Extracting and Selecting Distinctive EEG Features for Efficient Epileptic Seizure Prediction

Ning Wang, *Member, IEEE*, and Michael R. Lyu, *Fellow, IEEE*

***Abstract*—This paper presents compact yet comprehensive feature representations for the electroencephalogram (EEG) signal to achieve efficient epileptic seizure prediction performance. The initial EEG feature vectors are formed by acquiring the dominant amplitude and frequency components on an epoch-by-epoch basis from the EEG signals. These extracted parameters can reveal the intrinsic EEG signal changes as well as the underlying stage transitions. To improve the efficacy of feature extraction, an elimination-based feature selection method has been applied on the initial feature vectors. This diminishes redundant and noisy points, providing each patient with a lower dimensional and independent final feature form. In this context, our study is distinguished from that of others currently prevailing. Usually, these latter approaches adopted feature extraction processes, which employed time-consuming high-dimensional parameter sets. Machine learning approaches that are considered as state of the art have been employed to build patient-specific binary classifiers that can divide the extracted feature parameters into preictal and interictal groups. Through out-of-sample evaluation on the intracranial EEG recordings provided by the publicly available Freiburg dataset, promising prediction performance has been attained. Specifically, we have achieved 98.8% sensitivity results on the 19 patients included in our experiment, where only one of 83 seizures across all patients was not predicted. To make this investigation more comprehensive, we have conducted extensive comparative studies with other recently published competing approaches, in which the advantages of our method are highlighted.**

***Index Terms*—Amplitude and frequency modulation features, electroencephalogram (EEG) signal representation, epileptic seizure prediction, feature selection.**

I. Introduction

A. Epilepsy

EPILEPSY affects around 1% of the world's population [1]. This neurological disease is caused by sudden disturbances of the brain function [2]. Prominent characteristics of epilepsy are recurring seizures. During a seizure cycle, stage transitions take place. The main stages are preictal—the period of time before the seizure onset, ictal—the interval during which the seizure occurs, postictal—the period immediately succeed the end of a seizure, and interictal—the time between two consecutive seizures [3]. However, transitions from the interictal stage to the ictal stage turn out to be not abrupt [4], [5]. Recent studies also indicate the existence of premonitory symptoms for a majority of epilepsy patients [6], [7]. Experienced neurologists can forecast upcoming seizures by inspecting changes in EEG recordings [8]. The typical attributes of an EEG signal are its rhythmic activity and transients [9]. An example of seizure stage transitions conveyed by EEG signals is given in Fig. 1.

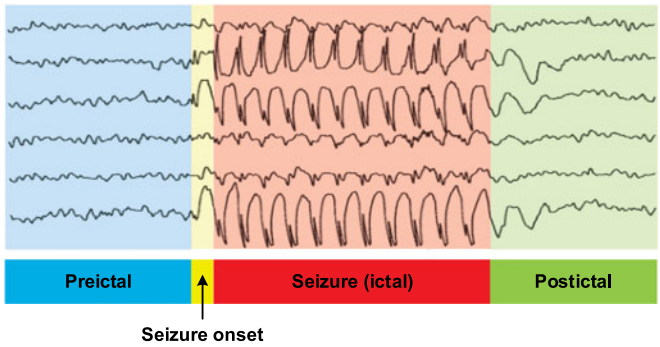


Fig. 1. Preictal–postictal stage transition in an epileptic seizure cycle manifested by multiple-channel EEG signals.

B. Seizure Prediction

Epileptic seizure event detection is concerned more with how to accurately detect seizure occurrence, without considering how late they were reported for medical treatment [10], while seizure onset detection algorithms expect to make the earliest possible alarms once a seizure takes place [11]. An even more challenging seizure diagnosis task is prediction, which aims to forecast impending seizure onset. A perfect epileptic seizure predictor can make predictions with the highest possible sensitivity as well as the least false alarms, and give warnings as early as possible for urgent actions to be taken. The period of time during which a seizure is supposed to occur is defined as seizure occurrence period (SOP), while the time interval after an alarm but before the SOP is called seizure prediction horizon (SPH) [12]. The SPH can last from a few minutes to hours. This minimum window of time between an alarm and the beginning of the SOP is essential for rendering a therapeutic intervention or a behavioral adjustment [13]. Once a seizure occurs within the SOP, the prediction is regarded successful, otherwise, it is determined as a false alarm.

Manuscript received February 26, 2014; revised June 8, 2014 and September 10, 2014; accepted September 11, 2014. Date of publication September 17, 2014; date of current version September 1, 2015. The work described in this paper was fully supported by the National Grand Fundamental Research 973 Program of China (No. 2014CB340401 and No. 2014CB340405), the Research Grants Council of the Hong Kong Special Administrative Region, China (No. 415212 of the General Research Fund), the Royal Society International Exchange grant IE130681, and the European Union Seventh Framework Programme (FP7/2007–2013) grant 288899.

N. Wang is with the Shenzhen Key Laboratory of Rich Media Big Data Analytics and Applications, Shenzhen Research Institute, The Chinese University of Hong Kong, Hong Kong, China, and the School of Computing and Mathematics, Plymouth University, U.K. (e-mail: nwang@cse.cuhk.edu.hk).

M. R. Lyu is with the Shenzhen Key Laboratory of Rich Media Big Data Analytics and Applications, Shenzhen Research Institute, The Chinese University of Hong Kong, and the Department of Computer Science & Engineering, The Chinese University of Hong Kong, Hong Kong, China (e-mail: lyu@cse.cuhk.edu.hk).

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/JBHI.2014.2358640

Nowadays, machine learning techniques are state of the art approaches to perform seizure diagnosis tasks. Among them, artificial neural networks [14], [15], Gaussian mixture models [16], Adaboost [17], and support vector machines (SVMs) [18], [19] have been employed widely. Under the machine-learning-based framework, binary classifiers are trained to decide whether a feature vector belongs to a preictal stage or an interictal stage. Regarding feature extraction, currently reported EEG features for seizure prediction tasks mostly consist of a large number of parameters. For example, in [11] and [20], each 6-s EEG epoch was characterized by a total of 432 spectral parameters, and in [14] and [21], 6300 dimensional bivariate feature vectors were generated for each 5-min time span. Besides these, the power spectra of time- and space-differential EEG signals [22], spatio-temporal correlation structures of multichannel EEG data [18], wavelet coefficients [23], [24] and autoregression parameters [25] were also taken as seizure indicators. Considering the sophisticated rhythmic activities and other physiological mechanisms underlying an EEG signal, a parametric representation for EEG signals is highly desirable. The amplitude–frequency modulation theory is found powerful in digging out dynamic mechanisms of narrow-band signals like speech resonances [26], and EEG rhythms [27]. Therefore, in our previous work on seizure prediction [28], we have made efforts to extract the most dominant amplitude and frequency characteristics of EEG epochs under the modulation theory to provide feature vectors, and have proven their efficacy in seizure prediction tasks.

Seizure prediction algorithms are always expected to be not only accurate but also employable in real-time scenarios. However, it is known that in biomedical applications involving a large amount of data, the number of training patterns is noticeably smaller than the dimensionality of the feature space, whereas the learning and test on these huge parameter sets usually consume substantial computing resources and greatly limit their engagement in real-world applications. On the other hand, it was found in a variety of data mining studies that a portion of features can be discarded without deteriorating the classification performance [29], [30]. Dimension reduction, which reduces the amount of features, speeds up the learning process, and enhances model comprehensibility, is therefore, highly desirable for various classifiers, including SVMs. One commonly used method is to project the data onto their first few principal directions, thereby producing new features, which are linear combinations of the original features. Methods in this class, like principal component analysis have been applied, providing more compact seizure prediction feature vectors [15], [18]. One disadvantage of the projection methods, however, is that no original feature parameters can be discarded. A pruning-technique-based feature selection method is another important technique for dimension reduction, in which process a subset of original features will be selected according to relevant criteria. In this paper, to further refine the features we have proposed in [28], and to improve the efficiency of preictal/interictal feature classification models in a diagnosis task in discovering new knowledge and making intelligent decisions, we consider attaching a pruning-based feature selection step to the original feature extraction front-end. This type of method, which can eliminate irrelevant or noisy input features and retain a small subset of relevant features, has achieved success in many application domains, such as gene selection [31], EEG channel selection [32], and variable selection for epilepsy diagnostics [17], [21], [33], [34]. One prominent method in this category is recursive feature elimination (RFE) [35], where the main principle is to include initially all attribute points about a specific subject, and to gradually remove those nondiscriminative points identified in the classification process, and finally, retain a core set of distinctive features.

C. Research Highlights

In this paper, we make efforts to explore efficient and compact EEG feature representations suitable for epilepsy prediction by employing an integrated framework. At first, a feature extractor is engaged that can decompose an EEG signal according to its frequency contents and identify the most dominant amplitude and frequency quantities therein. Second, the attached feature selection module will discriminate and discard the less relevant parameters from others in these extracted short-term amplitude–frequency components. Eventually, the remaining parameters constitute representational vectors of EEG features. In order to examine the efficacy of the proposed framework, we demonstrate experimentally that the parameters produced in this way yield superior preictal/interictal classification performance and are physically meaningful for the seizure prediction task. In summary, the main contributions of this paper are highlighted as follows.

1) Introduction of an analytical signal representation to extract primary EEG amplitude and frequency components.
2) Application of an integrated framework to produce a systematic feature extraction, selection, integration, and intelligent classification working program.
3) Derivation of comprehensive yet compact patient-specific feature forms to capture distinctive seizure-related EEG parameters.
4) Comparative studies with state of the art competing seizure prediction approaches to provide a comprehensive insight for researchers in relevant domains.

II. Seizure-Related Parameters

In this section, we describe the process of generating seizure-related feature parameters from EEG signals.

A. Modulation Feature Derivation

Conventional studies on rhythmic characteristics for seizure diagnostic purposes are concerned more with extracting signal energy and magnitude [11], [23], [36], rather than on the underlying amplitude–frequency patterns in the EEG rhythms. An analytic representation for EEG rhythms, on the other hand, cares more about the inclusive components contained in the signal. In a *monocomponent* amplitude and frequency modulating (AM–FM) signal like an EEG rhythm, the two determining parameters are the amplitude and frequency. Therefore, the kth rhythm $s_k(n)$ in an EEG signal $s(n)$ could be formulated as the product of AM and FM terms as follows:

$$s_k(n) = A_k(n)cos\big[\Theta_k(n)\big] \tag{1}$$

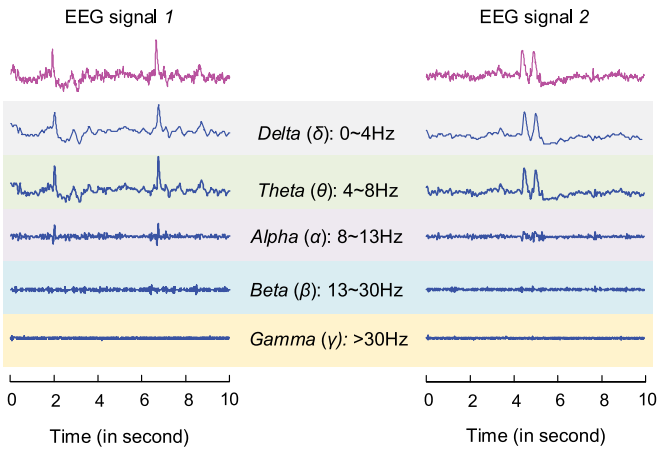


Fig. 2. Manifestation of rhythms in two consecutive 10-s-long EEG segments.

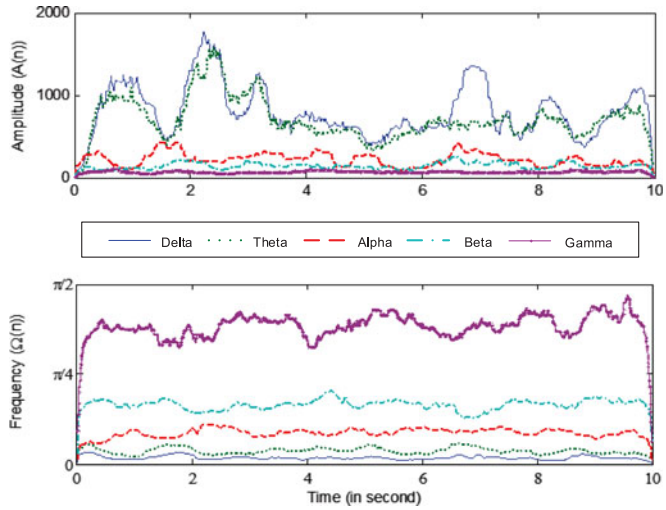


Fig. 3. Instantaneous amplitude estimate $A(n)$ and frequency estimate $\Omega(n)$ in detected EEG rhythms of the first 10-s-long EEG signal.

with the EEG rhythm being characterized by two time-varying sequences

1) $A_k(n)$: amplitude;
2) $\Theta_k(n)$: phase.

Teagers explored speech resonance characteristics through employment of a *multicomponent* AM–FM model [37]. Similarly, a multicomponent EEG signal can be represented as a summation of AM–FM components as follows:

$$s(n) = \sum_{k=1}^{K} A_k(n)\cos[\Theta_k(n)] + \eta(n)$$
$$= \sum_{k=1}^{K} A_k(n)\cos\left\{\left[\Omega_c(k)n + \sum_{r=1}^{n} q_k(r)\right]\right\} + \eta(n). \quad (2)$$

In this formulation, $A_k(n)$ and $\Theta_k(n)$ denote the instantaneous amplitude and phase of the kth primary component, respectively. By taking backward difference between $\Theta_k(n)$ and $\Theta_k(n-1)$, the instantaneous frequency sequence can be defined as $\Omega_k(n) = \Omega_c(k) + q_k(n) = \frac{2\pi}{f_s} f_c(k) + q_k(n)$, where f_s is the sampling frequency, and $q_k(n)$ is the frequency modulation component. $\Omega_c(k) = \frac{2\pi}{f_s} f_c(k)$ is the angular center frequency of the kth AM-FM subband signal. $\eta(n)$ stands for additive noise and errors of the modeling.

B. Feature Extraction

We first inspect EEG characteristics by observing EEG recordings from epilepsy patients. Two consecutive 10-s-long EEG segments from the first preictal stage of Patient 1 in the Freiburg database [38] are shown in Fig. 2. In this figure, the EEG rhythms detected through a bank of filters have also been illustrated. In Fig. 3, instantaneous amplitude and frequency estimates of these subbands from the first 10-s-long EEG signal have been shown. We can see the decomposed signals in Delta and Theta bands are of high amplitude, while the others are of relatively small value. The estimates of instantaneous frequency vary widely. However, the amplitude and frequency of each subband are still dominated by a primary component, respectively.

TABLE I
FEATURE EXTRACTION PARAMETERS AND VARIABLES

Parameter	Value	Description
n_s	1280	No. of samples per 5-s epoch
n_c	6	No. of EEG channels
K	5	No. of subbands in signal decomposition
n_e	60	No. of epochs contained by an 5-min AIE or AIF vector
d_i	600	Dimension of AIEF_{Ini} vector
Variable	**Range**	**Description**
α	$10 \sim 270$	No. of AIE parameters contained by an AIEF_{Ref} vector
β	$10 \sim 270$	No. of AIF parameters contained by an AIEF_{Ref} vector
$d_r = \alpha + \beta$	100, 200, or 300	Dimension of AIEF_{Ref} vector; being patient specific

We are, therefore, motivated to estimate these primary characteristics as initial EEG features. Notations of the parameters and variables engaged in producing the related feature vectors and their values/ranges are provided in Table I.

By applying the multiband AM–FM model on the EEG signal, two sets of characteristic sequences: the averaged instantaneous envelope (AIE) and averaged instantaneous frequency (AIF) parameters have been produced as follows.

1) *Signal Segmentation:* EEG signal in each channel segmented into nonoverlapping 5-s epochs.
2) *Signal Decomposition:* 48th ordered finite impulse response (FIR) filter bank employed to divide each epoch into $K = 5$ subbands: *delta* (0–4 Hz), *theta* (4–8 Hz), *alpha* (8–13 Hz), *beta* (13–30 Hz), and *gamma* (>30 Hz).
3) *Multiband Demodulation:* Energy separation algorithm proposed in [39] applied to obtain instantaneous envelope (IE) sequence $|A(n)|$ and instantaneous angular frequency (IF) $\Omega(n)$ one epoch by another for each subband signal.
4) *Sequence Smoothing:* 21-point median filter applied to remove abrupt impulses in IE and IF sequences.

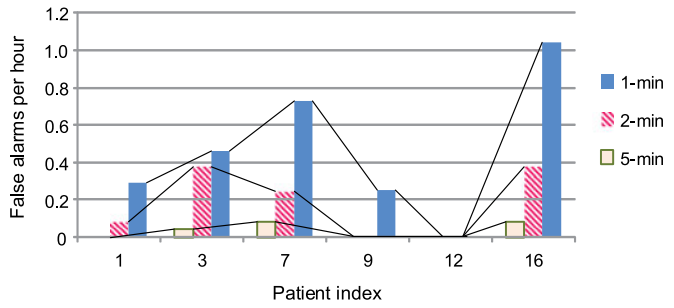


Fig. 4. False alarm rates obtained from AIEF feature vectors with various durations: 1 min, 2 min, and 5 min.

5) *Spatio-Temporal Averaging:* Two-step calculation for feature generalization.
   a) *Temporal averaging:* Averaging operation undertaken on smoothed IE and IF sequences to remove fluctuations over time. For each subband, a mean parameter is produced to represent the IE and IF sequences, respectively. The number of dimensions goes from $(K \times n_s \times n_c)$ to $(K \times n_c)$. (Notations explained in Table I.)
   b) *Spatial averaging:* Temporal IE and IF mean values averaged across different channels to compensate for possible channel variability. As a result, the IE and IF mean vectors across all channels are downsized to one, respectively. Therefore, for each EEG epoch, the dimension of IE/IF feature vectors is equal to the subband number K.

C. Initial Feature Form

The parameter sets AIE and AIF are generated from nonoverlapping 5-s-long epochs of all five EEG subbands. Two aggregation procedures are taken on these 5-D vectors to acquire more distinct and comprehensive feature forms. In the first step, we incorporate a pair of AIE and AIF vectors one after another to form an AIEF vector, whose dimension is, therefore, equal to the sum of the individual ones. As suggested in [21] to concatenate consecutive chunks of EEG features sequentially in time to provide longer decision interval, in the second step, we search for a suitable concatenation degree for AIEF vectors in a heuristic way. For this purpose, the 5-s-long AIEF vectors are temporally integrated into 1-min, 2-min, and 5-min vector forms to test their seizure prediction performance. In Fig. 4, the false alarm results on six patients from these newly generated AIEF features are shown. In this pilot study, the 5-min AIEF feature vectors are found more discriminative than the others. Therefore, we take 5-min durations for all AIE, AIF, and AIEF features in the following evaluation studies.

III. Seizure Prediction Features

In order to provide a compact feature form for the epilepsy prediction task, this section describes the process we take to screen the most discriminative parameters out of the feature vectors produced in Section II.

A. Recursive Feature Elimination (RFE)

Machine-learning-based predictors trained from input features often degrade when learning from many irrelevant parameters, which will finally deteriorate the performance of seizure prediction results on unseen trials [35]. Therefore, screening the input parameters for a more compact and discriminative subset of features before performing machine learning tasks is of critical importance to obtain predictors with reliable generalization performance.

In pattern recognition tasks, feature selection schemes are combined with or take part in the feature extraction process in order to reduce redundancy and dimensionality of the feature vectors. In our feature extraction front end, for an AIEF vector of 5-min duration, the number of dimensions is 600. Considering that the AIEF features may contain some irrelevant or redundant information, and also considering the variability among patients, we investigate performing elimination-based feature selection on a patient-by-patient basis. Feature selection schemes generally fall into three categories: filters, wrappers, and embedded methods [35], [40]. Filter methods like sequential forward selection and backward selection (SFS/SBS) are independent of classifiers involved in the classification tasks, while wrapper methods like RFE incorporate feature selection as parts of the training process to discard irrelevant features [35]. In biomedical/biological works such as gene selection, RFE was widely applied to combine with classifiers like SVM to provide effective classification [31]. In this paper, we consider an approach of RFE-SVM to eliminate noisy features, as well as to take advantage of the discrimination power of SVM classifiers. The basic philosophy of the RFE is to include initially all characteristic points about a specific subject, and to gradually exclude points that do not contribute in discriminating patterns from different classes. Whether a parameter in the current feature set contributes enough to be retained is dependent on its weight value resulted from training a classifier with the current set of features. Feature elimination usually progresses gradually and includes cross-validation steps. For each patient, in the leave-one-record-out cross-validation process, the feature evaluation and selection conducted in each validation iteration is independent of others. The feature set that produces the best overall performance among all iterations will be retained. The implementation of the RFE is carried out with the Spider toolbox for MATLAB [41].

B. Feature Selection Strategy

According to Section II-C, an AIEF feature vector contains the short-term AIE and AIF parameter sets first concatenated sequentially with a ratio of 1:1, and then, aggregated into vectors spanning over a continuous 5-min EEG duration along the time line. Under this task, the candidate feature forms will have to consider the following three factors in this patient-specific selection process: the feature specification, the feature dimension, and the time span coverage. Under a machine-learning-based seizure prediction framework, these factors are expected to result in a threefold impact on the system: first, the dimensionality can be reduced; second, the feature specifications can be patient

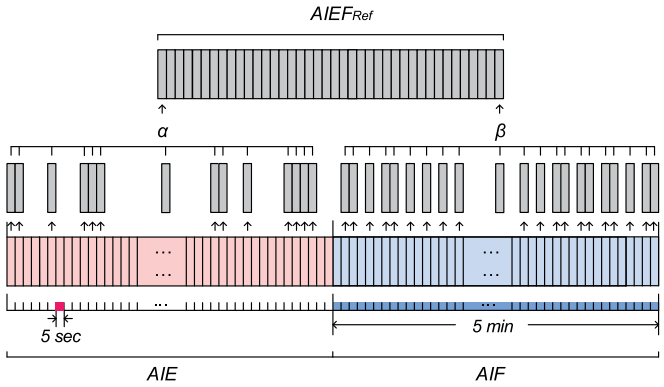


Fig. 5. Feature selection scheme of forming an AIEF_{Ref} vector from corresponding AIE, AIF parameter sets. For a 5-min long EEG signal, an AIEF_{Ini} vector is formed by concatenating 5-s AIE and AIF sets sequentially together. These initial AIE/AIF parameters are subject to individual screening process that ends up with α AIE parameters and β AIF parameters, respectively, and determines the form of an $\text{AIEF}_{[\alpha,\beta]}$ feature vector. $\text{AIEF}_{[\alpha,\beta]}$ with the highest performance in the prediction test will be picked out as the final form, noted as AIEF_{Ref}.

specific; and third, the feature selection process can be automatically performed.

Fig. 5 shows the process of selecting parameters from the entire AIE and AIF sets to form a new AIEF feature vector. To avoid ambiguity, we denote the *initial* 600-dimensional AIEF vector as AIEF_{Ini}, and this *refined* version as AIEF_{Ref}. A single AIE or AIF vector contains five parameters each, and is extracted from a continuous 5-s long EEG segment. As indicated by Section II-C, these 5-s vectors are concatenated sequentially to form a 300-dimension 5-min AIE or AIF vector. Parameters extracted from these two information sources are very different in magnitude; therefore, to ensure balance between these two sources, we are motivated to perform a screening algorithm on them separately. Two independent parameters α and β are engaged to control the amount of AIE and AIF samples that retain after the elimination process, respectively. As a result, the AIEF_{Ref} vector will contain α parameters from the 5-min AIE sequence and β parameters from the 5-min AIF sequence. The dimension d of AIEF_{Ref} vector is, therefore, equal to $\alpha+\beta$. The AIEF_{Ref} vector delivers AIE + AIF information in a certain period, say, k minutes of EEG data. The screening results largely depend on the choice of α and β. As a preliminary exploration, we set the following rules to decide these two variables:

1) $\alpha:\beta$ (feature specification) $= 1:9 \sim 5:5 \sim 9:1$;
2) $\alpha+\beta$ (feature dimension) $= 100$, 200, or 300;
3) k (time span) $= 10$ s ~ 5 min.

According to these rules, there are in total 27 sets of different $[\alpha,\beta]$ explored in this study. The newly generated $\text{AIEF}_{[\alpha,\beta]}$ features from each $[\alpha,\beta]$ set are evaluated under seizure prediction protocol to compete their performance. To find out the best-performed $\text{AIEF}_{[\alpha,\beta]}$ feature form for each patient, we take an experimental method, which contains three steps: 1) screening: to retain α AIE and β AIF parameters by discarding those less relevant points in the original AIE and AIF feature sets, respectively; 2) integration: to combine the retained α AIE points and β AIF points with a specific $\alpha:\beta$ to produce new vectors $\text{AIEF}_{[\alpha,\beta]}$; and (3) selection: to pick out the $\text{AIEF}_{[\alpha,\beta]}$ feature set that outperforms others as the final form, namely AIEF_{Ref}.

The time span k of a final form AIEF_{Ref} vector may vary from 10 s to 5 min. In the case of $\alpha:\beta = 1:9$ and $\alpha+\beta = 100$, ten parameters are from AIE sequence. If it happens that all these ten AIE parameters come from two consecutive AIE vectors, then they actually convey the whole AIE information for a 10-s EEG signal. The AIEF_{Ref} may convey the information from the entire 5-min EEG interval in other cases. By varying weights α and β following the described rules, a variety of cases may be approached. Each subject can also have their own feature forms, considering the feature selection and integration scheme is a patient-specific process. The variables involved in this feature selection process are also specified in Table I.

IV. Prediction Experimental Paradigm

A. Database

The investigated Freiburg EEG database [38] is a popular epileptic seizure dataset. It is a publicly available intracranial EEG data set, which contains invasive EEG recordings of 21 patients suffering from medically intractable focal epilepsy. The dataset was recorded during an invasive presurgical epilepsy monitoring at the Epilepsy Center of the University Hospital of Freiburg, Germany. The EEG data were acquired using a Neurofile NT digital video EEG system with 128 channels, a 256-Hz sampling rate, and a 16-bit analogue-to-digital converter. For each of the patients, there are two sets of data that contain EEG signals from ictal and interictal stages, respectively. For prediction purposes, at least 50 min of preictal data were retained prior to each epileptic seizure. As for the interictal states, approximately 24 hour of EEG recordings without seizure activity were provided. At least 24 hour of continuous interictal recordings were available from 13 patients. For the remaining patients, interictal invasive EEG data consisting of less than 24 hour were joined together, so as to end up with at least 24 hour of interictal recordings per patient. For each patient, the recordings of three focal and three extra-focal electrode contacts were provided.

B. Automatic Seizure Prediction System Architecture

The physical monitoring of epileptic seizure is conducted through observing physiological data continuously. For each targeted subject, two models will be built: One is for the between-seizure state (i.e., interictal) and the other is for the preictal state immediately before an upcoming seizure. A subject-specific binary classification is conducted continuously to classify the input feature vectors into interictal or preictal groups [11], [42]. Once preseizure observations are found to last for a certain period, which is 5 min in this study, alarms are raised to clinical caregivers immediately. The binary classification scheme is implemented with the SVM, where nonlinear decision boundaries are generated to separate the data by using a radial basis function (RBF) kernel [28].

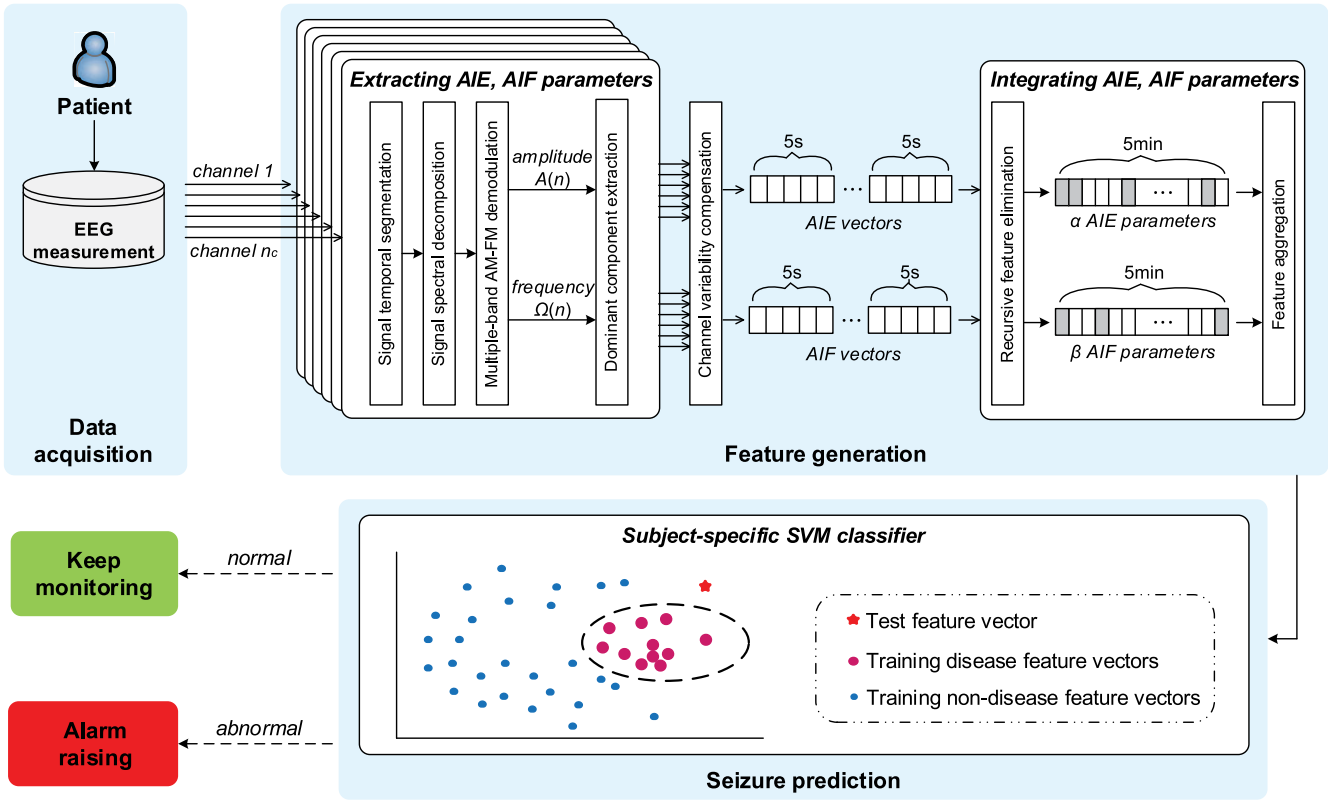


Fig. 6. Machine-learning-based disease diagnosis/monitoring system in a typical seizure prediction application. It mainly contains three modules: EEG data acquisition, feature generation that includes parameter extraction, selection and integration, and seizure prediction with SVM-based classification. This infrastructure provides necessary signal processing steps to handle various other physiological signals than EEG, such as electrocardiogram (ECG) and electromyogram (EMG). By replacing the SVM-based pattern recognition module with other machine learning tools when necessary, it is extendable to new application scenarios.

Fig. 6 gives an overall illustration of a machine-learning-based disease prediction experimental paradigm. The input data are n_c channel intracranial EEG measurements from the patients with epilepsy on a continuous-time basis. These time-varying sequences are then processed by a series of signal processing steps, producing respective health profiles of the concerned subject. Through implementing these sequentially connected procedures, which include temporal segmentation, spectral decomposition, and multiband demodulation on the EEG signals, their instantaneous amplitude and frequency sequences, $A(n)$ and $\Omega(n)$, respectively, are picked out. The most dominant amplitude and frequency components present in the n_c channel signals are extracted as intrinsic characteristics, where the variability across channels are then compensated, and these characteristics are aggregated as described in Section III-B to produce forms of seizure indicators. To provide subject-centric medical management, SVM classifiers are trained separately for each subject. A separation boundary is learned from past datasets of the subject regarding whether a seizure is impending or not. Once a set of unseen data samples arrive, the categories they should fall into will be decided accordingly. In case an upcoming seizure is forecasted, alarms will be raised to notify the concerned parties; otherwise, the system keeps monitoring the health condition of the respective subject.

C. Evaluation Methodology

We measured the performance of our seizure prediction framework in terms of sensitivity and specificity. In each test, the sensitivity is calculated as the percentage of successfully forecast seizures, while the specificity refers to the amount of false alarms occurred per hour. To estimate the seizure predictor's performance for each patient, a leave-one-record-out cross-validation scheme is employed. A record here indicates either a 50-min preictal recording or a 1-hour interictal recording in the Freiburg EEG dataset. To establish an optimal SVM classifier in training, five-fold cross validation is performed. Specifically, a grid search is executed in each cross-validation training round to seek optimal parameter set $[C, \gamma]$, where C is the SVM cost parameter and γ is the RBF kernel parameter. In this 21×21 grid search, the $\log_2 C$ and $\log_2 \gamma$ both vary between -10 and 10, respectively. The evaluation results are then generated by adopting these selected parameters in relevant tests.

In a disease detection test that screens people for a disease, each subject taking the test either has or does not have the disease. The test outcome could be positive or negative, which indicates that the subject is sick or not sick, respectively. The test results for each subject in this setting can be as follows.

1) *True positive (TP):* Sick people correctly diagnosed as sick.

TABLE II
PATIENT-SPECIFIC SEIZURE PREDICTION RESULTS WITH INITIAL FEATURE SETS

Patient Id.	Gender	Age	Seizure type	Seizure No.	Interictal Hr.	Sen. (%)	FA/hr
01	F	15	SP, CP	4	24	100	0.000
02	M	38	SP, CP, GTC	3	24	66.7	0.000
03	M	14	SP, CP	5	24	100	0.000
04	F	26	SP, CP, GTC	5	24	100	0.000
05	F	16	SP, CP, GTC	5	24	100	0.708
06	F	31	CP, GTC	3	24	100	0.042
07	F	42	SP, CP, GTC	3	25	100	0.000
09	M	44	CP, GTC	5	26	100	0.000
10	M	47	SP, CP, GTC	5	25	100	0.082
11	F	10	SP, CP, GTC	4	25	75.0	0.042
12	F	42	SP, CP, GTC	4	26	100	0.000
14	F	41	CP, GTC	4	25	100	0.294
15	M	31	SP, CP, GTC	4	25	100	0.200
16	F	50	SP, CP, GTC	5	24	100	0.208
17	M	28	SP, CP, GTC	5	25	100	0.083
18	F	25	SP, CP	5	27	100	0.232
19	F	28	SP, CP, GTC	4	25	100	0.041
20	M	33	SP, CP, GTC	5	25	60.0	0.284
21	M	13	SP, CP	5	26	100	0.401
Total						Mean	
19				83	473	95.2	0.138

2) *False positive/alarm (FP/FA):* Healthy people incorrectly identified as sick.
3) *True negative (TN):* Healthy people correctly identified as healthy.
4) *False negative (FN):* Sick people incorrectly identified as healthy.

The sensitivity, specificity, and the overall accuracy are calculated in the following manner:

$$\text{Sensitivity} = \frac{\sum \text{TP}}{\sum \text{TP} + \sum \text{FN}}$$

$$\text{Specificity} = \frac{\sum \text{TN}}{\sum FA + \sum \text{TN}}$$

$$\text{Accuracy} = \frac{\sum \text{TP} + \sum \text{TN}}{\sum \text{TP} + \sum \text{FA} + \sum \text{FN} + \sum \text{TN}}. \quad (3)$$

On the other hand, the specificity in disease prediction tasks is by usage referred to as $1 - \text{Specificity}$, namely, the smaller the better, while sensitivity indicates the same meaning as defined in (3). In general, for this type of task, the data falling into the two classes are typically unbalanced in number, which makes the overall accuracy not an ideal choice in this condition. Instead, F_β measure is considered suitable to serve the purpose. Being defined by the following equation, F_β measures binary classifiers from their TP, FN, and FA:

$$F_\beta = \frac{(1+\beta^2) \cdot \text{TP}}{(1+\beta^2) \cdot \text{TP} + \beta^2 \cdot \text{FN} + \text{FA}} \quad (4)$$

where β is a weighting factor. We set β to be 2 in this study.

V. PERFORMANCE

The previously described seizure prediction algorithm is examined on the Freiburg EEG dataset. 19 out of 21 patients with no less than 3 recorded seizures are involved in the experiments.

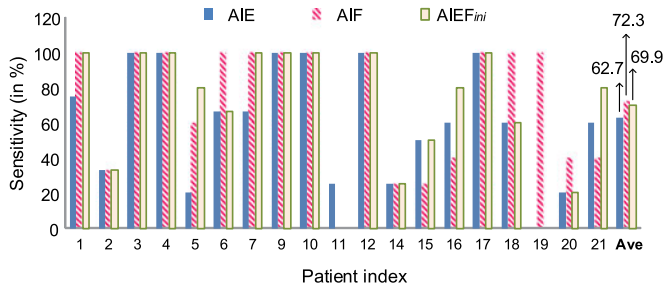


Fig. 7. Individual seizure prediction sensitivity performance by three sets of initially formed feature sets: AIE, AIF, and AIEF$_{Ini}$ (in %).

Relevant patient and EEG database characteristics are tabulated in Table II. To make the performance evaluation outcome more comprehensive, extensive comparisons with competing approaches published recently are conducted and shown in several aspects.

A. Initial Results

In evaluating the features, we have conducted two sets of parallel experiments, of which Exp$_{Acc}$ is targeted to maximize the overall accuracy, and Exp$_{F2}$ aims to optimize the F_2 measurement. These two sets of experiments are totally independent of each other. Fig. 7 illustrates the prediction sensitivity for all 19 patients and their averaged results for the AIE, AIF, and AIEF$_{Ini}$ parameters, respectively, under Exp$_{Acc}$ test protocol. It is observed that on average the AIF parameters demonstrate greater discriminative power than that of the AIE parameters. Besides, the aggregation of AIE and AIF parameter sets in the form of AIEF$_{Ini}$ vector cannot reflect and strengthen the advantage of AIE and AIF parameters in enhancing the prediction outcome. Another observation is that by employing F_2 measure

TABLE III
PATIENT-SPECIFIC SEIZURE PREDICTION RESULTS WITH OUR REFINED FEATURE VECTORS AND IN OTHER COMPETING APPROACHES

Feature Dim.	*Brown (in 2011)*		*Park (in 2011)*		*Ayinala (in 2012)*		*Williamson (in 2012)*		*Proposed*	
	120		1260		150		400		125	
Patient Id.	Sen. (%)	FA/hr	Sen. (%)	FA/hr	Sen. (%)	FA/hr	Sen. (%)	FA/hr	Sen. (%)	FA/hr
01	100	0.095	100	0.080	100	0	100	0.040	100	0
02	N/A	N/A	N/A	N/A	N/A	N/A	100	0	66.7	0
03	100	0.250	100	0	100	0	100	0	100	0
04	100	0.040	100	0.040	100	0	100	0	100	0
05	80.0	0.276	100	0.790	60.0	0.792	60.0	0.090	100	0.208
06	100	0.340	100	0.040	100	0.083	100	0.040	100	0
07	100	0.053	100	0.040	100	0	100	0.040	100	0
09	100	0.242	100	0.340	100	0.250	100	0	100	0
10	100	0.508	100	0.200	100	0.167	80.0	0	100	0.082
11	50.0	0.502	75.0	0.170	75.0	0	100	0.040	100	0.042
12	100	0.080	100	0.040	100	0	100	0.080	100	0
14	100	0.800	100	0.220	75.0	0.083	75.0	0.040	100	0.042
15	100	0.610	100	0.380	100	0.167	100	0.040	100	0
16	40.0	0.606	100	0.420	100	0.167	80.0	0	100	0
17	80.0	0.347	100	0	100	0.125	100	0	100	0.042
18	40.0	0.273	100	0.160	100	0.167	80.0	0	100	0.039
19	75.0	0.680	75.0	0.900	N/A	N/A	50.0	0.090	100	0.041
20	60.0	0.685	100	0.680	N/A	N/A	20.0	0.080	100	0.122
21	60.0	0.577	100	0.380	100	0.167	100	0.040	100	0.401
Average	81.3	0.386	97.5	0.271	91.5	0.135	85.5	0.033	98.8	0.054

rather than overall accuracy, most patients' performance can be evidently improved.

Table II provides detailed prediction results for all patients. The results recorded therein are from the best performing feature sets among AIE, AIF, and AIEF$_{Ini}$ parameter sets. On average, the overall sensitivity obtained across all patients is 95.2%, in which 79 out of 83 seizures in the evaluation set have been successfully predicted. Due to the fact that a significant amount of isolated positive detections turn out to be false alarms, a two-in-a-row filtering step is taken to diminish single positives, leaving only consecutive positive alarms (at least two) counted in the results. We end up achieving 0.138 FAs per hour specificity result with these sorts of features. These initial results were reported in our previous work [28].

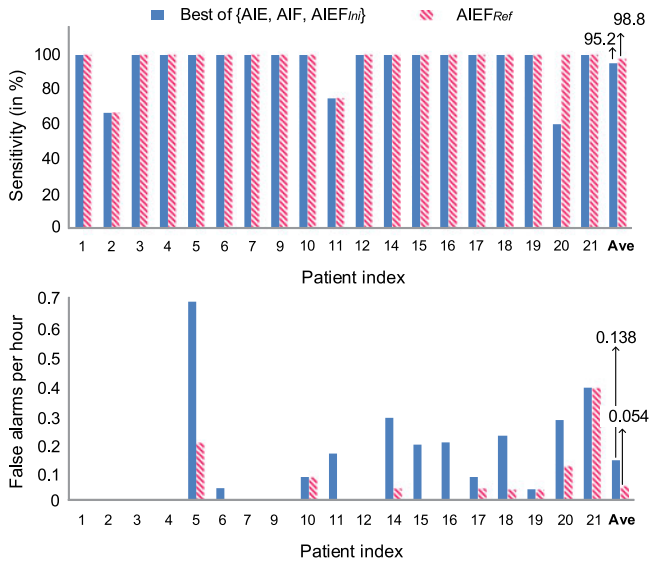


Fig. 8. Our current and previous sensitivity and specificity results, where previous results come from the best among {AIE, AIF, and AIEF$_{Ini}$}, and the current results are produced by the AIEF$_{Ref}$ features. Top: sensitivity rate in %; bottom: specificity rate in FA/hr.

B. Effects of Feature Selection

The final prediction outcomes in terms of sensitivity and specificity after the feature selection process are shown in Table III. In order to compare our approach with other competing approaches published recently, and to evaluate them under similar classification methods, Table III also includes their performance for the convenience of readers.

1) Comparison with our Previous Results: In Fig. 8, we compare our previous results recorded in Table II and the updated results in Table III. The patient-by-patient sensitivity and FA/h performance as well as their overall results are numerically noted. It is found that in terms of average sensitivity, the AIEF$_{Ref}$ features outperform the best performing feature set from {AIE, AIF, and AIEF$_{Ini}$} with a relative improvement of 3.78%, while the overall specificity in FA/h is greatly improved by 60.9%.

2) Comparison With Others: As indicated by Table III, Brown *et al.* in [43] have recently applied a divergence measurement-based feature selection process to retain a pool of 120 features, and finally, achieved an overall sensitivity of 81.3% and 0.386 FA/h as a result. Ayinala *et al.* [17], on the other hand, have employed Adaboost to make classification on 150-dimensional feature vectors and obtained a 91.5% sensitivity, 0.135 FA/h for a pool of 16 patients out of the 19 patients in our evaluation set. By contrast, the performance of Park's [36]

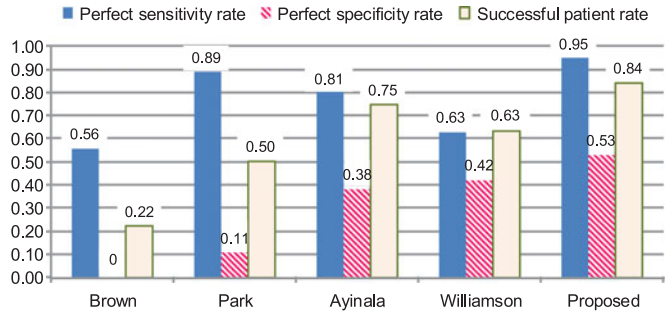


Fig. 9. Amount of successful prediction cases for all referred approaches. Perfect sensitivity/specificity rate indicates the percentage of patients for whom no sensitivity/specificity errors appear during the test; successful patient rate relates with the proportion of patients where the prediction is counted as a success.

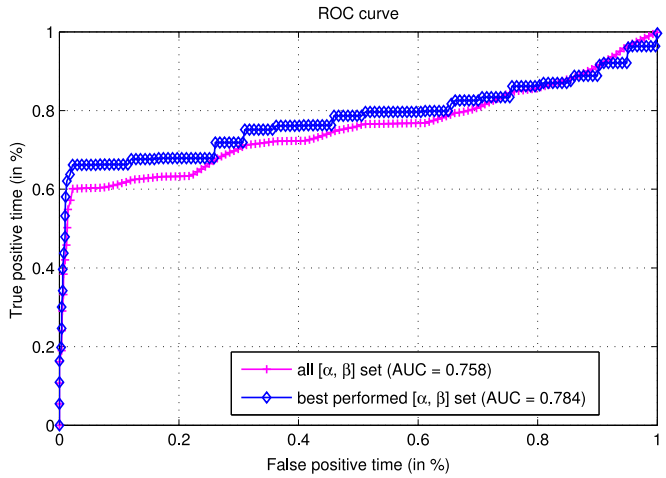


Fig. 10. ROC curve: TP time (in %) *versus* FP time (in %).

and Williamson's [18] prediction systems involve a much larger number of feature parameters than the aforementioned ones. As a result, Park *et al.* have produced an overall sensitivity as high as 97.5%, and Williamson *et al.* have reported a specificity result as low as 0.033 FA/h. However, neither of them have achieved a good balance between the prediction sensitivity and specificity, that is, producing satisfactory results in both terms simultaneously. In comparison, the automatic feature integration method proposed in this study not only produced feature vectors with lower dimension, but also demonstrated more proficient overall performance than other state of the art competing approaches. Although slight differences on evaluation procedures among approaches referred in Table III exist, the results are believed to have faithfully reflected the practical situation. Further inspections on these results from different perspectives will be addressed in Section V-C.

C. Results Analysis

1) Sensitivity and Specificity of Prediction: To make comprehensive inspections, we have discarded less test data than other approaches, e.g., we used all 473 interictal hours of EEG recordings involved in the Freiburg database to measure when false alarms might occur (Brown and Park uses 433.2 hours, and Williamson uses 448.3 hours), and we observed sensitivity on 83 marked seizures (Brown and Park uses 80 seizures, while Ayinala uses 71 seizures). As per results in Table III, we achieved perfect sensitivity, where all seizures have been successfully forecast beforehand for 18 out of all 19 patients. On average, the overall sensitivity obtained across all patients is 98.8%, in which 82 out of 83 seizures in the evaluation set have been correctly predicted. We achieved perfect specificity for 10 out of 19 patients, and this proportion is regarded as the largest among existing methods in Table III. Fig. 9 shows that for 95% patients we have achieved perfect sensitivity, and for 53% patients no false alarms have been reported during the whole interictal period. Our approach is clearly superior to the others.

2) Receiver Operating Characteristics Curve: Receiver operating characteristics (ROC) is one of the best-known indication of two-class discrimination in terms of TP rate as well as FP rate in percentage. It provides a good measurement on overall separability of the relevant classifier. With ROC statistics, a threshold is continuously updated to traverse all possible TP and FP on a 0–1 space. With a varying threshold, the variation of TP versus FP is recorded, and the resulting curve is termed as an ROC curve. To quantify the degree to which two groups can be discriminated by the classifier, the area under ROC curve (i.e., AUC) can be calculated. Varying between 0 and 1, the higher the value of AUC the better, where AUC $= 0.5$ means a random classification performance.

In this paper, we have evaluated our preictal/interictal classification algorithm by the ROC curve as well as AUC quantity as shown in Fig. 10. In this figure, the proportion of instances with TP detections versus those with FP detections is indicated. Two ROC curves have been given, which are both obtained from the summary results of all 19 patients. On a patient-by-patient basis, the best performing $[\alpha, \beta]$ sets in each patient's test are recorded, and their discrimination results accumulated together produce the curve with a $\diamond$ mark, while without a selection process, test results obtained from all sets mark the ROC curve with †. The AUC values indicate an improvement from 0.758 to 0.784.

3) Successful Patient Rate: Chiang *et al.* [32], [44] have proposed another evaluation scheme to compare the success of a seizure prediction algorithm, where the proportion of successful patients among all test populations is considered as an essential metric. If satisfying the following two conditions, a patient is counted as a successful patient.

1) False alarm rate per hour is less than 0.2.
2) Each single seizure has been predicted beforehand.

According to these criteria, the amount of successful patients for the referred approaches in Table III are shown in Fig. 9. It is found that the proposed method in this study achieved the highest successful patient rate, which is 16/19 $=$ 84.2%. It means, in a general scenario, our method can provide early warnings for all seizures that occur, and can report false alarms below a threshold rate, for 84.2% of the patients involved. The successful patient rate, rather than overall sensitivity or specificity

TABLE IV
PATIENT-SPECIFIC STATISTIC RESULTS ON FREQUENCY OF SUCCESS FOR THE REFERRED METHODS OF BROWN, PARK, AYINALA, WILLIAMSON, AND OUR PROPOSED APPROACH

	01	02	03	04	05	06	07	09	10	11	12	14	15	16	17	18	19	20	21
Brown	✓	*N/A*	*X*	✓	*X*	*X*	✓	*X*	*X*	*X*	✓	*X*	*X*	*X*	*X*	*X*	*X*	*X*	*X*
Park	✓	*N/A*	✓	✓	*X*	✓	✓	*X*	✓	*X*	✓	*X*	*X*	*X*	✓	✓	*X*	*X*	*X*
Ayinala	✓	*N/A*	✓	✓	*X*	✓	✓	*X*	✓	*X*	✓	*X*	✓	✓	✓	✓	*N/A*	*N/A*	✓
Williamson	✓	✓	✓	✓	*X*	✓	✓	✓	*X*	✓	✓	*X*	✓	*X*	✓	*X*	*X*	*X*	✓
Proposed	✓	*X*	✓	✓	*X*	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	*X*
Freq. of success	1.0	0.50	0.80	1.0	0	0.80	1.0	0.40	0.60	0.40	1.0	0.20	0.60	0.40	0.80	0.60	0.25	0.25	0.40
Hard patient?	No	Yes	No	No	Yes	No	No	Yes	No	Yes	No	Yes	No	Yes	No	No	Yes	Yes	Yes

would better indicate the predictor's performance in practical engagements.

D. Patient-Specific Study

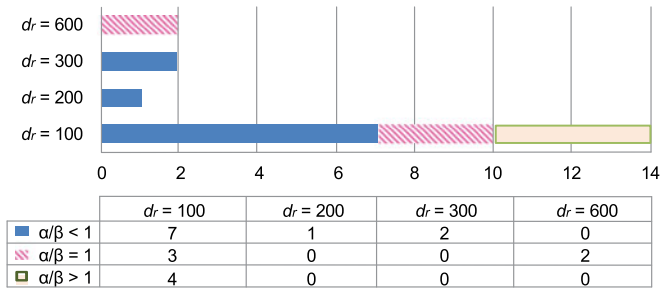
An ideal predictor is expected to produce a sensitivity as high as possible, and the lowest possible FAs measured in FA/h. By giving an observation along with all the evaluated patients, and across different approaches on their prediction outcome shown in Table III, it is obvious that for some patients, good performance results are always hard to obtain. According to the patient-specific records, we have made some statistics on the frequency of success across different approaches referred to in Table III, and the relevant observations are shown in Table IV.

In Table IV, the frequency of success for each patient is calculated as $\frac{\text{\# of successful predictions}}{\text{\# of predictions}}$. If a patient's frequency of success falls below 0.5, he/she is counted as a hard patient. Therefore, the following nine patients are found harder than others in making successful prediction: Patients 2, 5, 9, 11, 14, 16, 19, 20, and 21. Considering that 1) Patients 2, 19, and 20 are discarded from evaluation in some of the mentioned works, and 2) Patients 9, 11, 14, and 16 are considered successful in our approach, the typical hard patients with the poorest results turn out to be Patients 5 and 21.

E. Feature Selection Parameters

α and β variables control the screening of raw AIE and AIF parameters in constituting an $\text{AIEF}_{[\alpha,\beta]}$ vector. The $[\alpha, \beta]$ set that yields optimal performance determines the final form of AIEF_{Ref} for a certain subject. Considering the fact that the feature selection for AIE and AIF parameters are conducted separately, α and β together evidently affect the final prediction outcomes. With the dimension d_r of AIEF_{Ref} varying among patients, we have an average d_r equal to 125 calculated over all patients involved in the evaluation pool.

Fig. 11 shows a brief illustration on the experimentally determined $[\alpha, \beta]$ sets over the evaluation database. It is found that the majority of patients can rely on an amount of AIE/AIF parameters as small as 100 to achieve satisfactory prediction results. Only 2 out of the 19 patients engage the entire AIE and AIF parameter sets to provide final outcomes. To find out a suitable proportion of AIE and AIF parameters in feature selection, the ratios between α and β in the finally selected AIEF_{Ref} vectors are also noted. Sightly more than a half (10 out of 19) of



	dr = 100	dr = 200	dr = 300	dr = 600
α/β < 1	7	1	2	0
α/β = 1	3	0	0	2
α/β > 1	4	0	0	0

Fig. 11. Statistics on feature dimension d_r and ratio α/β in the final forms of AIEF_{Ref} feature vectors over all patients.

the patients have their β larger than α, that is, $\alpha/\beta < 1$, which is consistent with our previous observations on sole AIE and AIF vectors. These resultant statistics on feature selection parameters are expected to offer new observations on EEG signal feature derivation.

VI. CONCLUSION

Timely and accurate predictions of impending seizures are crucial for epilepsy patients and others caring for them. To carry out efficient analysis and prediction on seizure activities, distinct seizure-related representatives from EEG signals are highly desirable. Under a machine-learning-based binary classification protocol, this paper focuses on deriving compact yet comprehensive feature vectors from EEG signals to perform efficient prediction. We first present a new EEG representation to extract primary amplitude and frequency components underlying an EEG signal. We subsequently examine the relevance of these two sources of information with a patient's seizure occurrence. In order to diminish potential redundancy within the amplitude and frequency components, we execute a parameter screening process on them. Considering personal variety might exist, the feature selection process is carried out on a patient-specific basis. The screened feature vectors not only benefit from the discriminative power of the dominant EEG amplitude and frequency components in a flexible manner, but are proven more effective than the raw vectors. Through theoretical analysis and experimental evaluation on a standard EEG dataset, the proposed EEG features are found very promising in making timely predictions with very low false alarm rates. In comparison with state of the art methods under similar setup conditions, our

approach is identified as being much superior in several aspects including prediction sensitivity, prediction specificity, and successful patient rates. As a result, the presented feature extraction method leads to an overall sensitivity as high as 98.8%, and a false alarm rate as low as 0.054 FA/h. In particular, we have ended up with perfect (100%) sensitivity for 18 out of all the 19 patients included in our experiment, only missing one from 83 involved seizures across all patients. The study described in this paper is distinguished from most of the current seizure prediction approaches, which employ high-dimensional and exhaustive feature extraction processes. It is, therefore, expected to provide new perspectives in modern disease diagnostics. Considering the finding that simple narrowband features can provide good prediction performance, the investigation on combining the feature vectors proposed here with other conventional multivariate features might be one of the future directions.

REFERENCES

[1] E. Kandel, J. Schwartz, and T. Jessell, *Principles of Neural Science*. New York, NY, USA: McGraw-Hill, 2000.

[2] J. Engel, *Seizures and Epilepsy*, 2nd ed. New York, NY, USA: Oxford Univ. Press, 2013.

[3] "Types of seizures: What you need to know about different types and different stages of seizures," Epilepsy Foundation of America, Landover, MD, USA, 2009.

[4] H. Lange, J. Lieb, J. Engel, and P. Crandall, "Temporo-spatial patterns of pre-ictal spike activity in human temporal lobe epilepsy," *Electroencephalography Clinical Neurophysiol.*, vol. 56, no. 6, pp. 543–555, Dec. 1983.

[5] F. Bartolomei, F. Wendling, J. Régis, M. Gavaret, M. Guye, and P. Chauvel, "Pre-ictal synchronicity in limbic networks of mesial temporal lobe epilepsy," *Epilepsy Res.*, vol. 61, no. 1–3, pp. 89–104, Sep.-Oct. 2004.

[6] P. Rajna, B. Clemens, and E. Csibri "Hungarian multicentre epidemiologic study of the warning and initial symptoms (prodrome, aura) of epileptic seizures," *Seizure*, vol. 6 no. 5, pp. 361–368, Oct. 1997.

[7] A. Schulze-Bonhage, C. Kurth, A. Carius, B. Steinhoff, and T. Mayer, "Seizure anticipation by patients with focal and generalized epilepsy: a multicenter assessment of premonitory symptoms," *Epilepsy Res.*, vol. 70, no. 1, pp. 83–88, Jul. 2006.

[8] B. Litt and K. Lehnertz, "Seizure prediction and the preseizure period," *Current Opinion Neurol.*, vol. 15, no. 2, pp. 173–177, Apr. 2002.

[9] R. Rangayyan, *Biomedical Signal Analysis*. New York, NY, USA: Wiley, 2002, pp. 177–235.

[10] S. Ghosh-Dastidar, H. Adeli, and N. Dadmehr, "Mixed-band wavelet-chaos-neural network methodology for epilepsy and epileptic seizure detection," *IEEE Trans. Biomed. Eng.*, vol. 54, no. 9, pp. 1545–1551, Sep. 2007.

[11] A. Shoeb and J. Guttag, "Application of machine learning to epileptic seizure detection," in *Proc. Int. Conf. Mach. Learn.*, Haifa, Israel, 2010, pp. 975–982.

[12] B. Schelter, M. Winterhalder, T. Maiwald, A. Brandt, A. Schad, J. Timmer, and A. Schulze-Bonhage, "Do false predictions of seizure depend on the state of vigilance? A report from two seizure-prediction methods and proposed remedies," *Epilepsia*, vol. 47, no. 12, pp. 2058–2070, 2006.

[13] M. Winterhalder, T. Maiwald, H. Voss, R. Aschenbrenner-Scheibe, J. Timmer and A. Schulze-Bonhage, "The seizure prediction characteristic: A general framework to assess and compare seizure prediction methods," *Epilepsy Behavior*, vol. 4, pp. 318–325, 2003.

[14] P. Mirowski, Y. LeCun, D. Madhavan, and R. Kuzniecky, "Comparing SVM and convolutional networks for epileptic seizure prediction from intracranial EEG," in *Proc. Machine Learning for Signal Processing Workshop*, New York, NY, USA, 2008, pp. 244–249.

[15] S. Ghosh-Dastidar, H. Adeli, and N. Dadmehr, "Principal component analysis-enhanced cosine radial basis function neural network for robust epilepsy and seizure detection," *IEEE Trans. Biomed. Eng.*, vol. 55, no. 2, pp. 512–518, Feb. 2008.

[16] A. Zandi, G. Dumont, M. Javidan, and R. Tafreshi, "Epileptic seizure prediction using variational mixture of Gaussians," in *Proc. IEEE Eng. Med. Biol. Soc.*, 2011, pp. 7549–7552.

[17] M. Ayinala and K. Parhi, "Low complexity algorithm for seizure prediction using Adaboost," in *Proc. IEEE Eng. Med. Biol. Soc.*, 2012, pp. 1061–1064.

[18] J. Williamson, D. Bliss, D. Browne, and J. Narayanan, "Seizure prediction using EEG spatiotemporal correlation structure," *Epilepsy Behavior*, vol. 25, pp. 230–238, 2012.

[19] J. Yoo, L. Yan, D. El-Damak, M. Altaf, A. Shoeb, and A. Chandrakasan, "An 8-channel scalable EEG acquisition SoC with patient-specific seizure classification and recording processor," *IEEE J. Solid-State Circuits*, vol. 48, no. 1, pp. 214–228, Jan. 2013.

[20] A. Shoeb, A. Kharbouch, J. Soegaard, S. Schachter, and J. Guttag, "A machine-learning algorithm for detecting seizure termination in scalp EEG," *Epilepsy Behavior*, vol. 22, pp. S36–S43, Dec. 2011.

[21] P. Mirowski, D. Madhavan, Y. LeCun, and R. Kuzniecky, "Classification of patterns of EEG synchronization for seizure prediction," *Clinical Neurophysiol.*, vol. 120, no. 11, pp. 1927–1940, Nov. 2009.

[22] Y. Park "Reduced complexity epileptic seizure prediction with EEG," Ph.D. dissertation University of Minnesota, Minneapolis, MN, USA, Jan. 2012.

[23] P. Jahankhani, V. Kodogiannis, and K. Revett, "EEG signal classification using wavelet feature extraction and neural networks," in *Proc. IEEE John Vincent Atanasoff Int. Symp. Modern Comput.*, 2006, pp. 52–57.

[24] H. Adeli, S. Ghosh-Dastidar, and N. Dadmehr, "A wavelet-chaos methodology for analysis of EEGs and EEG subbands to detect seizure and epilepsy," *IEEE Trans. Biomed. Eng.*, vol. 54, no. 2, pp. 205–211, Feb. 2007.

[25] 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. Bio-med. Eng.*, vol. 57, no. 5, pp. 1124–1132, May 2010.

[26] A. Potamianos and P. Maragos, "Speech formant frequency and bandwidth tracking using multiband energy demodulation," *J. Acoustical Soc. Amer.*, vol. 99, no. 6, pp. 3795–3806, Jun. 1996.

[27] M. Díaz, J. Viola, and R. Esteller, "Analysis of instantaneous amplitude and frequency of intracranial EEG signal to characterize epileptic seizure stages," in *Proc. IEEE Eng. Med. Biol. Soc.*, Lyon, France, 2007, pp. 1290–1293.

[28] N. Wang and M. Lyu, "Exploration of instantaneous amplitude and frequency features for epileptic seizure prediction," in *Proc. IEEE Int. Conf. Bioinformat. Bioeng.*, 2012, pp. 292–297.

[29] A. Ng, "Feature selection, L1 versus L2 regularization, and rotational invariance," in *Proc. Int. Conf. Mach. Learn.*, 2004, p. 78.

[30] D. Donoho, "For most large underdetermined systems of linear equations, the minimal L1-norm solution is also the sparsest solution," *Comm. Pure Appl. Math.*, vol. 59, pp. 907–934, 2006.

[31] I. Guyon, J. Weston, and S. Barnhill, "Gene selection for cancer classification using support vector machines," *Mach. Learning*, vol. 46, pp. 389–422, 2002.

[32] N. Chang, T. Chen, C. Chiang, and L. Chen "Channel selection for epilepsy seizure prediction method based on machine learning," in *Proc. IEEE Eng. Med. Biol. Soc.*, 2012, pp. 5162–5165.

[33] K. Shen, C. Ong, X. 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, Jul. 2007.

[34] W. Kerr, A. Anderson, H. Xia, E. Braun, E. Lau, A. Cho and M. Cohen, "Parameter selection in mutual information-based feature selection in automated diagnosis of multiple epilepsies using scale EEG," in *Proc. Int. Workshop Pattern Recog. NeuroImag.*, 2012, pp. 45–48.

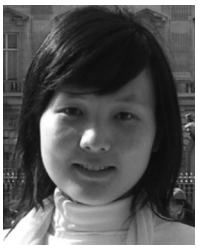
[35] I. Guyon and A. Elisseeff, "An introduction to variable and feature selection," *J. Mach. Learn. Res.*, pp. 1157–1182, 2003.

[36] Y. Park, L. Luo, 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.

[37] H. Teager, "Some observations on oral air flow during phonation," *IEEE Trans. Acoust. Speech*, vol. 28, no. 5, pp. 599–601, Oct. 1980.

[38] R. Andrzejak, K. Lehnertz, F. Mormann, C. Rieke, P. David, and C. Elger, "Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state," *Phys. Rev. E*, vol. 64, no. 6, pp. 061907, Nov. 2001.

[39] P. Maragos, J. Kaiser, and T. Quatieri, "Energy separation in signal modulations with application to speech analysis," *IEEE Trans. Signal Process.*, vol. 41, no. 10, pp. 3024–3051, Oct. 1993.
[40] R. Kohavi and G. John, "Wrappers for feature subset selection," *Artif. Intell.*, vol. 97, pp. 273–324, 1997.
[41] J. Weston, A. Elisseeff, G. Baklr, and F. Sinz. (2005). The spider machine learning toolbox. [Online]. Available: http://www.kyb.tuebingen.mpg.de/bs/people/spider.
[42] G. Orrù, W. Pettersson-Yeo, A. Marquand, and G. Sartori, "Using support vector machine to identify imaging biomarkers of neurological and psychiatric disease: A critical review," *Neurosci. Biobehavioral Rev.*, vol. 36, pp. 1140–1152, 2012.
[43] M. Brown, T. Netoff, and K. Parhi, "A low complexity seizure prediction algorithm," in *Proc. IEEE Eng. Med. Biol. Soc.*, 2011, pp. 1640–1643.
[44] C. Chiang, N. Chang, T. Chen, H. Chen, and L. Chen, "Seizure prediction based on classification of EEG synchronization patterns with on-line retraining and post-processing scheme," in *Proc. IEEE Eng. Med. Biol. Soc.*, 2011, pp. 7564–7569.



Ning Wang (M'11) received the B.Eng. degree in measurement and control technologies and devices from the College of Automation, Northwestern Polytechnical University, Xi'an, China, in 2005, the M.Phil. and Ph.D. degrees in electronic engineering from the Department of Electronic Engineering, Chinese University of Hong Kong, Hong Kong, in 2007 and 2011, respectively.

She was working as Postdoctoral Fellow at the Department of Computer Science & Engineering, Chinese University of Hong Kong, from 2011 to 2013, and she is currently with the School of Computing and Mathematics, Plymouth University, Plymouth, U.K.. Her research interests include signal processing and machine learning, with applications in robust speaker recognition, biomedical pattern recognition, intelligent data analysis, and human–robot interaction.

Michael Rung-Tsong Lyu (F'04) received the B.Sc. degree from National Taiwan University, Taipei, Taiwan, the M.Sc. degree from the University of California, Santa Barbara, CA, USA, and the Ph.D. degree in computer science from University of California, Los Angeles, CA.

He is currently a Professor at Computer Science and Engineering Department, The Chinese University of Hong Kong, Hong Kong. He worked at the Jet Propulsion Laboratory, University of Iowa, Bellcore, and Bell Laboratories. He has published 450 refereed journal and conference papers in these areas, which recorded 14500 Google Scholar citations and an H-index of 60. He has published two books: Software Fault Tolerance (New York, NY, USA: Wiley, 1995) and Handbook of Software Reliability Engineering (New York, NY, USA: IEEE and McGraw-Hill, 1996). His research interests include software reliability engineering, distributed systems, fault-tolerant computing, service computing, multimedia information retrieval, and machine learning.

Prof. Lyu received the Best Paper Awards of International Symposium on Software Reliability Engineering (ISSRE) 1993, ISSRE 1998, ISSRE 2003, Best Poster Paper Award of WWW 2002, ACM SigSoft Distinguished Paper Award of the International Conference on Software Engineering 2010, Best Student Paper Award of the International Conference on Web Services (ICWS) 2010, and Vannevar Bush Best Paper Award of the ACM/IEEE Joint Conference on Digital Libraries 2012. He initiated the first ISSRE in 1990. He was the Program Chair for ISSRE'96, Program Cochair for WWW10, General Chair for ISSRE'2001, General Cochair for Pacific Rim International Symposium on Dependable Computing 2005, General Chair for the International Conference on Dependable Systems and Networks 2011, and Program Cochair for ICWS2013. He served as an Associate Editor of the IEEE TRANSACTIONS ON RELIABILITY, IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, and *Journal of Information Science and Engineering*. He is currently on the editorial boards of IEEE TRANSACTIONS ON SERVICE COMPUTING and*Software Testing, Verification and Reliability Journal*. He became an AAAS Fellow in 2007 and Croucher Senior Research Fellow in 2008, and was named as the IEEE Reliability Society Engineer of the Year in 2010.