of preictal signals corresponding to the signal right before an occurrence of the seizure, and the other class consists of normal EEG signals, also referred as interictal signals. Identifying features that can differentiate or discriminate the preictal state (time period before a seizure) from the interictal state (time period between seizures) is the key to seizure prediction.

Although the mechanism of the sudden occurrence of a seizure still remains unclear, it is known that the patterns do vary during preictal and interictal periods in most of the cases [5]–[9]. Significant amount of research in seizure prediction has been directed towards identifying these discriminating patterns or features. Examples of these features include power spectral density [10], [11], [12], autoregressive coefficients [13], power of the wavelet coefficients [14], mean phase coherence [15], statistical features [16], instantaneous amplitude, frequency, or phase [17] of the EEG signal. Past research has demonstrated the feasibility of predicting seizures from scalp EEG (sEEG) or intracranial EEG (iEEG) signals [18]–[22]. Recent research efforts have focused on developing a real-time automated seizure prediction system that can predict seizures for patients over long periods of time [23]–[26]. It is known that the power spectral density (PSD) of the EEG signal is altered before and during seizures [10], [23].

PSD features have been used to design programmable devices that detect seizure activity and deliver a responsive electrical stimulation in an attempt to disrupt the seizure activity. The pivotal trial by Neuropace showed a 37.9% median seizure reduction compared with baseline in the blinded phase, and a 53% median seizure reduction at the end of the 2-year open label phase

[27]. While the Neuropace device can benefit many epileptic patients, more efficient prediction algorithms that can lead to a median seizure reduction in 80% to 90% range prior to seizure onsets can greatly benefit significantly larger number of epileptic patients. However, the main drawback of using spectral powers is the high false positive ratio (FPR) as the PSD increases often in the interictal periods as well. Furthermore, preictal and interictal patterns vary substantially over different patients. Even for a single patient, preictal and interictal patterns may vary substantially from seizure to seizure and from hour to hour. Past seizure prediction algorithms suffer most from the high false positive rate or low sensitivity. These algorithms suffer from several other drawbacks. For example, some algorithms are designed without cross-validation, i.e., the datasets are not split randomly into disjoint training and testing groups and simply estimating the model based on the whole data. Such algorithms are "overtrained", and may not be able to predict future seizures. Other algorithms are validated using few patients, and are not tested on large datasets containing many patients.

Most recently, various studies have achieved a relatively high sensitivity ($>$90%) and a low false positive rate ($<$0.25 FP/hour) compared with the previous efforts [13], [17], [28]–[31]. However, from the perspective of hardware implementation, these algorithms suffer from several drawbacks. Some algorithms suffer from high dimensionality of the features. Other algorithms use computationally intensive features that require increased area and power when implemented in hardware. Other algorithms use nonlinear classifiers such as radial basis function kernel support vector machine (RBF-SVM) whose area and power consumption can be significantly higher than other classifiers as these are not only dependent on feature dimensions, but also on the number of support vectors.

C. Significance

The key contribution of this paper is that this paper develops a patient-specific algorithm that can reliably predict seizures using either one or two electrodes. The proposed algorithm achieves an overall sensitivity higher than 90% and a false positive (FP) rate less than 0.125 FP/hour. The algorithm also requires a low hardware complexity for extracting features and classification.

Features such as absolute spectral powers, relative spectral powers, and spectral power ratios have been explored in [10], [23][32], respectively. These prior works show that such features can detect or predict seizure activities. However, these studies suffer form several drawbacks. For example, false positive rate (FPR) is as high as 0.27 FP/hour in [10]. Sensitivity is as low as 75.8% for the method in [23]. The predictability of the spectral power ratios is not shown in [32]. In [32], hundreds of features including relative spectral powers and all possible ratios of spectral powers are computed and are used for seizure detection. However, the proposed method only achieved a sensitivity of 89%. This paper shows that combining the PSD features such as absolute spectral powers, relative spectral powers and spectral power ratios as a feature set and then carefully selecting a small number of these features from one or two electrodes according to the linear separability criteria can achieve a good prediction performance of the subsequent classifier. This paper also shows that as more features are selected, preictal feature vectors and ictal feature vectors become more linearly separable. Therefore, a linear classifier can be used to separate preictal features from interictal features. Since all these features can be extracted by performing the fast Fourier transform (FFT) on the signals from one or two electrodes and the classifier is linear, the proposed algorithm can be implemented in hardware with low complexity and low power consumption.

In low-power and low-complexity hardware design, the *first* key consideration is the number of sensors used to collect EEG signals. Electrode selection is an essential step before feature selection as sensors and analog-to-digital converters (A/D) can be highly power consuming for an implantable or wearable biomedical device. The *second* key consideration is selecting useful features that are computationally simple and are indicative of upcoming seizure activities. The *third* key consideration is the choice of classifier. Based on the selection of the classifier, a criteria for electrode and feature selection should be chosen accordingly in order to achieve the best classification performance. It is shown in [33] that linear classifiers have significantly lower power consumptions than the nonlinear ones and are dependent on the feature dimensions only. Therefore, only linear classifiers are considered in this paper. Thus, instead of selecting electrodes by their locations, which has been used in other studies, this paper selects electrodes and features in a way such that the preictal features are as linearly separable from the interictal features as possible.

In the proposed approach, we first compute the spectrogram of the input EEG signals from one or two electrodes. A window based PSD computation is used with a 4-second sliding window with half overlap. Thus, the effective window period is 2 seconds. Spectral powers and spectral ratios are extracted as features and are input to a classifier. A postprocessing step is used to remove undesired fluctuations of the decision output of the classifier. The feature signals are then subjected to feature selection and classification where two strategies are used. One is the single feature selection and the other is the multi-dimensional feature selection. While a seizure prediction system using a single feature requires low hardware complexity and power consumption, systems using multi-dimensional features achieve a higher prediction reliability. In this paper, multi-dimensional features are selected for patients where systems using a single feature can not achieve a predetermined requirement.

This paper makes three contributions. First, the feature set includes spectral power in different bands, *relative* spectral power in different bands, and *ratio* of spectral power in different bands. These three types of features as a feature set have not been used for seizure prediction in past literature. Second, use of *two-step* feature selection is introduced, where a feature basis is first selected by the *scatter matrix* method, and a subset of these features are selected in the second step by *branch and bound* method. An electrode selection method based on the scatter matrix method selects either one or two electrodes. Third, linearly separable features are input to a linear support vector machine (SVM) classifier; the energy consumption of the classifier is three orders of magnitude less than that of a radial-basis function SVM (RBF-SVM).

The remainder of this paper is organized as follows. Section II describes the proposed algorithm for seizure prediction and the database used to validate the algorithm. Section III evaluates the performance of the proposed algorithm. Section IV describes the system architecture for the proposed algorithm and estimates of energy consumption of the system architecture. Section V presents a discussion on the results, describes prior work, and provides comparisons with prior work. Section VI concludes the paper.

II. Materials and Methods

A. EEG Databases

We have trained and tested our algorithm on the two databases: Freiburg intracranial EEG (iEEG) database [34] and MIT Physionet scalp EEG (sEEG) database [35].

According to [34], the Freiburg EEG database contains electrocorticogram (ECoG) or iEEG from 21 patients with medically intractable focal epilepsy. The Freiburg database contains signals from six electrodes, three near the seizure focus (focal) and the other three distal to the focus (afocal). Seizure onset times and artifacts were identified by certified epileptologists. The data were collected at 256 Hz sampling frequency ($f_s = 256$ Hz) with 16 bit analog-to-digital converters, except Patient No. 12 whose data was sampled at 512 Hz but was downsampled to 256 Hz.

According to [35], the MIT Physionet EEG database, collected at the Children's Hospital Boston, consists of EEG recordings from pediatric subjects with intractable seizures. The International 10–20 system of EEG electrode positions and nomenclature were used for these recordings. Recordings are grouped into 23 cases. Each case contains between 9 and 42 hours' continuous recordings from a single subject. In order to protect the privacy of the subjects, all protected health information (PHI) in the original files have been replaced with bipolar signals (one channel minus another). All signals were sampled at 256 samples per second with a 16-bit resolution. Most files contain 23 bipolar-channel EEG signals.

For both databases, patients who have less than three seizures are not analyzed in this paper. The reason for not including these patients is that training using preictal data from only one seizure is likely to lead to a model overfitting to that particular seizure and may not be able to predict the other ones. Therefore, at least two seizures must be selected in the training set and another seizure is used for testing.

For both databases, we use the following categorization: 60 minutes' recordings preceding seizure onsets are categorized as preictal (C1); 3 minutes' and 30 minutes' recordings postceding seizure onsets are categorized as ictal (C2) and post-ictal (C3), respectively; the rest of the recordings are categorized as interictal (C0). The goal of seizure prediction is to separate C1 from C0, regardless of C2 and C3.

B. Feature Extraction

This section describes the method for feature extraction, feature selection and postprocessing, which include spectral power computation, spectral power ratio computation and Kalman filter.

1) Window-Based Signal Processing: In window-based signal processing, the input signal, $s(n)$, is divided into the input segments and the signal is processed segment by segment. Let M denote the length of each segment and L denote the total number of segments. Let

$$s_l(n) = s\left(n + (l-1)M/2\right) \quad n = 0, \ldots, M-1,\ l = 1, \ldots, L$$

denote the windowed signal in the l-th segment. Each segment has a 50% overlap with its neighbour segment. The main advantage of the window-based signal processing is that the number of observation points computed is reduced by a factor of $M/2$ as compared with sample-by-sample processing. This reduces the computation complexity of the classifier. Another advantage lies in the real-time implementation of the seizure prediction system with low latency. In this paper, the window size is chosen as four seconds ($M = 4 * f_s$) and each segment is categorized as interictal (C0), preictal (C1), ictal (C2), or post-ictal (C3) according to the criteria described in Section II-A.

2) Spectral Power and Spectral Power Ratios: Three types of features are extracted from the windowed signal. These include absolute spectral power, relative spectral power and spectral power ratio.

a) Absolute spectral power: Absolute spectral power in a particular frequency band represents the power of a signal in that frequency band. The rhythmic activity in an EEG signal is typically described in terms of the standard frequency bands, but the γ band is further split into 5 sub-bands. The bands considered include: (1) θ (4–8 Hz), (2) α (8–13 Hz), (3) β (13–30 Hz), (4) $\gamma 1$ (30–50 Hz), (5) $\gamma 2$ (50–70 Hz), (6) $\gamma 3$ (70–90 Hz), (7) $\gamma 4$ (90–110 Hz), (8) $\gamma 5$ (110–128 Hz). For Freiburg database, to eliminate power line hums at 50 Hz and its harmonics, spectral powers in the band of 47–53 Hz and 97–103 Hz are excluded in spectral power computation. For MIT database, spectral powers in the band of 57–63 Hz and 117–123 Hz are excluded.

To compute the (absolute) spectral powers in the above eight frequency bands, PSD of the input signal needs to be estimated. The PSD of a signal $s(n)$ describes the distribution of the signal's total average power over frequency. In this paper, the spectral power of a signal in a frequency band is computed as the logarithm of the sum of the PSD coefficients within that frequency band. Mathematically, the spectral power in the i-th frequency band is computed as

$$P_i = \log \sum_{\omega \in\ band\ i} PSD_s(\omega),\ i = 1, 2, \ldots, 8.$$

For window-based signal processing, spectral power needs to be computed for each windowed segment $s_l(n)$

$$P_i(l) = \log \sum_{\omega \in\ band\ i} PSD_{s_l}(\omega),\ i = 1, 2, \ldots, 8.$$

Therefore, $P_i(l)$ is a time series whose l-th element represents the spectral power of the input signal in the l-th segment in band i.

b) Relative spectral power: The relative spectral power measures the ratio of the total power in the i-th band to the total power of the signal in logarithm scale, which is computed as follows:

$$Q_i(l) = \log \frac{\sum_{\omega \in \text{ band } i} PSD_{s_l}(\omega)}{\sum_{\text{all } \omega} PSD_{s_l}(\omega)},\ i = 1, 2, \ldots, 8.$$

c) Spectral power ratio: Let $R_{i,j}(l) = P_i(l) - P_j(l)$ represent the spectral power ratio of the spectral power in band i over that in band j in the l-th window. These ratios indicate the change of power distribution in frequency domain from interictal to preictal periods, which have been shown in [32] to be good features for seizure detection and in seizure prediction [36]. For a single channel EEG signal, all possible combinations of eight spectral powers lead to a total number of $\binom{8}{2} = 28$ possible ratios.

In summary, for each electrode, 44 features which include 8 absolute spectral power, 8 relative spectral powers and 28 spectral power ratios are extracted every 2 seconds.

The key advantage of spectral power ratio features over the spectral power features is that certain ratio features are strong indicators of an upcoming seizure activity while the latter are not indicative of such activity at all as the spectral power usually fluctuates a lot during both interictal and preictal periods. The ratio feature *amplifies* the simultaneous increase in the spectral power of one band and decrease in that of another band. For instance, Fig. 1 illustrates the spectral power in $\gamma 2$ band (top panel), the spectral power in $\gamma 1$ band (middle panel) and the spectral power ratio of $\gamma 2$-to-$\gamma 1$ after postprcossing using the EEG recordings in electrode No. 1 of Patient No. 19 in the MIT Physionet database, where the red vertical lines represent the seizure onsets. While the spectral power features in both bands are indiscriminate of the preictal and interictal periods, the ratio between them shows strong predictability of the upcoming seizure activities as this ratio always increases significantly prior to the seizure onsets.

3) Postprocessing: The noise of a process, which degrades the prediction capabilities, can be reduced by smoothing its irregular effects. Kalman filter was shown in [10] to be very effective in smoothing undesired fluctuations. The Kalman filter is a statistical method that can estimate the state of a linear system by means of minimizing the variance of the estimation error, so the estimates tend to be close to the true values of measurements.

In order to apply the Kalman filter to remove the noise from a signal, the process must be described as a linear system. This paper uses the same state-space model as the model described in [13] and in supplementary document of [10]. Detailed algorithm for a second-order Kalman filter is described in [37]. As a result, Kalman fitler generates a much smoother output feature.

C. Single Feature Selection and Classification

Flow chart of a single feature selection is shown is Fig. 2, where $f(l)$ represents the l-th feature sample. The feature basis selection step is followed by electrode selection. The best electrode is selected using scatter matrix method. A second round of feature selection is performed to further reduce the number of features. The linear separability criteria J is computed for all features from all electrodes and the best feature is selected whose J is the maximum. Its corresponding electrode is then used for seizure prediction.

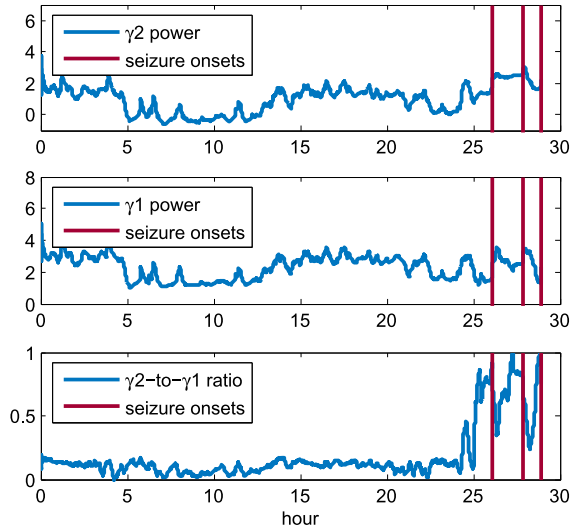

Fig. 1. Spectral power in $\gamma 2$ band (top panel), spectral power in $\gamma 1$ band (middle panel) and the spectral power ratio of $\gamma 2$-to-$\gamma 1$ after postprcossing using the EEG recordings in electrode No. 1 of Patient No. 19 in the MIT Physionet database.

Feature selection is important in limiting the number of the features input to a classifier in order to achieve a good classification performance and a less computationally intense classifier. In this section, features are ranked and a single feature is selected in a patient-specific manner. A universal spectral power ratio such as δ-to-α ratio (DAR) has been explored in [38], [39] for abnormality detection. However, ratio features or PSD features have to be chosen in a *patient-specific* manner [36]. One feature that works well for one patient may not work well for another patient.

A single feature is first selected for seizure prediction. The key reason for finding a single feature that provides acceptable prediction results is that systems using a single feature have the lowest hardware complexity and power consumption. To extract a single spectral power ratio feature from a single electrode, only one sensor needs to be implanted or placed and only spectral powers in two frequency bands need to be computed from the sensor. Therefore, this section describes the criteria used for the single feature selection and the classification method.

1) Feature Selection Criteria: Class separability is introduced to select the suboptimal group of linearly independent features. Let $\boldsymbol{f} = [f_1, f_2, \ldots, f_m]^T$ represents an m-dimensional feature vector. Define within-class scatter matrix ($\boldsymbol{S_w}$) and between-class scatter matrix ($\boldsymbol{S_b}$) as follows:

$$\boldsymbol{S_w} = \sum_{i=1}^{i=c} p_i \boldsymbol{\Sigma_i}$$

$$\boldsymbol{S_b} = \sum_{i=1}^{i=c} p_i (\boldsymbol{\mu_i} - \boldsymbol{\mu_0})(\boldsymbol{\mu_i} - \boldsymbol{\mu_0})^T$$

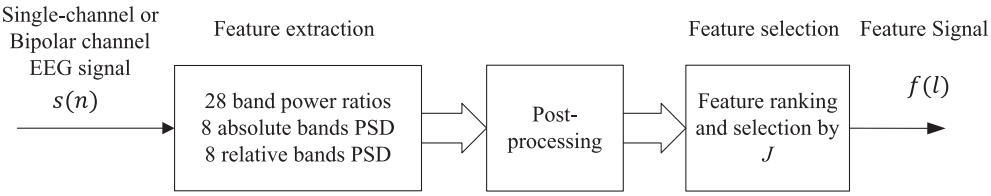

Fig. 2. Flow chart of single feature selection.

where c represents the number of classes, $\boldsymbol{\Sigma_i} = E[(\boldsymbol{f}-\boldsymbol{\mu_i})(\boldsymbol{f}-\boldsymbol{\mu_i})^T]$ represents the covariance matrix for class i, p_i represents the probability of class i, $\boldsymbol{\mu_0}$ represents the global mean vector, and $\boldsymbol{\mu_i}$ represents the mean vector for class i, respectively. The criterion

$$J = \frac{|\boldsymbol{S_w} + \boldsymbol{S_b}|}{|\boldsymbol{S_w}|}$$

takes a large positive value when samples in the m-dimensional space are well clustered within each class, and the clusters of the different classes are well separated [40]. The notation $|\boldsymbol{A}|$ represents the determinant of the matrix $\boldsymbol{A}$. To select a single feature, J is computed for all features from all electrodes and the feature that achieves the maximum J is selected.

The application of the class separability criteria is illustrated for Patient No. 1 from Freiburg database. For this patient, $\gamma 5$-to-$\gamma 4$ ratio of electrode No. 1 was selected as the best feature. Fig. 3 illustrates the $\gamma 5$-to-$\gamma 4$ ratio of electrode No. 1 before and after postprocessing using the (a) ictal and (b) interictal recordings of Patient No. 1 in the Freiburg EEG database, where the blue curves represent the feature signals before Kalman filter, the orange curves represent the outputs of the Kalman filter, and the red lines represent the thresholds and the black dashed lines represent seizure onsets, respectively. The feature in Fig. 3(a) corresponds to four different seizures where each seizure onset occurs at exactly 3000 second time stamp. The feature in Fig. 3(b) corresponds to interictal period of about 1 day duration. This particular ratio feature is shown to be a good seizure predictor for this patient as the feature always exceeds the threshold before seizure onset and is always below the threshold during interictal period.

2) Single Feature Classification: Since a feature input to the classifier is a one-dimensional signal, thresholding is used as the classifier. Receiver operating characteristic (ROC) is used to achieve the threshold. This classifier can be easily implemented in hardware with low power consumption.

The receiver operating characteristic (ROC) curve in classification theory finds the optimal thresholds by a plot of true positives (or sensitivity) versus false positives (or 1-speciticity). Regardless of the distribution of the two classes of data, the ROC tries to find optimal threshold between the two sets of data [40]. The reason for choosing this classification is that although finding the optimal threshold may take a long time during the training phase, the time to make a decision during the testing phase is very fast once the threshold is found by the algorithm.

During ROC analysis, the sensitivity is plotted as a function of false positive rate for each possible cut-off point. Therefore, each point on the curve corresponds to a particular cut-off

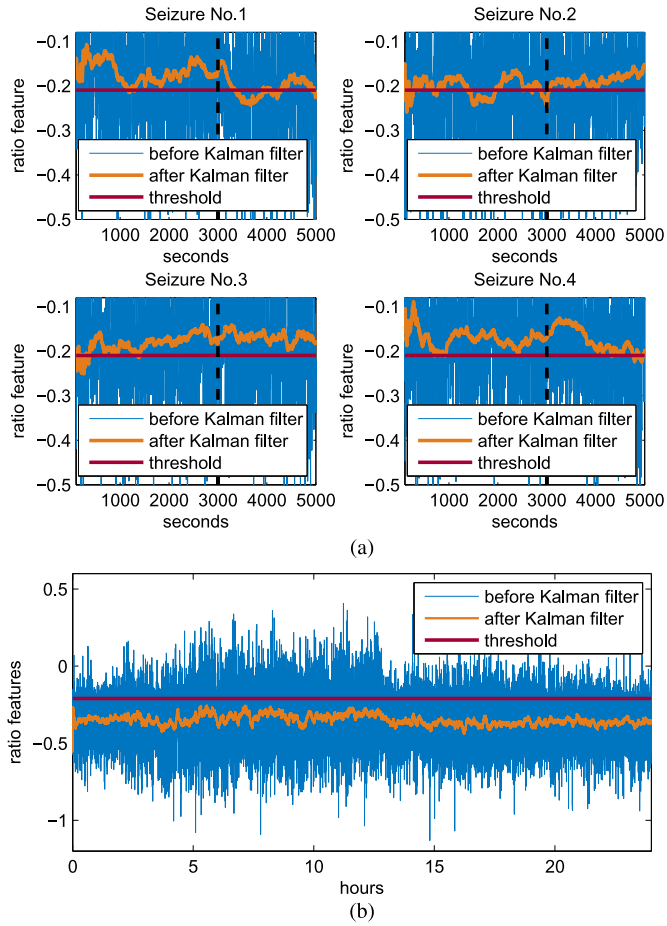

Fig. 3. Examples to illustrate the single ratio feature selected for seizure prediction and the power of the Kalman filter using the (a) ictal and (b) interictal recordings from Patient No. 1 in the Freiburg database.

threshold and specific values of sensitivity and specificity. A perfect classifier has an ROC curve that passes through the upper left corner or coordinate (0,1), which represents 100% sensitivity and 100% specificity. In general, the optimal point on the curve should be the one that is closest to the coordinate (0,1) on the curve and the optimal threshold is the one that corresponds to that point. Fig. 4 shows an example of ROC analysis where Patient No. 1's feature signal from the MIT EEG database is trained. The circled point on the figure corresponds to the optimal cut-off point found by the ROC algorithm.

D. Multi-Dimensional Feature Selection and Classification

While a single feature from a single electrode requires low hardware complexity and low power consumption, it only achieves good prediction results for patients whose seizures originate from the same location of the brain and are of the

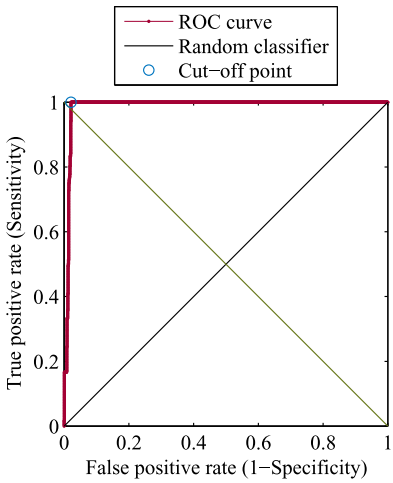

Fig. 4. ROC analysis using Patient No. 1's feature signal from the MIT EEG database.

same type. For patients who have multiple types of seizures that originate from multiple locations of the brain, multi-dimensional features from multiple electrodes need to be used to predict seizures. This section describes a novel *two-step* feature selection method for finding patient-specific multi-dimensional features that achieve acceptable prediction results for these patients. The multi-dimensional feature selection process is shown in Fig. 5, which includes feature basis selection, electrode selection, and optimal feature selection. The feature basis selection and optimal feature selection steps form the two steps of the proposed method. The electrode selection step is carried out before the second step and after the first step. Branch and bound (BAB) algorithm is used for optimal feature selection whose performance is then compared with that of the least absolute shrinkage and selection operator (LASSO) method. The output $\boldsymbol{f}(l)$ represents the l-th feature vector with dimension equal to r. The classifier used for prediction is cost-sensitive linear support vector machine (c-LSVM) [41], [42].

1) Feature Basis Selection: This section describes the method for selecting feature basis for each electrode. The goal is to select R linearly independent features that achieve the maximum linear separability criteria for each electrode, where R is determined by eigenvalue analysis. Feature basis selection is an essential step before electrode selection and before optimal feature selection for the reason that the input vectors to the BAB algorithm are required to be linearly independent. As described before, for each electrode, 44 features (8 absolute spectral powers, 8 relative spectral powers and 28 spectral power ratios) are extracted. An eigenvalue analysis of the covariance matrix of the features from each electrode is performed to find the maximum number of features that are linearly independent of each other. Fig. 6 shows the eigenvalues of the covariance matrix of the features sorted in a descending order from electrode No. 1 using patient No. 14's data from the MIT sEEG database. The largest nine eigenvalues are significantly higher than the remaining eigenvalues, which indicates that only nine out of the 44 features are linearly independent and the remaining features are redundant. Therefore, R is chosen to be 9.

The class separability method described in Section II-C-1 is used to select linearly independent features. The linearly independent features are selected sequentially in a greedy manner, which can be described as starting from an empty feature set, sequentially adding each of the features not yet selected such that the new feature combined with the selected features maximizes the objective function J until R features are selected. This process is repeated for each electrode. Such sequential selection scheme will produce a suboptimal group of features that are linearly independent. Detailed feature reduction scheme is described in Algorithm 1, where k represents the electrode number, K represents the total number of electrodes, f represents a feature selected out of the remaining features from electrode k only, and $J(k)$ represents the criteria value for electrode k. Algorithm 1 selects the R best features for each electrode such that the J value is maximized for each electrode.

Algorithm 1 Algorithm for feature basis selection

for electrode number $k = 1$ to K **do**
  Start with the empty set $S_0 = \{\phi\}, i = 0$
  for $i = 1$ to R **do**
    Select the next best feature $f^* = \arg\max_{f \notin S_{i-1}} J(S_{i-1} \cup \{f\})$
    $S_i = S_{i-1} \cup \{f^*\}$
  end for
  Compute $J(k)$
end for

However, it should be noted that this criterion takes infinite value when features are linearly dependent as S_w is rank-insufficient or ill-conditioned. To address this issue, the following modified criterion is used:

$$J = \begin{cases} \frac{|\boldsymbol{S_w}+\boldsymbol{S_b}|}{|\boldsymbol{S_w}|} & \text{if } \boldsymbol{S_w} \text{ is well-conditioned} \\ 0 & \text{otherwise} \end{cases}$$

where J is set to zero if the selected features are not linearly independent.

2) Electrode Selection: Electrode selection is then performed to limit the power consumed in sensing the signals from different locations of the brain. The criteria for electrode selection used in this paper can be described as selecting k electrodes such that features computed from the selected k electrodes satisfy maximum linear separability criteria J, where k represents the number of electrodes selected out of total electrodes, K. For example, if $k = 2$ and $K = 16$, J is computed for all *possible* pairs of electrodes out of 16 electrodes and the pairs with highest J is selected. The electrode selection and the second-step feature selection followed by classification are repeated iteratively until the classifier meets the specifications. The experimental results presented in Section III demonstrate that two iterations always suffice, i.e., no more than two electrodes need to be selected.

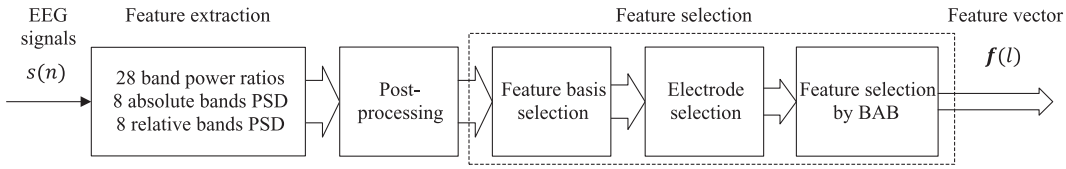

Fig. 5. Flow chart of single feature selection.

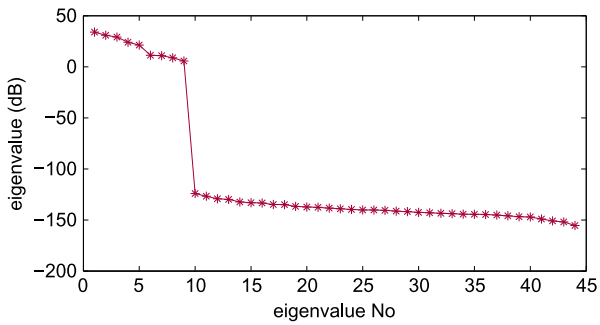

Fig. 6. Eigenvalues of the covariance matrix of the features using Patient No. 14's data from the MIT sEEG database.

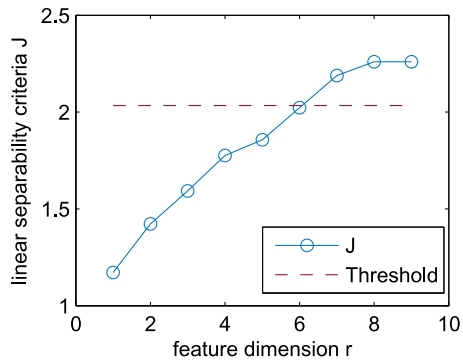

Fig. 7. Linear separability criteria J of the subset of features with different feature dimensions using Patient No. 14's recordings in electrode No. 14 from the MIT database.

3) Optimal Feature Selection by Branch and Bound: This section describes the method for the second round of feature selection after feature reduction and electrode selection to further reduce the number of features from R to r using branch and bound algorithm. Let $\boldsymbol{f}(l) = [f_1(l), f_2(l), \ldots, f_R(l)]^T$ represent the l-th column feature vector that consists of R selected feature samples computed from l-th windowed signal. Let y_l represent the class label for segment l. The goal of optimal feature selection is to select a subset of features (with dimension equal to r) that can produce the best classification result or achieve the maximum separability criteria. Such a problem could be extremely computationally intensive and usually, in practice, the number r is not even known *a priori*.

To simplify the proposed problem, a regression problem is introduced to select the subset of the features. Define $\boldsymbol{y} = [y_1, y_2, \ldots y_L]^T$ as the class label vector and define the feature matrix $\boldsymbol{F}$ as follows:

$$\begin{aligned}\boldsymbol{F} &= [\boldsymbol{f}(1), \boldsymbol{f}(2), \ldots, \boldsymbol{f}(L)]^T \\ &= \begin{bmatrix} f_1(1) & f_2(1) & \ldots & f_R(1) \\ f_1(2) & f_2(2) & \ldots & f_R(2) \\ . & . & \ldots & . \\ f_1(L) & f_2(L) & \ldots & f_R(L) \end{bmatrix} \\ &= [\boldsymbol{f}_1, \boldsymbol{f}_2, \ldots, \boldsymbol{f}_R]\end{aligned}$$

where $f_i(j)$ represents the feature i corresponding to segment j. Each row of $\boldsymbol{F}$ corresponds to the feature vector for segment l and each column of $\boldsymbol{F}$ represents a time series of a feature variable. Let $\boldsymbol{G_r} = [\boldsymbol{f}_{i_1}, \boldsymbol{f}_{i_2}, \ldots, \boldsymbol{f}_{i_r}]$ represent an r-variable subset of $\boldsymbol{F}$ where $i_1, i_2, \ldots, i_r$ represent the feature indices. The criteria used for feature selection in this paper is described as selecting a subset of features such that the least square fitting $\boldsymbol{y} = \boldsymbol{G_r} * \boldsymbol{q}$ achieves the minimum error. Mathematically, it can be described as finding $i_1, i_2, \ldots, i_r$ such that the following objective function

$$\varepsilon(\boldsymbol{G_r}) = \|\boldsymbol{y} - \boldsymbol{G_r} * \boldsymbol{q}\|$$

is minimized, where $\boldsymbol{q} = (\boldsymbol{G}^T\boldsymbol{G})^{-1}\boldsymbol{G}^T\boldsymbol{y}$ is the optimal projection vector.

In [43], an efficient branch and bound (BAB) algorithm is developed to solve the problem of selection of the globally optimal variables. The proposed BAB algorithm identifies the globally best feature variable subset such that the regression error ε is minimized.

As mentioned, the number of features r is not known *a priori*. The following steps are used to find r:

1) for each possible value of r, ($r \in \{1, 2, .., R\}$), use BAB to find the optimal subset of features with dimension equal to r.
2) evaluate the linear separability criteria J for all subsets of features.
3) select the subset of features with the minimum dimension of r^* such that its linear separability criteria J is greater than a predetermined threshold.

Fig. 7 shows the plot of linear separability criteria J versus feature dimension r using Patient No. 14's recordings in electrode No. 14 from the MIT database, where the red line represents the threshold equal to $\min\{3, 0.9\max(J)\}$. The value of r is chosen such that J exceeds the minimum of predetermined value of J_0 ($J_0 = 3$) and $0.9J_{max}$, where J_{max} is the maximum value of J over R features. As shown in the figure, the minimum r which achieves an objective function J greater than the threshold is 7. Therefore, the number of optimal features used for prediction is 7 ($r^* = 7$).

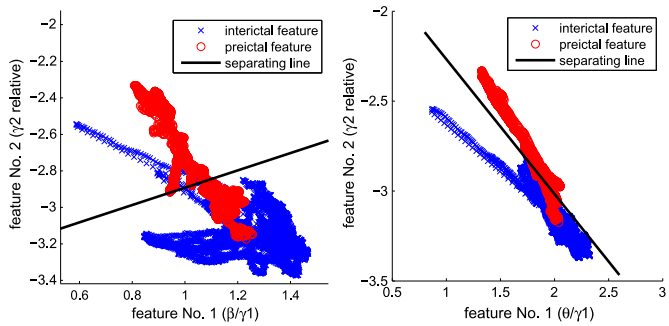

Fig. 8. Comparison the feature selection results of (a) LASSO and (b) BAB for Patient No. 15 in the Freiburg database.

4) Optimal Feature Selection by LASSO: Least absolute shrinkage and selection operator (LASSO) is one of the widely used selection methods for linear regression problem. It minimizes the total squared error with a penalty added to the number of the variables [44]. This paper uses LASSO as a baseline for feature variable selection and compares the performance of the BAB feature variable selection algorithm with LASSO. Therefore, the number of feature variables selected by LASSO is chosen to be same as the number chosen by the BAB algorithm.

For a given value of λ, a nonnegative parameter, LASSO solves the problem

$$\min J(\boldsymbol{q}) = \frac{1}{2L}\|\boldsymbol{y} - \boldsymbol{F} * \boldsymbol{q}\|^2 + \lambda|\boldsymbol{q}|$$

where L represents the number of observations, λ represents a nonnegative regularization parameter, and $|\boldsymbol{q}|$ represents the L^1 norm of the vector $\boldsymbol{q}$. As λ increases, less feature variables are selected as the number of nonzero components of $\boldsymbol{q}$ decreases. In this paper, λ is increased until the number of the nonzero components is the same as the number of feature variables selected by the BAB algorithm. This ensures a fair comparison between BAB and LASSO with respect to feature selection.

5) Comparison of BAB and LASSO: Fig. 8 compares the feature selection results of (a) LASSO and (b) BAB for Patient No. 15 in the Freiburg database. Fig. 8(a) illustrates the scatter plot of the 2-dimensional feature of $\gamma 2$ spectral power versus β-to-$\gamma 1$ spectral power ratio of electrode No. 2 selected by LASSO, where the cross points, cirle points and the black line represent the interictal features, preictal features and separating line, respectively. The 2-dimensional feature achieved a sensitivity of 100% and 3 FPs with a 30-minute refractory period. Fig. 8(b) illustrates the scatter plot of the 2-dimensional feature of $\gamma 2$ spectral power versus θ-to-$\gamma 1$ spectral power ratio of electrode No. 2 selected by BAB. The 2-dimensional feature achieved a sensitivity of 100% and 0 FPs for same refractory period. This example demonstrates that BAB performs better than LASSO with a 30-minute refractory period. A refractory period, which specifies a time period during which the system ignores all the subsequent alarms once it's triggered, is introduced to reduce the number of FPs in a short time period. The refractory period is set to be 30 minutes.

6) SVM and Classification: Recently, among all linear classifiers, Support Vector Machine (SVM) has attracted significant attention. Detailed descriptions of cost-sensitive linear SVM (c-LSVM) can be found in [40]. Generally speaking, the SVM seeks to find the solution to the following optimization problem:

$$\min J(\boldsymbol{w}, w_0, \boldsymbol{\xi}) = \frac{1}{2}\|\boldsymbol{w}\|^2 + C^+ \sum_{i\in C_1}^{N} \xi_i + C^- \sum_{j\in C_2}^{N} \xi_j$$
$$\text{subject to } y_i(\boldsymbol{w}^T \boldsymbol{x}_i + w_0) \geq 1 - \xi_i,\ i = 1, 2, \ldots, N$$
$$\xi_i \geq 0,\ i = 1, 2, \ldots, N$$

where $\boldsymbol{x}_i$ represents the r-dimensional feature vector, N represents the total number of feature vectors used for training the classifier, $\boldsymbol{w}$ represents the orientation of the discriminating hyperplane and w_0 represents the offset of the plane from the origin, y_i represents the class indicator ($y_i = +1$ if $\boldsymbol{x}_i$ is from class 1, otherwise $y_i = -1$), ξ_i represents the slack variable, and C^+, C^- represent the misclassification costs for two classes, respectively. After training, the decision function of a linear SVM is given by

$$f(\boldsymbol{x}) = sign\left(\sum_{i=1}^{N} \alpha_i y_i \boldsymbol{x_i}^T \boldsymbol{x} + b\right)$$

where $\boldsymbol{x}$ represents a new feature vector. The above equation can be simplified as follows:

$$f(\boldsymbol{x}) = sign(\boldsymbol{w}^T \boldsymbol{x} + b)$$

where $\boldsymbol{w} = \sum_i \alpha_i y_i \boldsymbol{x_i}$.

III. Experimental Results

The details for the proposed algorithm are described as follows:

1) Due to the imbalance between the data size of the preictal features and the interictal features, *random subsampling*, which refers to randomly selecting a subset of the feature objects, are performed on the interictal features. In our experiments, 20% of the interictal feature objects are randomly selected for training and the rest of the data are used for testing.
2) Leave-one-out cross validation is used in the training phase to (a) train a number of classifiers with feature vectors preceding the seizure left out in each turn (b) test on the remaining data. Final classifier which has the lowest FP rate on the interictal dataset is selected.
3) Three important criteria for performance evaluation include sensitivity (SS), false positive rate (FPR, the number of FP per hour) and seizure prediction horizon (SPH, time interval before a seizure when it's predicted). Min. SS and Max. FPR for each patient are predetermined as 80% and 0.125/hr, respectively. Multi-dimensional feature selection and classification are performed for patients where a single

TABLE I
PREDICTION PERFORMANCE OF THE PROPOSED SYSTEM USING A SINGLE FEATURE FOR FREIBURG DATABASE

Patient #	electrode #	Power ratio	# of SZ	SS	FPR	Max/Min SPH(min.)
1	1	$\gamma5/\gamma4$	4	100	0	47/33
3	6	$\gamma4/\beta$	5	100	0	47/16
4	1	$\gamma5/\beta$	5	100	0	50/40
7	4	$\gamma5/\beta$	3	100	0	50/50
9	5	$\gamma4/\gamma3$	5	100	0.083	50/50
10	5	α/θ	5	100	0.083	47/33
11	1	$\gamma1/\beta$	4	100	0.125	47/25
12	6	$\gamma4/\gamma5$	4	100	0	50/24
14	6	$\gamma1/\gamma2$	4	100	0.042	50/25
16	1	$\gamma4/\alpha$	5	100	0.042	40/16
17	4	$\theta/\gamma1$	5	100	0	45/25
21	5	β/α	5	100	0.083	27/20

TABLE II
PREDICTION PERFORMANCE OF THE PROPOSED SYSTEM USING A SINGLE FEATURE FOR MIT DATABASE

Patient #	electrode #	Power ratio	# of SZ	SS	FPR	Max/Min SPH(min.)
1	17	$\alpha/\gamma4$	6	100	0.024	60/3
8	20	$\alpha/\gamma4$	5	100	0.1	60/30
11	14	$\gamma5/\gamma3$	3	100	0.086	18/12
18	1	$\gamma3/\theta$	4	100	0.114	75/3
19	1	$\gamma2/\gamma1$	3	100	0	48/18
20	12	θ/β	6	100	0.071	60/20
21	1	$\gamma1/\beta$	3	100	0.065	78/3

feature is not able to achieve the predetermined requirements.

4) Window size is chosen as 4 seconds. Since sampling frequency is 256 Hz for both databases, each segment contains $4 * f_s = 1024$ samples.
5) The cost value C in SVM is selected from the set $\{4^{-6}, 4^{-5}, 4^{-4}, \ldots, 4^5, 4^6\}$. The cost ratio C^+/C^- is selected from the set $\{2^{-3}, 2^{-2}, \ldots, 2^2, 2^3\}$.

Systems using a single feature achieved a sensitivity of 100% and FPR less than 0.1 for 12 patients in the Freiburg database and for 7 patients in the MIT database. Test Results for these 12 patients in the Freiburg database and for the 7 patients in the MIT database are shown in Table I and in Table II, respectively, where "SZ" stands for seizures. Details about the spectral power ratio used for prediction are shown in the third column, where the symbol $\alpha/\gamma3$, for instance, indicates that the spectral power ratio between power in α band and power in $\gamma3$ band is used. For the rest of the patients, single feature classification can not achieve a minimum sensitivity of 80% or a FPR less than 0.125.

Test Results using multi-dimensional features for the remaining 6 patients in Freiburg database and for the remaining 10 patients in MIT database are shown in Table III and in Table IV, respectively. Details about the spectral power ratios, relative spectral powers, absolute spectral powers used for prediction are shown in the 3rd, 4th and 5th columns, respectively.

Summary of the overall prediction performance for both databases is shown in Table V. For Freiburg intra-cranial EEG database, the proposed algorithm achieved a sensitivity of 100% and a FPR of 0.032 using 1.167 electrodes and 2.78 features on average. For MIT scalp EEG database, the proposed

TABLE III
PREDICTION PERFORMANCE OF THE PROPOSED SYSTEM USING BAB FOR FREIBURG DATABASE

Patient #	electrode No.	Power ratio	Rel. power	Abs. Power	# of SZ	SS	FPR	Max/Min SPH(min.)
5	1	$\frac{\theta}{\beta}, \frac{\gamma2}{\gamma5}, \frac{\gamma3}{\gamma4}$		$\gamma1$	5	100	0.039	54/39
	6	$\frac{\theta}{\alpha}, \frac{\alpha}{\gamma2}, \frac{\alpha}{\gamma4}$	θ					
6	2	$\frac{\theta}{\alpha}, \frac{\beta}{\gamma3}, \frac{\gamma1}{\gamma2}$	$\gamma2$		3	100	0.042	46/30
15	2	$\frac{\theta}{\gamma1}$	$\gamma2$		4	100	0	50/36
18	2	$\frac{\gamma1}{\gamma5}$	$\gamma4$		5	100	0	50/50
19	1	$\frac{\theta}{\gamma5}, \frac{\alpha}{\beta}, \frac{\beta}{\gamma2}, \frac{\beta}{\gamma5}, \frac{\gamma2}{\gamma3}$	θ	$\gamma4$	4	100	0.042	50/41
	2	$\frac{\theta}{\gamma4}, \frac{\beta}{\gamma5}, \frac{\gamma1}{\gamma3}, \frac{\gamma4}{\gamma5}$		$\gamma5$				
20	1	$\frac{\alpha}{\beta}, \frac{\gamma3}{\gamma4}$		$\gamma2$	5	100	0	50/43
	2	$\frac{\theta}{\alpha}, \frac{\theta}{\gamma3}, \frac{\gamma1}{\gamma5}$	$\gamma1$,$\gamma4$	β,$\gamma2$				

TABLE IV
PREDICTION PERFORMANCE OF THE PROPOSED SYSTEM USING BAB FOR MIT DATABASE

Patient #	electrode No.	Power ratio	Rel. power	Abs. Power	# of SZ	SS	FPR	Max/Min SPH(min.)
2	18	$\frac{\gamma1}{\gamma4}, \frac{\gamma2}{\gamma5}$	θ,α		3	100	0.029	60/39
3	7	$\frac{\beta}{\gamma2}, \frac{\gamma1}{\gamma4}, \frac{\gamma3}{\gamma4}$	θ	$\gamma4$	5	100	0	68/15
	8	$\frac{\gamma1}{\gamma4}, \frac{\gamma2}{\gamma3}$		$\gamma2$				
5	8	$\frac{\theta}{\alpha}, \frac{\theta}{\gamma2}, \frac{\theta}{\gamma3}, \frac{\beta}{\gamma1}, \frac{\beta}{\gamma4}, \frac{\gamma4}{\gamma5}$	θ,$\gamma2$		5	100	0.051	60/10
6	8	$\frac{\gamma1}{\gamma2}, \frac{\gamma3}{\gamma4}$		$\gamma2$	6	83.3	0.045	72/21
	21	$\frac{\alpha}{\beta}, \frac{\alpha}{\gamma5}, \frac{\gamma2}{\gamma3}, \frac{\gamma4}{\gamma5}$		$\gamma1$,$\gamma5$				
9	12	$\frac{\theta}{\gamma3}, \frac{\alpha}{\beta}, \frac{\gamma1}{\gamma5}, \frac{\gamma2}{\gamma4}$	$\gamma4$		3	100	0.046	69/33
	18	$\frac{\gamma1}{\gamma3}, \frac{\gamma2}{\gamma4}$		α				
10	1	$\frac{\beta}{\gamma3}, \frac{\gamma3}{\gamma4}, \frac{\gamma3}{\gamma5}$	$\gamma2$	α,$\gamma2$	7	100	0.060	69/24
	18	$\frac{\theta}{\gamma4}, \frac{\alpha}{\beta}, \frac{\alpha}{\gamma2}$	$\gamma5$	θ,$\gamma3$				
13	1	$\frac{\gamma1}{\gamma2}, \frac{\gamma3}{\gamma4}$	$\gamma2$,$\gamma5$	θ	6	100	0.030	68/18
	18	$\frac{\gamma1}{\gamma2}, \frac{\gamma3}{\gamma4}$	$\gamma2$,$\gamma5$	θ				
14	14	$\frac{\theta}{\alpha}, \frac{\theta}{\gamma4}, \frac{\gamma3}{\gamma4}, \frac{\gamma4}{\gamma5}$	θ	β,$\gamma2$	5	100	0.039	60/6
16	27	$\frac{\beta}{\gamma3}, \frac{\gamma1}{\gamma4}, \frac{\gamma1}{\gamma5}, \frac{\gamma2}{\gamma5}$		θ	3	100	0	58/42
22	11	$\frac{\theta}{\gamma3}, \frac{\beta}{\gamma2}, \frac{\gamma1}{\gamma5}, \frac{\gamma2}{\gamma4}, \frac{\gamma3}{\gamma5}$	θ		3	100	0.032	78/11

TABLE V
OVERALL PREDICTION PERFORMANCE OF THE PROPOSED SYSTEM FOR FREIBURG AND MIT DATABASE

Database	EEG type	Mean # of electrodes	Mean# of features	SS	FPR
Freiburg	iEEG	1.167	2.78	100	0.0324
MIT	sEEG	1.294	5.05	98.68	0.0465

algorithm achieved a sensitivity of 98.68% and a FPR of 0.0465 using 1.29 electrodes and 5.05 features on average.

Table VI and Table VII compare the prediction performance between LASSO and BAB for the Freiburg database and MIT database, respectively. Three criteria are used to measure the prediction performance, which include sensitivity, number of false positives (FP) and number of support vectors (SV). As shown in Table VI for the Freiburg database, the LASSO method not only leads to a larger number of FPs, but also requires a significantly larger number of SVs except for patient No. 6. As shown in Table VII for the MIT database, LASSO has about the same number of SVs as BAB, but has a lower sensitivity and a larger number of FPs.

IV. SYSTEM ARCHITECTURE

This section describes the system architecture using the methods described in the previous sections. Based on the methods proposed in the previous sections, the seizure prediction system contains 3 parts which include (1) PSD estimation, (2) feature extraction, and (3) classifier.

TABLE VI
COMPARISON OF PREDICTION PERFORMANCE BETWEEN BAB AND LASSO FOR FREIBURG DATABASE

Patient #	# of SZ	SS		# of FP		# of SVs	
		BAB	LASSO	BAB	LASSO	BAB	LASSO
5	5	100	100	1	3	4191	6391
6	3	100	100	1	2	2526	2482
15	4	100	100	0	3	1699	4411
18	5	100	100	0	0	2244	5012
19	4	100	100	1	1	2123	2540
20	5	100	100	0	0	3679	4471

TABLE VII
COMPARISON OF PREDICTION PERFORMANCE BETWEEN BAB AND LASSO FOR MIT DATABASE

Patient #	# of SZ	SS		# of FP		# of SVs	
		BAB	LASSO	BAB	LASSO	BAB	LASSO
2	3	100	100	1	3	3719	4771
3	5	100	100	0	0	5027	5470
5	5	100	100	2	2	6780	6751
6	6	83.3	83.3	3	5	4454	4524
9	3	100	100	2	2	3988	3921
10	7	100	85.71	3	3	8212	8546
13	6	100	100	1	3	9696	10452
14	5	100	100	1	1	7727	7643
16	3	100	100	0	0	3331	3412
22	3	100	100	1	2	4943	4307

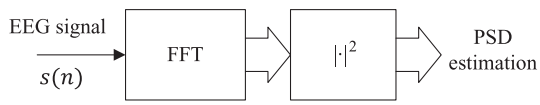

Fig. 9. System architecture for PSD estimation.

A. PSD Estimation

Fig. 9 illustrates the system architecture for PSD estimation. The PSD of the input signal is estimated by first computing the fast Fourier transform (FFT) of the input segmented signal and then computing the magnitude square of the FFT coefficients. A 1024-point real FFT is required in the system as each input segment is 4 seconds long and thus contains $4 * 256 = 1024$ samples.

Fig. 10 shows the proposed fully-real serial 1024-point FFT architecture in [45]. Table VIII presents the synthesis results obtained for the proposed real FFT architectures in [45]. The two designs were synthesized using a clock speed of 100 MHz in Synopsys Design Compiler with 45 nm NCSU PDK. The interleaved architecture can process FFT computations of two electrodes using same pipelined hardware in an interleaved manner. The proposed 1024-point real number FFT (RFFT) architecture in [45] reuires $\log_2 512 - 3 = 6$ complex multipliers and $3 * 1024/2 - 5 = 1531$ delay elements to compute the FFT coefficients. It requires an area of $0.284327\ \mathrm{mm}^2$ and a power of 14.8012 mW. Therefore, computing FFT coefficients for a single input segment requires a total energy of $14.8\ \mathrm{mW}/100\ \mathrm{MHz} * 1531 = 226.6\ \mathrm{nJ}$ as the operations are completed in 1531 clock cycles.

TABLE VIII
SYNTHESIS RESULTS OF 1024-POINT SERIAL RFFT FOR 100 MHZ CLOCK FREQUENCY

Fully real	0.284327 mm^2	14.8012 mW
Fully real-Interleaved by factor 2	0.375221 mm^2	17.7314 mW

B. Feature Extractor

Fig. 11 illustrates the system architectures for extracting (a) a single absolute spectral power in a specific band, (b) a relative spectral power in a specific band, and (c) a ratio of spectral powers in two bands from the PSD coefficients computed in the previous step. As shown in Fig. 11, extracting these features from the PSD coefficients requires far less number of multipliers than the PSD estimation.

C. Classifier

This section illustrates the architecture for linear SVM, computes the approximate energy for linear SVM and RBF-SVM, and shows the reason why kernel SVM such as radial basis function kernel SVM (RBF-SVM) is not preferred. Fig. 12 illustrates the system architectures for a linear SVM. In [33], a low-energy architecture based on approximate computing by exploiting the inherent error resilience in the SVM computation was proposed. According to [33], the computational complexity of a linear SVM only depends on the feature dimension. However, the computational complexity of a RBF-SVM consists of 2 parts, which include kernel computation and decision variable computation. The computational complexity of a RBF-SVM classifier is not only proportional to the feature dimension, but also to the number of support vectors (SVs). Table IX compares the number of support vectors after training using linear SVM and RBF-SVM for Patient No. 10 and Patient No. 13 in the MIT database. The fourth and fifth columns of Table IX show the approximate estimates of the energy in kernel computation and decision variable computation per test vector using the results in [33]. The last column shows the total energy per test vector. As shown in the table, even though RBF-SVM requires significantly less number of SVs than the linear SVM, its energy requirement is 3 orders of magnitude larger than the linear SVM.

Thus, regardless of the energy required in sensors and analog-to-digital converters (ADC), the total energy required in feature extraction and classification using a single electrode is approximately 227 nJ when linear SVM is used. That number is increased to $2 * 227 = 454$ nJ for Patient No. 10 and for Patient No. 13 in the MIT database as the interleaved architecture requires twice the number of clock cycles for feature extraction. When RBF-SVM is used, the energy consumption increases to 586 nJ and 490 nJ per test vector for Patient No. 10 and for Patient No. 13 in the MIT database, respectively. These energy consumption estimates are obtained by interpolating the energy estimates in [33], [45]. The energy consumption of the Kalman filter is not included in this analysis. The RBF-SVM not only requires more energy consumption, it also requires additional hardware for approximately 23900 multiplications and 1992 RBF kernel computations for Patient No. 10, and for 6000 multiplications and 585 RBF kernel computations for Patient No. 13. The number of multiplications increases by a factor of

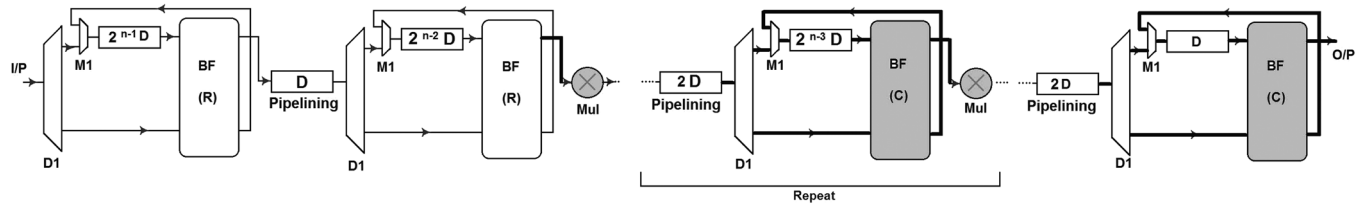

Fig. 10. Fully real serial FFT architecture.

TABLE IX
COMPARISON OF ENERGY CONSUMPTION BETWEEN LINEAR SVM AND RBF-SVM FOR MIT DATABASE

Patient #	# of features	# of SVs		kernel		decision variable		classifier energy		total energy	
		SVM	RBF-SVM	SVM	RBF-SVM	SVM	RBF-SVM	SVM	RBF-SVM	SVM	RBF-SVM
10	12	8212	1992	–	108 nJ	32 pJ	24 nJ	32 pJ	132 nJ	454 nJ	586 nJ
13	10	9696	585	–	30 nJ	30 pJ	6 nJ	30 pJ	36 nJ	454 nJ	490 nJ

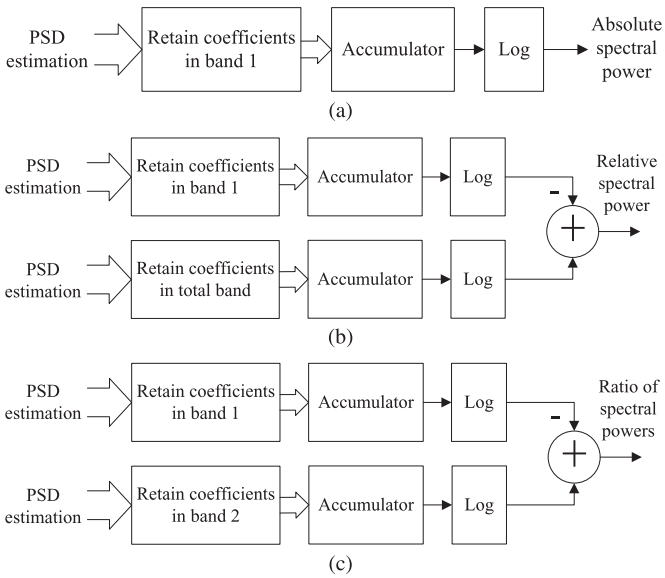

Fig. 11. System architectures for extracting (a) a single absolute spectral in a specific band, (b) a relative spectral power in a specific band, and (c) a ratio of spectral powers in two bands from the PSD coefficients.

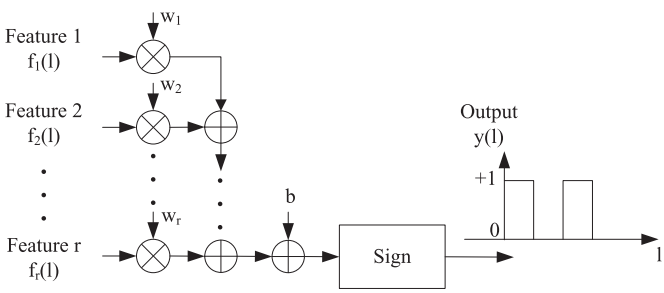

Fig. 12. System architecture for linear SVM.

N_{sv} for RBF-SVM, where N_{sv} represents the number of support vectors. Furthermore, N_{sv} additional kernel evaluation are needed in the RBF-SVM.

V. DISCUSSION

Many approaches have been presented for predicting seizures in epileptic patients. Various types of linear and nonlinear features have been used for seizure prediction. Our results are compared directly to several other studies that have tested prediction algorithms using the same Freiburg EEG database, [10], [13], [17], [28]–[31], [46]–[48] or MIT EEG database [49]. Our results may also be compared to studies using other databases [16], [23]. We demonstrate high sensitivity, low FPR, and low feature dimension for these two databases.

Table X compares the system performance of the proposed algorithm with prior works. The proposed algorithm for seizure prediction, using the least number of features selected by the BAB algorithm (for iEEG), achieves the highest sensitivity (for iEEG) and the lowest FPR.

Even though the proposed algorithm has been tested on short duration EEG data, future work will be directed towards analysis on long term EEG recordings.

Another evaluation criterion, successful patient rate, was proposed in [50] and is used to evaluate the success of a seizure prediction algorithm. A patient is considered as a successful patient if the sensitivity is 100% and the FP rate is lower than 0.2. We achieved a FPR of 0 for 10 out of 19 patients in the Freiburg database and for 3 out of 17 patients for the MIT database. We also achieved a successful patient rate of 100% for the Freiburg database and a successful patient rate of 94.1% for the MIT database.

System performance is degraded for the scalp EEG recordings as the MIT (sEEG) database has a lower sensitivity, a lower successful patient rate, and a higher FP rate than the Freiburg (iEEG) database. This is caused by the fact that intracranial EEG recordings usually have a higher spatial resolution and signal-to-noise ratio due to greater proximity to neural activity. Therefore, sEEG is a much noisier measurement of the neural activity and is highly suspectable to the interferences from the outer environment than the iEEG, which leads to the decrease of sensitivity and the increase of FP rate. However, since iEEG is an invasive signal, the process to obtain invasive EEG recordings brings the risk of infections. Furthermore, the patient's hospital stay for surgery to implant these electrodes can be expensive. In addition, the sEEG has a larger coverage of the brain than iEEG.

In addition, the proposed seizure prediction algorithm using BAB for feature selection has several advantages over using LASSO for feature selection. The BAB algorithm achieves a

TABLE X
COMPARISON TO PRIOR WORK

Reference	EEG type	Sensitivity	FPR	Feature Type	No. of features
Chisci *et al.* 2010	iEEG	100	0.17	AR coeff.	36
Wang *et al.* 2014	iEEG	98.80	0.054	Amplitude and Frequency	125
Park *et al.* 2011	iEEG	97.5	0.27	PSD	36
Ozdemir *et al.* 2014	iEEG	96.55	0.21	Hilbert Spectrum	14.49
Ayinala *et al.* 2012	iEEG	94.37	0.14	PSD	4.8
Aarabi *et al.* 2014	iEEG	92.60	0.15	Model parameters	72
Williamson *et al.* 2011	iEEG	90.8	0.095	correlation	36
Aschenbrenner *et al.* 2003	iEEG	84.2	1.0	correlation	25
Zheng *et al.* 2013	iEEG	80	0.17	Phase Coherence	3
Maiwald *et al.* 2004	iEEG	41.5	0.15	correlation	25
Bandarabadi *et al.* 2014	iEEG	75.8	0.1	PSD	9.9
Alexandre *et al.* 2014	iEEG	50	0.15	Various	22
Khammari *et al.* 2012	sEEG	85	–	PSD	30
proposed iEEG	iEEG	100	0.032	PSD ratio	2.78
proposed sEEG	sEEG	98.68	0.047	PSD ratio	5.05

higher sensitivity and a lower FPR for both databases. The BAB algorithm also requires a smaller number of SVs than LASSO on the Freiburg database.

Finally, the total energy consumption of the system using linear SVM is reduced by 8% to 23% compared to system using RBF-SVM. In analysis of long-term EEG data, number of support vectors will increase proportionally to the number of total feature vectors. Thus, the energy consumption of a RBF-SVM will be greatly increased when long-term EEG is analyzed, and the reduction in total energy consumption of the system using linear SVM will be greatly increased compared to the system using RBF-SVM.

VI. CONCLUSION

In this paper, a patient-specific algorithm for seizure prediction using unipolar or bipolar EEG signals from either one or two channels has been proposed. This algorithm achieves a sensitivity of 100%, a successful patient rate of 100% a FP rate of 0.032 per hour on average for iEEG recordings, and achieves a sensitivity of 98.68%, a successful patient rate of 94.1% and a FP rate of 0.047 per hour on average for iEEG recordings. Compared with the results in [10], [13], [17], [28]–[31], [46]–[49], the proposed algorithm uses the fewest number of features and achieves a high sensitivity and a lower FP rate. The proposed approach reduces the complexity and area by about 2 to 3 orders of magnitude. We conclude that using discriminative sparse important features and using a simple classifier such as linear SVM can lead to higher sensitivity and specificity compared to processing hundreds of features with a complex classifier such as RBF-SVM.

Many algorithms that work well on short EEG recordings (like one day) fail to work on longer recordings (i.e., several days to weeks). Future work will be directed towards validating the proposed approach on longer term recordings. The spectral powers in eight subbands are sufficient for signals sampled at 256 Hz. However, further research needs to be directed to find out how many subbands are sufficient for high-frequency recordings such as 1 kHz or 2 kHz.

VII. DISCLOSURE

Part of the work in this paper was carried out at Leanics Corporation. The contents of this paper are covered in the U.S. Patent Application [51].

ACKNOWLEDGMENT

The authors are grateful to all three anonymous reviewers for their constructive comments that improved the quality of this paper.

REFERENCES

[1] H. M. de Boer, M. Mula, and J. W. Sander, "The global burden and stigma of epilepsy," *Epilepsy Behav.*, vol. 12, no. 4, pp. 540–546, 2008.
[2] M. Leonardi and T. B. Ustun, "The global burden of epilepsy," *Epilepsia*, vol. 43, no. s6, pp. 21–25, 2002.
[3] J. Engel, *Seizures and Epilepsy*. Cambridge, U.K.: Oxford Univ. Press, 2013, vol. 83.
[4] J. E. Leestma, M. B. Kalelkar, S. S. Teas, G. W. Jay, and J. R. Hughes, "Sudden unexpected death associated with seizures: Analysis of 66 cases," *Epilepsia*, vol. 25, no. 1, pp. 84–88, 1984.
[5] P. Perucca, F. Dubeau, and J. Gotman, "Intracranial electroencephalographic seizure-onset patterns: Effect of underlying pathology," *Brain*, vol. 137, no. 1, pp. 183–196, 2014.
[6] F. Mormann, T. Kreuz, C. Rieke, R. G. Andrzejak, A. Kraskov, P. David, C. E. Elger, and K. Lehnertz, "On the predictability of epileptic seizures," *Clin. Neurophysiol.*, vol. 116, no. 3, pp. 569–587, 2005.
[7] H. Witte, L. D. Iasemidis, and B. Litt, "Special issue on epileptic seizure prediction," *IEEE Trans. Biomed. Eng.*, vol. 50, no. 5, pp. 537–539, 2003.
[8] F. Mormann, R. G. Andrzejak, C. E. Elger, and K. Lehnertz, "Seizure prediction: The long and winding road," *Brain*, vol. 130, no. 2, pp. 314–333, 2007.
[9] B. Litt and J. Echauz, "Prediction of epileptic seizures," *Lancet Neurol.*, vol. 1, no. 1, pp. 22–30, 2002.
[10] Y. Park, L. Luo, K. K. Parhi, and T. Netoff, "Seizure prediction with spectral power of EEG using cost-sensitive support vector machines," *Epilepsia*, vol. 52, no. 10, pp. 1761–1770, 2011.
[11] G. Alarcon, C. Binnie, R. Elwes, and C. Polkey, "Power spectrum and intracranial EEG patterns at seizure onset in partial epilepsy," *Electroencephalogr. Clin. Neurophysiol.*, vol. 94, no. 5, pp. 326–337, 1995.
[12] P. Celka and P. Colditz, "A computer-aided detection of EEG seizures in infants: A singular-spectrum approach and performance comparison," *IEEE Trans. Biomed. Eng.*, vol. 49, no. 5, pp. 455–462, 2002.
[13] L. Chisci, A. Mavino, G. Perferi, M. Sciandrone, C. Anile, G. Colicchio, and F. Fuggetta, "Real-time epileptic seizure prediction using AR models and support vector machines," *IEEE Trans. Biomed. Eng.*, vol. 57, no. 5, pp. 1124–1132, 2010.

[14] M. Saab and J. Gotman, "A system to detect the onset of epileptic seizures in scalp EEG," *Clin. Neurophysiol.*, vol. 116, no. 2, pp. 427–442, 2005.
[15] F. Mormann, K. Lehnertz, P. David, and C. E Elger, "Mean phase coherence as a measure for phase synchronization and its application to the EEG of epilepsy patients," *Physica D*, vol. 144, no. 3, pp. 358–369, 2000.
[16] C. Alexandre Teixeira, B. Direito, M. Bandarabadi, M. Le Van Quyen, M. Valderrama, B. Schelter, A. Schulze-Bonhage, V. Navarro, F. Sales, and A. Dourado, "Epileptic seizure predictors based on computational intelligence techniques: A comparative study with 278 patients," *Comput. Methods Programs Biomed.*, vol. 114, no. 3, pp. 324–336, 2014.
[17] N. Wang and M. R. Lyu, Extracting and Selecting Distinctive EEG Features for Efficient Epileptic Seizure Prediction, 2014.
[18] M. Le Van Quyen, J. Martinerie, M. Baulac, and F. Varela, "Anticipating epileptic seizures in real time by a non-linear analysis of similarity between EEG recordings," *Neuroreport*, vol. 10, no. 10, pp. 2149–2155, 1999.
[19] L. D. Iasemidis, D.-S. Shiau, W. Chaovalitwongse, J. C. Sackellares, P. M. Pardalos, J. C. Principe, P. R. Carney, A. Prasad, B. Veeramani, and K. Tsakalis, "Adaptive epileptic seizure prediction system," *IEEE Trans. Biomed. Eng.*, vol. 50, no. 5, pp. 616–627, 2003.
[20] M. D'Alessandro, R. Esteller, G. Vachtsevanos, A. Hinson, J. Echauz, and B. Litt, "Epileptic seizure prediction using hybrid feature selection over multiple intracranial EEG electrode contacts: A report of four patients," *IEEE Trans. Biomed. Eng.*, vol. 50, no. 5, pp. 603–615, 2003.
[21] G. A. Worrell, L. Parish, S. D. Cranstoun, R. Jonas, G. Baltuch, and B. Litt, "High-frequency oscillations and seizure generation in neocortical epilepsy," *Brain*, vol. 127, no. 7, pp. 1496–1506, 2004.
[22] L. D. Iasemidis, D. S. Shiau, P. M. Pardalos, W. Chaovalitwongse, K. Narayanan, A. Prasad, K. Tsakalis, P. R. Carney, and J. C. Sackellares, "Long-term prospective on-line real-time seizure prediction," *Clin Neurophysiol.*, vol. 116, no. 3, pp. 532–544, 2005.
[23] M. Bandarabadi, C. A. Teixeira, J. Rasekhi, and A. Dourado, "Epileptic seizure prediction using relative spectral power features," *Clin Neurophysiol.*, 2014.
[24] M. J. Cook, T. J. O'Brien, S. F. Berkovic, M. Murphy, A. Morokoff, G. Fabinyi, W. D'Souza, R. Yerra, J. Archer, and L. Litewka *et al.*, "Prediction of seizure likelihood with a long-term, implanted seizure advisory system in patients with drug-resistant epilepsy: A first-in-man study," *Lancet Neurol.*, vol. 12, no. 6, pp. 563–571, 2013.
[25] A. Yang, D. H. Arndt, R. A. Berg, J. L. Carpenter, K. E. Chapman, D. J. Dlugos, W. B. Gallentine, C. C. Giza, J. L. Goldstein, and C. D. Hahn *et al.*, "Development and validation of a seizure prediction model in critically ill children," *Seizure*, 2014.
[26] S. Ramgopal, S. Thome-Souza, M. Jackson, N. E. Kadish, I. S. Fernández, J. Klehm, W. Bosl, C. Reinsberger, S. Schachter, and T. Loddenkemper, "Seizure detection, seizure prediction, closed-loop warning systems in epilepsy," *Epilepsy Behav.*, vol. 37, pp. 291–307, 2014.
[27] C. N. Heck, D. King-Stephens, A. D. Massey, D. R. Nair, B. C. Jobst, G. L. Barkley, V. Salanova, A. J. Cole, M. C. Smith, and R. P. Gwinn *et al.*, "Two-year seizure reduction in adults with medically intractable partial onset epilepsy treated with responsive neurostimulation: Final results of the rns system pivotal trial," *Epilepsia*, vol. 55, no. 3, pp. 432–441, 2014.
[28] N. Ozdemir and E. Yildirim, "Patient specific seizure prediction system using hilbert spectrum and bayesian networks classifiers," *Comput. Math. Methods Med.*, vol. 2014, 2014.
[29] M. Ayinala and K. K. Parhi, "Low complexity algorithm for seizure prediction using adaboost," in *Proc. IEEE Eng. Med. Biol. Soc. Conf.*, 2012, pp. 1061–1064.
[30] A. Aarabi and B. He, "Seizure prediction in hippocampal and neocortical epilepsy using a model-based approach," *Clin. Neurophysiol.*, vol. 125, no. 5, pp. 930–940, 2014.
[31] J. R. Williamson, D. W. Bliss, and D. W. Browne, "Epileptic seizure prediction using the spatiotemporal correlation structure of intracranial EEG," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process*, 2011, pp. 665–668.
[32] M. Bandarabadi, C. Teixeira, T. Netoff, K. K. Parhi, and A. Dourado, "Robust and low complexity algorithms for seizure detection," in *Proc. IEEE Eng. Med. Biol. Soc. Conf.*, 2014.
[33] M. Ayinala and K. K. Parhi, "Low-energy architectures for support vector machine computation," in *Proc. IEEE Asilomar Conf. Signals Syst. Comput. Conf. Rec.*, 2013, pp. 2167–2171.
[34] Seizure Prediction Project Freiburg, Germany [Online]. Available: https://epilepsy.uni-freiburg.de/freiburg-seizure-predictionproject/eeg-database
[35] A. L. Goldberger, L. A. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley, "PhysioBank, PhysioToolkit, PhysioNet: Components of a new research resource for complex physiologic signals," *Circulation,* vol. 101, no. 23, pp. e215–e220, 2000, 10.1161/01.CIR.101.23.e215 [Online]. Available: http://circ.ahajournals.org/cgi/content/full/101/23/e215 PMID:1085218
[36] K. K. Parhi and Z. Zhang, "Seizure prediction using ratio of spectral power from single EEG electrode," in *Proc. 6th Int. Workshop Seizure Prediction*, 2013, p. 39.
[37] S. S. Haykin, *Adaptive Filter Theory*, 4th ed. Englewood Cliffs, NJ, USA: Prentice Hall, 2002.
[38] J. Leon-Carrion, J. F. Martin-Rodriguez, J. Damas-Lopez, J. M. Barroso y Martin, and M. R. Dominguez-Morales, "Delta-alpha ratio correlates with level of recovery after neurorehabilitation in patients with acquired brain injury," *Clin. Neurophysiol.*, vol. 120, no. 6, pp. 1039–1045, 2009.
[39] J. Claassen, L. J. Hirsch, K. T. Kreiter, E. Y. Du, E. Sander Connolly, R. G. Emerson, and S. A. Mayer, "Quantitative continuous EEG for detecting delayed cerebral ischemia in patients with poor-grade subarachnoid hemorrhage," *Clin. Neurophysiol.*, vol. 115, no. 12, pp. 2699–2710, 2004.
[40] S. Theodoridis and K. Koutroumbas*, Pattern Recognition*. New York, NY, USA: Academic, 2008.
[41] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, 1995.
[42] Y. Lin, Y. Lee, and G. Wahba, "Support vector machines for classification in nonstandard situations," *Mach. Learn.*, vol. 46, no. 1–3, pp. 191–202, 2002.
[43] V. Kariwala, L. Ye, and Y. Cao, "Branch and bound method for regression-based controlled variable selection," *Comput. Chem. Eng.*, vol. 54, pp. 1–7, 2013.
[44] R. Tibshirani, "Regression shrinkage and selection via the lasso," *J. Royal Stat. Soc., Ser. B (Methodological)*, pp. 267–288, 1996.
[45] A. Chinnapalanichamy and K. K. Parhi, "Serial and interleaved architectures for computing real FFT," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, 2015, pp. 1066–1070.
[46] T. Maiwald, M. Winterhalder, R. Aschenbrenner-Scheibe, H. U. Voss, A. Schulze-Bonhage, and J. Timmer, "Comparison of three nonlinear seizure prediction methods by means of the seizure prediction characteristic," *Physica D*, vol. 194, no. 3, pp. 357–368, 2004.
[47] R. Aschenbrenner-Scheibe, T. Maiwald, M. Winterhalder, H. U. Voss, J. Timmer, and A. Schulze-Bonhage, "How well can epileptic seizures be predicted? an evaluation of a nonlinear method," *Brain*, vol. 126, no. 12, pp. 2616–2626, 2003.
[48] Y. Zheng, G. Wang, K. Li, G. Bao, and J. Wang, "Epileptic seizure prediction using phase synchronization based on bivariate empirical mode decomposition," *Clin. Neurophysiol.*, vol. 125, no. 6, pp. 1104–1111, 2013.
[49] H. Khammari and A. Anwar, "A spectral based forecasting tool of epileptic seizures," *IJCSI Int. J. Comput.*, 2012.
[50] K.-Q. Shen, C.-J. Ong, X.-P. Li, Z. Hui, and E. Wilder-Smith, "A feature selection method for multilevel mental fatigue EEG classification," *IEEE Trans. Biomed. Eng.*, vol. 54, no. 7, pp. 1231–1237, 2007.
[51] K. K. Parhi and Z. Zhang, "Method and apparatus for prediction and detection of seizure activity," U.S. Patent Application Number 14/121,426, Sep. 5, 2014.

Zisheng Zhang (S'10) received the B.S. degree in electrical engineering from Shanghai Jiaotong University, Shanghai, China, in 2010.

Currently, he is working toward the Ph.D. degree in the Department of Electrical and Computer Engineering, University of Minnesota-Twin Cities, Minneapolis, MN, USA. His research interests include digital signal processing, seizure detection, and seizure prediction.

Keshab K. Parhi (S'85–M'88–SM'91–F'96) received the B.Tech. degree from IIT Kharagpur, Kharagpur, India, in 1982, the M.S.E.E. degree from the University of Pennsylvania, Philadelphia, PA, USA, in 1984, and the Ph.D. degree from the University of California, Berkeley, Berkeley, CA, USA, in 1988.

He has been with the University of Minnesota, Minneapolis, MN, USA, since 1988, where he is currently a Distinguished McKnight University Professor and an Edgar F. Johnson Professor with the Department of Electrical and Computer Engineering. He has authored more than 550 papers, is the inventor or coinventor of 29 patents, has authored the textbook, VLSI Digital Signal Processing Systems (New York, NY, USA: Wiley, 1999), and coedited the reference book, Digital Signal Processing for Multimedia Systems (Boca Raton, FL, USA: CRC Press, 1999). His current research interests include the VLSI architecture design and implementation of signal processing, communications and biomedical systems, error control coders and cryptography architectures, high-speed transceivers, stochastic computing, secure computing, and molecular computing. He is also currently working on intelligent classification of biomedical signals and images for applications such as seizure prediction and detection, schizophrenia classification, biomarkers for mental disorders, brain connectivity, and diabetic retinopathy screening.

Dr. Parhi has served on the editorial boards of the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS, PART I AND PART II, IEEE TRANSACTIONS ON VERY LARGE SCALE INTEGRATION (VLSI) SYSTEMS, IEEE TRANSACTIONS ON SIGNAL PROCESSING, IEEE SIGNAL PROCESSING LETTERS, and IEEE SIGNAL PROCESSING MAGAZINE, and served as the Editor-in-Chief of IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS, PART I from 2004 to 2005. He currently serves on the editorial board of the *Journal of Signal Processing Systems*. He has served as the Technical Program Cochair of the 1995 IEEE VLSI Signal Processing Workshop and the 1996 Application Specific Systems, Architectures, and Processors Conference, and as the General Chair of the 2002 IEEE Workshop on Signal Processing Systems. He was a Distinguished Lecturer of the IEEE Circuits and Systems Society from 1996 to 1998. He served as a Board of Governors Elected Member of the IEEE Circuits and Systems Society from 2005 to 2007. He is a recipient of numerous awards, including the 2013 Distinguished Alumnus Award from IIT Kharagpur, the 2013 Graduate/Professional Teaching Award from the University of Minnesota, the 2012 Charles A. Desoer Technical Achievement Award from the IEEE Circuits and Systems Society, the 2004 F. E. Terman Award from the American Society of Engineering Education, the 2003 IEEE Kiyo Tomiyasu Technical Field Award, the 2001 IEEE W. R. G. Baker Prize Paper Award, and the Golden Jubilee Medal from the IEEE Circuits and Systems Society in 2000.