

Seizure Prediction by Graph Mining, Transfer Learning, and Transformation Learning

Nimit Dhulekar¹, Srinivas Nambirajan², Basak Oztan³, and Bülent Yener¹

¹ Computer Science Department, Rensselaer Polytechnic Institute, Troy, NY

² Department of Mathematical Sciences, Rensselaer Polytechnic Institute, Troy, NY

³ American Science and Engineering Inc, Billerica, MA

Abstract. We present in this study a novel approach to predicting EEG epileptic seizures: we accurately model and predict non-ictal cortical activity and use prediction errors as parameters that significantly distinguish ictal from non-ictal activity. We suppress seizure-related activity by modeling EEG signal acquisition as a cocktail party problem and obtaining seizure-related activity using Independent Component Analysis. Following recent studies intricately linking seizure to increased, widespread synchrony, we construct dynamic EEG synchronization graphs in which the electrodes are represented as nodes and the pair-wise correspondences between them are represented by edges. We extract 38 intuitive features from the synchronization graph as well as the original signal. From this, we use a rigorous method of feature selection to determine minimally redundant features that can describe the non-ictal EEG signal maximally. We learn a one-step forecast operator restricted to just these features, using autoregression (AR(1)). We improve this in a novel way by cross-learning common knowledge across patients and recordings using *Transfer Learning*, and devise a novel transformation to increase the efficiency of transfer learning. We declare imminent seizure based on detecting outliers in our prediction errors using a simple and intuitive method. Our median seizure detection time is 11.04 minutes prior to the labeled start of the seizure compared to a benchmark of 1.25 minutes prior, based on previous work on the topic. To the authors' best knowledge this is the first attempt to model seizure prediction in this manner, employing efficient seizure suppression, the use of synchronization graphs and transfer learning, among other novel applications.

1 Introduction

Epilepsy is one of the most common disorders of the central nervous system characterized by recurring seizures. An epileptic seizure is described by abnormally excessive or synchronous neuronal activity in the brain [24]. Epilepsy patients show no pathological signs of the disease during inter-seizure periods, however, the uncertainty with regards to the onset of the next seizure deeply affects the lives of the patients.

Seizure prediction refers to predicting the onset of epileptic seizures by analyzing electroencephalographic (EEG) recordings without any a priori knowledge of the exact temporal location of the seizure [22]. A method with the capacity to successfully predict the occurrence of an epileptic seizure would make it possible for the patient to be administered therapeutic treatments thereby alleviating the pain [21]. Many approaches have been suggested as possible seizure prediction algorithms, with modest levels of success. The first attempt at a seizure prediction algorithm was made by Viglione and Walsh in 1975 [71] and investigated spectral components and properties of EEG data. This was followed in the 80s decade by different groups attempting to apply linear approaches, in particular autoregressive modeling, to seizure prediction [59,60,62]. Moving forward 20 years, Mormann et al. posed the question whether characteristic features can be extracted from the continuous EEG signal that are predictive of an impending seizure [48]. In 2002, the First International Workshop on Seizure Prediction [41] was conducted to bring together experts from a wide range of background with the common goal of improving the understanding of seizures, and thus advancing the current state of seizure prediction algorithms on a joint data set [41].

Most general approaches to the seizure prediction problem share several common steps including (i) processing of multichannel EEG signals, (ii) discretization of the time series into fixed-size overlapping windows called epochs, (iii) extraction of frequency bands to analyze the signal in frequency and/or time domains using techniques such as wavelength transformation [1], (iv) extraction of linear and non-linear features from the signal or its transformations; these features can be univariate, computed on each EEG channel separately, or multivariate, computed between two

or more EEG channels, and (v) learning a model of the seizure statistics given the features by using supervised machine learning techniques such as Artificial Neural Networks [26,34,64,69,70], or Support Vector Machines [10,11]. A thorough survey of the various linear and non-linear features can be found in [23,25,57]. These features are usually calculated over epochs of predetermined time duration (around 20 seconds) via a moving window analysis. It has been found that univariate features, such as Lyapunov exponents, correlation dimension, and Hjorth parameters, calculated from the EEG recordings performed poorly as compared to bivariate and multivariate features [16,22,27,29,31,39,50]. This is understandable given that the seizure spreads to all the electrodes, whereas not all electrical activity in the brain may result in the onset of a seizure - it might be a localized discharge at a certain electrode. Although it has been shown that univariate features are less significant for seizure prediction, the importance of non-linear features over linear features is not quite as straightforward. It has recently been observed that non-linear techniques might not enhance the performance of the seizure prediction algorithm considerably over linear techniques, and also have considerable limitations with respect to computational complexity and description of epileptic events [14,15,43,49,51].

A phase-locking bivariate measure, which captures brainwave synchronization patterns, has been shown to be important in differentiating interictal from pre-ictal states [37,38]. In particular, it is suggested that the interictal period is characterized by moderate synchronization at large frequency bands while the pre-ictal period is marked by a decrease in the beta range synchronization between the epileptic focus and other brain areas, followed by a subsequent hypersynchronization at the time of the onset of the ictal period [47].

Many different approaches have been applied towards determining these features, such as frequency domain tools [8,53], wavelets [28,52], Markov processes [45], autoregressive models [3,11,68], and artificial neural networks [44]. If it were possible to reliably predict seizure occurrence then preventive clinical strategies would be replaced by patient specific proactive therapy such as resetting the brain by electrical or other methods of stimulation. While clinical studies show early indicators for a pre-seizure state including increased cerebral blood flow, heart rate change, the research in seizure prediction is still not reliable for clinical use.

Recently there has been an increased focus in analyzing multivariate complex systems such as EEG recordings using concepts from *network theory* [4,7,18,42,54,67], describing the topology of the multivariate time-series through *interaction networks*. The interaction networks enable characterization of the pair-wise correlations between electrodes using graph theoretical features over time [9,66]. In the spatio-temporal interaction networks, *nodes* (vertices) represent the EEG channels and the *edges* (links) represent the level of neuronal synchronization between the different regions of the brain. This approach has been exploited in the analysis of various neuropsychiatric diseases including schizophrenia, autism, dementia, and epilepsy [36,66,73]. Within epilepsy research, evolution of certain graph features over time revealed better understanding of the interactions of the brain regions and the seizures. For instance, Schindler et al. analyzed the change in path lengths and clustering coefficients to highlight the evolution of seizures on epileptic patients [61], Kramer et al. considered the evolution of local graph features including betweenness centrality to explain the coupling of brain signals at seizure onset [35], and Douw et al. recently showed epilepsy in glioma patients was attributed to the theta band activity in the brain [20]. In [46] authors independently suggest a similar approach that combines tensor decompositions with graph theory. Even with this significant body of research what remained unclear was whether the network-related-approach can adequately identify the inter-ictal to pre-ictal transition [36]. In this paper, we continue studying a form of interaction networks dubbed *synchronization graphs* [19] and introduce new features as the early indicators of a seizure onset, thereby identifying the inter-ictal to pre-ictal transition.

Summarily, the current approaches aim to develop features that are naturally characteristic of seizure activity. While these approaches are both intuitive and instructive, ictal activity is often a small portion of the available data, and statistical learning techniques, which require a large corpus of data for reliable prediction, can be expected to perform poorly as seizure-predictors. However, these techniques seem promising for accurate prediction of non-ictal activity with respect to which ictal activity may be identified as an anomaly. Provided that the only anomalous activity in the data is the seizure, or that other anomalies present with discernible signatures, this provides an equivalent method of predicting seizure. In general we operate under the paradigm that any feature or parameter that distinguishes between ictal and non-ictal activity is a mathematical characteristic of seizure, although it may not be a natural physiological indicator. We rigorously define the notion of what it means to be a good mathematical characteristic of a seizure, rate our seizure-discriminating parameter accordingly, systematically increase how well it discriminates between ictal and non-ictal activity, and qualify our predictions using such a discriminating parameter.

Furthermore, since cortical activity is continuously recorded as EEG signals, it can be represented as a time-series, and analyzed using time-series forecasting methods. The objective of time-series forecasting is to use equally spaced past and current observations to accurately predict future values. *Autoregressive* models (AR) are commonly used tools for time-series prediction, and have been used to capture the spatio-temporal properties of EEG signals [3,68]. We further improve on the AR model by using *Transfer Learning* [56] to learn the best forecast operator for a particular EEG recording from other EEG recordings. Transfer learning is a general form of learning such that there need not be any similarity in the distributions of the training and testing data. In our context, transfer learning does not require the past values and future values of the output variable to be correlated. In addition, transfer learning is particularly useful when data is only partially available or corrupted by noise, where such data can be effectively supplemented by clean data from a different experiment. We further improve on the transfer learning by modifying the transfer set into the most similar form to the dataset being investigated by means of a simple transformation (based on the Procrustes problem [72]).

The three main contributions of this work are as follows: (i) Formulating seizure prediction as a problem of anomaly detection and developing a discriminating parameter for the anomaly (ii) bridging the two concepts of AR modeling and interaction graphs by constructing an AR(1) model on the features extracted from the time-evolving EEG synchronization graphs (as well as other features obtained from the EEG signal itself), and (iii) introducing the concepts of transfer and transformation learning to improve the predictions of the AR(1) model.

The organization of the paper is as follows: in Section 2, we describe our methodology starting with the epileptic EEG dataset, initial noise removal, and procedure to construct EEG synchronization graphs and extract features from the graph. We then detail the working of a feature selection method based on quadratic programming, build autoregressive models on the selected features, and transfer and transformation learn on these features. Finally, we use an alarm-based detection system to signal the seizure. In Section 3 we present and discuss the results for seizure prediction, and comparison to a benchmark. We provide an overview and outline possible extensions to this study, in Section 4.

2 Methodology

Our seizure-prediction paradigm is centered around discriminating between seizure and non-seizure activity. So we attempt to learn from normal activity maximally and from seizure activity minimally. We then interpret consistent and clear deviations from our understanding of normal activity as seizure. To this end: we initially suppress seizure activity; develop synchronization graphs to describe the seizure-suppressed cortical activity well; select maximally descriptive, minimally redundant features; cross-learn common attributes of seizure-suppressed activities across patients; and measure how well we predict on data from which the seizure has not been suppressed. Conditioned on this sequence of operations performing well, we reason that any prediction error on the data where the seizure has not been suppressed is due to seizure-related activity. We develop a simple way to determine when such seizure-related activity is reliable enough to declare an imminent seizure, and use it to make such a declaration. The following sections describe the steps involved from taking the EEG signal as input to predicting the seizure. The steps are illustrated in the block diagram shown in Fig. 1.

2.1 Epileptic EEG Data Set

Our dataset consists of scalp EEG recordings of 41 seizures from 14 patients. All the patients were evaluated with scalp video-EEG monitoring in the international 10-20 system (as described in [30]), magnetic resonance imaging (MRI), fMRI for language localization, and position emission tomography (PET). All the patients had *Hippocampal Sclerosis* (HS) except one patient (Patient-1) who suffered from *Cortical Dysplasia* (CD). After selective amygdalo-hippocampectomy, all the patients were seizure free. The patient information is provided in Table 1. For 4 patients, the seizure would onset from the right, whereas for 10 patients the seizure would onset from the left.

The recordings include sufficient pre-ictal and post-ictal periods for the analysis. Two of the electrodes (A_1 and A_2) were unused and C_z electrode was used for referential montage that yielded 18-channel EEG recordings. A team of doctors diagnosed the initiation and the termination of each seizure and reported these periods as the ground truth for our analysis. An example of such a recording can be found in Fig. 2 in [63]. Seizures were 77.12 seconds long on

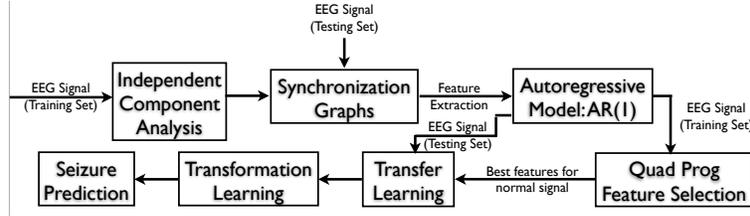


Fig. 1. Block diagram representing the entire sequence of steps involved in the methodology. We apply Independent Component Analysis (ICA) for artifact removal and noise reduction, which allows us to learn non-ictal activity. This step is carried out only on the training set of EEG recordings, and the testing set of EEG recordings is kept separate. Synchronization graphs are constructed by using Phase Lag Index as explained in Section 2.3. These graphs are constructed for both the training and testing sets. Based on the features extracted from the synchronization graphs and the signal itself, an autoregressive model is built 2.6, and this model allows us to identify predictive importance of features that are determined via a Quadratic Programming Feature Selection (QPFS) technique 2.5. The feature selection technique is applied only on the training set. The important features are then used for transfer and transformation learning 2.7 which improves the performance of seizure prediction 2.8.

average and their standard deviation was 48.94 seconds. The high standard deviation of the data is an indication of the vast variability in the data which makes the task of seizure prediction complicated.

Table 1. Patient Types. Almost all the patients (except one) exhibited hippocampal sclerosis (HS). There are two types of lateralizations in HS: left (L) and right (R). One patient (Patient-1) exhibited cortical dysplasia (CD).

Patient	Pathology	Lateralization	Number of Recordings	Length of Individual Recordings (in minutes)
Patient-1	CD	R	2	30
Patient-2	HS	R	2	30
Patient-3	HS	R	3	60
Patient-4	HS	R	5	60
Patient-5	HS	L	1	60
Patient-6	HS	L	1	30
Patient-7	HS	L	2	60
Patient-8	HS	L	2	60
Patient-9	HS	L	3	60
Patient-10	HS	L	3	30
Patient-11	HS	L	2	60
Patient-12	HS	L	5	41
Patient-13	HS	L	5	35
Patient-14	HS	L	5	35

2.2 Seizure Suppression

In order to suppress seizure-activity, we resort to modeling EEG signal acquisition as follows. We assume that: (1) seizure activity is statistically independent of normal activity and (2) there may be numerous statistically independent cortical activities, both seizure related and otherwise, that combine to provide the signal captured by a single electrode (3) the seizure activity is non-gaussian. Based on these two assumptions, we look to locate and discard the seizure-related activity, thereby suppressing the seizure. Under the assumptions stated above, the problem of extracting seizure-related activity is mathematically equivalent to the cocktail party problem exemplifying blind source separation, which is solved by the state of the art technique of Independent Component Analysis [12,13,33], which has thus far been used mainly to remove artifacts from EEG data [17,32,40]. Here we use ICA to locate seizure-related activity and remove

it in a manner similar to artifact-removal. Formally, given that $\mathbf{X} \in \mathbb{R}^{n,d}$ is a linear mixture of k d -dimensional independent components contained in $\mathbf{S} \in \mathbb{R}^{k,d}$, we may write

$$\mathbf{X} = \mathbf{A}\mathbf{S},$$

where $\mathbf{A} \in \mathbb{R}^{n \times k}$ is the *mixing matrix* and $\mathbf{S} \in \mathbb{R}^k$. In general, both \mathbf{A} and \mathbf{S} are unknown and we compute the independent components, with respect to an independence maximization measure, as $\mathbf{S} = \mathbf{W}\mathbf{X}$, where \mathbf{W} is the inverse of the mixing matrix.

Once \mathbf{A} is computed, we discard seizure related activity by zeroing the columns having the lowest euclidean norms. We reason this as follows: since much of the data is normal function, the independent components corresponding to seizure-related activity do not contribute to most of the data; their contribution is concentrated in time (corresponding to concentration in row-indices of \mathbf{A}). Due to the inherent scaling-degeneracy in the problem of blind source separation, we obtain an \mathbf{A} having unit row-norms. This leads to the coefficients corresponding to seizure-related independent components to be tightly controlled, resulting in columns corresponding to the seizure-related independent components being of low euclidean norm. We heuristically zero the lowest two columns of \mathbf{A} to form \mathbf{A}_o and declare

$$\mathbf{X}_o = \mathbf{A}_o\mathbf{S}$$

to be the seizure-suppressed EEG data. It is important to note that the seizure is not completely suppressed, but the independent components retrieved allow us to model the non-ictal activity more precisely.

2.3 Construction of EEG Synchronization Graphs

For the signal $\mathbf{f}X[i, m]$, we construct epochs of equal lengths with an overlap of 20% between the preceding and following epochs. The number of epochs, n , is equal to $1.25M/L$, where L is the duration of the epoch in same time units. Since the EEG recordings contain both temporal and spatial information, we construct *time-evolving EEG Synchronization Graphs* on the EEG datasets. A synchronization graph is constructed for each epoch, giving an indication of the spatio-temporal correspondence between electrodes - these relationships can then be utilized to obtain changes in the network by identifying descriptive features. The *nodes* represent the EEG electrodes and the *edges* represent a closeness relationship between the nodes in a given epoch. We use an epoch length of 5 seconds.

A sample time-evolving graph on an EEG recording is shown in Fig. 2. The pair-wise relationships between the electrodes during an epoch are used to construct the graph edges. If the pair-wise distance between two nodes i and k , where $i, k \in \{1, \dots, 18\}$, and $i \neq k$, for epoch t , given as $d_{i,k}^n$, is less than a specified threshold, τ , then an edge is inserted into the graph between the two nodes. Note that smaller threshold values seek higher correlation between the electrodes, thereby yielding sparser graphs. Similarly, higher threshold values would establish an edge even if there is small correlation between the data, thereby yielding denser graphs. For our analysis, we performed a parametric search and found the best value of τ to be 1.

Several synchronization measures have been proposed as plausible options for $d_{i,k}^n$ to set up the edges in the graph. Based on earlier results presented in [19], we chose Phase Lag Index (PLI) [65] for $d_{i,k}^n$. PLI is defined as follows:

$$PLI_{i,k}(n) = \frac{1}{Lf_s} \left| \sum_{m=1}^{Lf_s} \text{sgn}(\phi_i^n(m) - \phi_k^n(m)) \right| \quad (1)$$

where $\phi_i^n = \arctan\left(\frac{\hat{\mathbf{f}}x_i^n}{\mathbf{f}x_i^n}\right)$ is the angle of the Hilbert transform $\hat{\mathbf{f}}x_i^n$ of the signal $\mathbf{f}x_i^n$.

2.4 Feature Extraction from EEG Synchronization Graphs

We extract 26 features from the EEG synchronization graph for each epoch. These features quantify the compactness, clusteredness, and uniformity of the graph. Apart from these graph-based features, we compute two spectral features - the variance of the stationary distribution on an undirected markov chain on the graph, and the second largest eigenvalue of the Laplacian of the graph. In addition we compute certain natural statistics: the mean jump size between

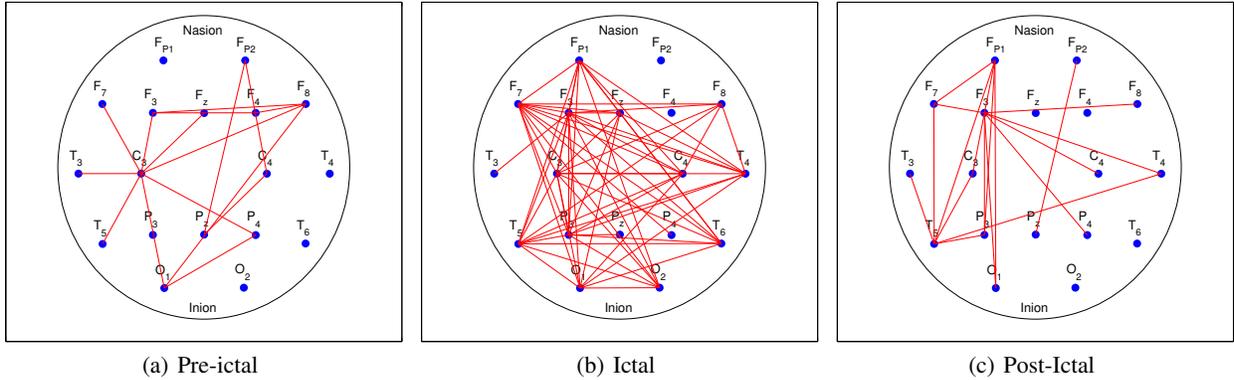


Fig. 2. Sample EEG Synchronization Graphs for pre-ictal, ictal, and post-ictal epochs. It is clearly seen that the ictal period has more coherence between different regions of the brain. The Phase Lag Index measure is used as the synchronization measure (or cost function) to set up edges between the nodes in the graph.

epochs and its variance, to measure the similarity to a Weiner process, and finally the *hinged mean* and *hinged variance*, defined as the mean and variance, respectively, of the signal at the current epoch centered/hinged at the mean of a strictly trailing window. These features arise naturally in change-point-detection and are motivated by the natural belief that, as a stochastic process, the EEG signal undergoes a statistical change when a seizure begins. To this feature set we also added time-domain and spectral features. The time-domain features include the Hjorth parameters - activity, mobility, and complexity, and the frequency-domain features include skewness of amplitude spectrum and spectral entropy [1]. In all, we calculated 38 features.

In subsequent text, we refer to the feature matrix as $\mathbf{D} \in \mathbb{R}^{n,d}$, with n epochs and d features. We refer to the feature vector at time t (row t of \mathbf{D}) as \mathbf{d}^t , and the time-series corresponding to feature i (column i of \mathbf{D}) as \mathbf{d}_i . A complete list of the features used in this work and their definitions is listed in Table 2. For further information regarding the features, we refer the reader to [1,6].

2.5 Determining The Significance of Features

The computed features were motivated by discussions with the subject matter experts, with the view of casting a meaningful but wide net to capture attributes of an epileptic seizure. However, this doesn't strictly preclude the possibility that certain features may be redundant or low in predictive importance. Furthermore, we wish to select features that are particularly descriptive of the non-ictal activity, of which the data is largely comprised. Therefore, we quantify the predictive significance of the features in a natural but effective way, and score the features to maximize their predictive importance for the entire data, and minimize redundancy, using the method in [58], which we summarize here. The primary advantages of using QPFS are as follows: (i) QPFS is based on efficient quadratic programming technique [5]. The quadratic term quantifies the dependence between each pair of variables, whereas the linear term quantifies the relationship between each feature and the class label. (ii) QPFS provides a considerable time complexity improvement over current methods on very large data sets with high dimensionality.

Measuring Redundancy Our notion of redundancy arises naturally from the interpretation of brain activity as a stochastic process, whence the usual notion of linear dependence is replaced with the notion of statistical correlation. Specifically, suppose the data matrix, $\mathbf{D} \in \mathbb{R}^{n,d}$, spanning n epochs and consisting of d features. We define, the correlation matrix, $\mathbf{Q} \in \mathbb{R}^{d,d}$, element-wise, where $\mathbf{Q}(i,j)$ is the Pearson correlation coefficient between the feature vectors $\mathbf{d}_i, \mathbf{d}_j \in \mathbb{R}^n$:

$$\mathbf{Q}(i,j) = \frac{\mathbf{d}_i^\top \mathbf{d}_j}{\|\mathbf{d}_i\| \|\mathbf{d}_j\|}.$$

Table 2. Names and description of EEG global graph features. Features 1–26 are computed on the synchronization graph, features 27 and 28 are signal-based. Features 29–31 are representative of change-point detection, and features 32 and 33 are spectral features. Features 34–36 and 37–38 are time-domain and frequency-domain features, respectively.

Index	Feature Name	Description
1	Average Degree	Average number of edges per node
2	Clustering Coefficient C	Average of the ratio of the links a node's neighbors have in between to the total number that can possibly exist
3	Clustering Coefficient D	Same as feature 2 with node added to both numerator and denominator
4	Average Eccentricity	Average of node eccentricities, where the <i>eccentricity</i> of a node is the maximum distance from it to any other node in the graph
5	Diameter of graph	Maximum of node eccentricities
6	Radius of graph	Minimum of node eccentricities
7	Average Path Length	Average hops along the shortest paths for all possible pairs of nodes
8	Giant Connected Component Ratio	Ratio between the number of nodes in the largest connected component in the graph and total the number of nodes
9	Number of Connected Components	Number of clusters in the graph excluding the isolated nodes
10	Average Connected Component Size	Average number of nodes per connected component
11	% of Isolated Points	% of isolated nodes in the graph, where an <i>isolated node</i> has a degree 0
12	% of End Points	% of endpoints in the graph, where an <i>endpoint</i> has a degree 1
13	% of Central Points	% of nodes in the graph whose eccentricity is equal to the graph radius
14	Number of Edges	Number of edges between all nodes in the graph
15	Spectral Radius	Largest eigenvalue of the adjacency matrix
16	Adjacency Second Largest Eigenvalue	Second largest eigenvalue of the adjacency matrix
17	Adjacency Trace	Sum of the adjacency matrix eigenvalues
18	Adjacency Energy	Sum of the square of adjacency matrix eigenvalues
19	Spectral Gap	Difference between the magnitudes of the two largest eigenvalues
20	Laplacian Trace	Sum of the Laplacian matrix eigenvalues
21	Laplacian Energy	Sum of the square of Laplacian matrix eigenvalues
22	Normalized Laplacian Number of 0's	Number of eigenvalues of the normalized Laplacian matrix that are 0
23	Normalized Laplacian Number of 1's	Number of eigenvalues of the normalized Laplacian matrix that are 1
24	Normalized Laplacian Number of 2's	Number of eigenvalues of the normalized Laplacian matrix that are 2
25	Normalized Laplacian Upper Slope	The sorted slope of the line for the eigenvalues that are between 1 and 2
26	Normalized Laplacian Trace	Sum of the normalized Laplacian matrix eigenvalues
27	Mean of EEG recording	Mean of EEG signal for each electrode and epoch
28	Variance of EEG recording	Variance of EEG signal for each electrode and epoch
29,30	Change-based Features	Mean and variance of jump size in EEG signal for each electrode and epoch
31	Change-based Feature 3	Variance of EEG signal for particular electrode in given epoch after subtracting the mean of up to 3 previous windows
32	Spectral Feature 1	Variance of eigenvector of the product of the adjacency matrix and the inverse of the degree matrix
33	Spectral Feature 2	Second largest eigenvalue of the Laplacian matrix
34,35,36	Hjorth parameters (time-domain)	Activity, Mobility, and Complexity
37,38	Frequency-domain features	Skewness of amplitude spectrum and Spectral entropy

The quadratic form $\mathbf{x}^\top \mathbf{Q} \mathbf{x}$ thus has the natural interpretation of yielding the sample-covariance of a compound feature, with coefficients contained in \mathbf{x} , which is the notion of redundancy that we wish to minimize.

Measuring Predictive Importance We first recall that the activity of the brain at time t is completely captured by \mathbf{d}^t . We define the predictive importance, f_i , of the feature i , as the r.m.s. influence of \mathbf{d}_i^t on \mathbf{d}_j^{t+1} , $1 \leq j \leq n$, measured by the coefficients in the forecast operator corresponding to i . Formally, let $\Psi \in \mathbb{R}^{d,d+1}$ be the forecast operator. Then our best prediction of \mathbf{d}^{t+1} is $\tilde{\mathbf{d}}^{t+1}$ where

$$\tilde{\mathbf{d}}^{t+1} = \mathbf{d}^t \Psi^\top.$$

The influence, $\mathbf{p}_i(j)$, of feature i on j , contained in $\mathbf{p}_i \in \mathbb{R}^d$, may be determined by predicting via Ψ using its indicator vector, \mathbf{e}_i :

$$\mathbf{p}_i = \mathbf{e}_i \Psi^\top,$$

whence the r.m.s. influence, f_i , of i is simply

$$f_i = \|\mathbf{p}_i\| = \|\Psi_i\|,$$

the column-norm of the forecast operator corresponding to column i . We define $\mathbf{f} \in \mathbb{R}^d$ such that $\mathbf{f}_i = \|\Psi_i\|$, as the predictive importance vector.

Optimizing Redundancy and Predictive Importance We obtain a significance-distribution over the features that maximizes predictive importance and minimize redundancy by solving

$$\begin{aligned} \mathbf{x}^* = & \arg \min q(\mathbf{x}); \\ & \text{subject to } \mathbf{x} \in \mathbb{R}^d, \mathbf{x} \geq \mathbf{0}, \sum_i \mathbf{x}_i = 1, \end{aligned} \quad (2)$$

where the constraints arise from forcing the resulting vector to be a distribution, from which we omit an appropriately sized tail, or select just the support if it is small. To make the objective function stable under scaling of the data, we normalize \mathbf{f} to obtain

$$\hat{\mathbf{f}} = \mathbf{f} / \|\mathbf{f}\|_\infty.$$

To effect a meaningful trade-off between minimizing redundancy and maximizing predictive importance, we take a convex combination of the corresponding terms:

$$q(\mathbf{x}) = (1 - \alpha)\mathbf{x}^\top \mathbf{Q}\mathbf{x} - \alpha\hat{\mathbf{f}}^\top \mathbf{x},$$

where α is chosen, as in [58] as

$$\alpha = \frac{\sum_{i,j} \mathbf{Q}(i,j)/d^2}{\sum_{i,j} \mathbf{Q}(i,j)/d^2 + \sum_k \mathbf{f}_k/d}.$$

Since both the predictive importance and the correlation matrix are statistical in nature, they are less affected by the relatively fleeting seizure. So we expect the significant features obtained via QPFS to be features that are highly predictive of non-ictal activity. We use the MATLAB utility quadprog to solve Eq. 2.

2.6 Autoregressive Modeling on Feature Data

Research has indicated that promising results regarding early detection or prediction of the seizure can be achieved by application of an autoregressive model (AR) to the EEG signal [2,11]. Also, AR models are linear and as shown in prior research are comparable to non-linear models in their predictive capability [14,15,43,49,51]. We expand on these earlier results by applying an autoregressive model to the features extracted both from the graph and the signal itself.

An autoregressive model of order 1, AR(1), is applied to the matrix \mathbf{D} , extracted from the time-evolving EEG synchronization graphs. For an AR(1) model the output at time t is only dependent on the values of the time-series at time $t - 1$. As a result, the implicit assumption when using AR(1) is that \mathbf{d}^t is a markov chain indexed by t . Formally

$$\mathbf{d}_i^t = \rho_{i0} + \rho_{i1}\mathbf{d}_1^{t-1} + \rho_{i2}\mathbf{d}_2^{t-1} + \dots + \rho_{im}\mathbf{d}_m^{t-1} + \epsilon_t \quad (3)$$

where ρ_{ij} are the linear coefficients computed via autoregression. In matrix form, (3) is $[\mathbf{D}]_1^t = \Psi \cdot [1, \mathbf{D}]_0^{t-1} + \epsilon$, where the notation $[\mathbf{A}]_a^b$ denotes a matrix containing all rows from a to b of \mathbf{A} , including rows a, b . We compute Ψ to minimize the error ϵ in euclidean norm,

$$\begin{aligned} \Psi &= \arg \min_{\mathbf{Z}} \left\| [\mathbf{D}]_1^t - \mathbf{Z} \cdot [1, \mathbf{D}]_0^{t-1} \right\|_F^2, \\ \Psi &= [\mathbf{D}]_1^t \cdot ([1, \mathbf{D}]_0^{t-1})^\dagger. \end{aligned} \quad (4)$$

where \mathbf{A}^\dagger denotes the moore-penrose pseudoinverse of \mathbf{A} . The role of the operator Ψ is to predict $\mathbf{D}(t)$ as a function of $\mathbf{D}(t - 1)$. Any operator that does this will be called subsequently as the *forecast operator*. Thus, using an AR(1) model we arrive at a forecast operator, Ψ .

2.7 Transfer Learning and Transformation Learning on Autoregressive Model

We critically improve this forecast operator obtained from AR(1) in two directions. First, we improve it under the assumption that the data obtained from different patients are not completely independent of each other; that data obtained from one patient holds some information common to all patients, along with information specific to the patient. Thus, we *transfer* knowledge from one patient to another, motivated by the existing work on Transfer Learning. Specifically, given a feature data set \mathbf{D} , the feature transfer set, $\hat{\mathbf{D}}$, and the corresponding forecast operators Ψ and $\hat{\Psi}$ respectively we transfer knowledge from $\hat{\mathbf{D}}$ to \mathbf{D} by regularizing (4) with

$$\lambda \|\Psi - \hat{\Psi}\|_F^2.$$

The parameter λ , playing the familiar role of the Tikhonov Regularizer, is the *transfer coefficient*, governing how much we learn from $\hat{\mathbf{D}}$ onto \mathbf{D} . The forecast operator obtained from this transfer learning is simply

$$\bar{\Psi} = \arg \min \left\{ \left\| [\mathbf{D}]_{t-1}^0 * \Psi - [\mathbf{D}]_t^1 \right\|_F^2 + \lambda \left\| \Psi - \hat{\Psi} \right\|_F^2 \right\}, \quad (5)$$

the analytical solution for which is:

$$\Psi = \hat{\Psi} + \left([\mathbf{D}]_{t-1}^0 \top [\mathbf{D}]_{t-1}^0 + \lambda \mathbf{I} \right)^\dagger \left([\mathbf{D}]_{t-1}^0 \right)^\top \left([\mathbf{D}]_t^1 - [\mathbf{D}]_{t-1}^0 \top \hat{\Psi} \right) \quad (6)$$

As more data is obtained for \mathbf{D} , the value of λ is reduced because now the core set is getting better at predicting its own future values. To test our estimates, we use the following split between the training and testing data. First, we split \mathbf{D} into training (TR) and testing (TE) sets. Within the training set, we create a further split thereby creating training prime (\overline{TR}) and validation (Val) sets. We then train our AR(1) model on \overline{TR} , and then use $\hat{\Psi}$ to improve this estimate by testing on Val . Then, we retrain the model using the learned parameters on the entire training set TR and finally test on TE .

Next, we account for the differences in the collected data that may arise as a result of non-uniformities in the process of acquiring data. We do this under the assumption that the spectral nature of the data is minimally variant with changes across the various setups for acquiring data from multiple patients, and that the flow of time is immutable. To learn from $\hat{\mathbf{D}}$ onto \mathbf{D} , we find the object, $\hat{\mathbf{D}}_{\mathbf{D}}$, retaining the spectral nature of $\hat{\mathbf{D}}$, and respecting the directionality of time, that is the closest to \mathbf{D} . Formally, we find a rotation $\Gamma(\hat{\mathbf{D}}, \mathbf{D})$ such that

$$\hat{\mathbf{D}}_{\mathbf{D}} = \hat{\mathbf{D}} \Gamma(\hat{\mathbf{D}}, \mathbf{D}),$$

$$\Gamma(\hat{\mathbf{D}}, \mathbf{D}) = \arg \min_{\mathbf{U}^\top \mathbf{U} = \mathbf{I}} \left\| \mathbf{D} - \hat{\mathbf{D}} \mathbf{U} \right\|.$$

This is the Procrustes problem that has been well-studied [72], and has a closed form solution in terms of the SVD of $\mathbf{D}, \hat{\mathbf{D}}$. Let these SVDs be

$$\mathbf{D} = \mathbf{U}_{\mathbf{D}} \Sigma_{\mathbf{D}} \mathbf{V}_{\mathbf{D}}^\top, \quad \hat{\mathbf{D}} = \mathbf{U}_{\hat{\mathbf{D}}} \Sigma_{\hat{\mathbf{D}}} \mathbf{V}_{\hat{\mathbf{D}}}^\top.$$

Then

$$\Gamma(\hat{\mathbf{D}}, \mathbf{D}) = \mathbf{V}_{\hat{\mathbf{D}}} \mathbf{V}_{\mathbf{D}}^\top, \Rightarrow \hat{\mathbf{D}}_{\mathbf{D}} = \mathbf{U}_{\hat{\mathbf{D}}} \Sigma_{\hat{\mathbf{D}}} \mathbf{V}_{\mathbf{D}}^\top.$$

We now transfer-learn using $\hat{\mathbf{D}}_{\mathbf{D}}$. In summary, we first notice that knowledge can be transferred from other but similar data and then transform such similar data sets into their *most learnable* forms using a simple transformation.

2.8 Declaration Of Imminent Seizure

We use the prediction errors incurred by the use of our forecast operator as the eventual *ictal discriminator*. We compute an estimate of the probability of deviation towards seizure using these errors, and declare that a seizure is imminent when this probability is reliably high. We outline how we compute this probability and quantify the sense of reliability we use, in that order.

Probability Of Deviation Towards Seizure Let $\epsilon(t)$ be the prediction error at time, t . Using a moving window of size $\Delta = 30$, we first use a simple statistical thresholding on the errors to determine if an *alarm* has to be thrown, which signifies an outlier to normal function, and a potential seizure. Specifically, let $\kappa(t)$ be the binary variable indicating whether or not an alarm is thrown - 1 when it is thrown and 0 when it is not. Let $\mu(t, \Delta), \sigma(t, \Delta)$ denote the mean and standard deviation of the sequence $\epsilon(t), \epsilon(t + 1), \dots, \epsilon(t + \Delta)$. Then

$$\kappa(t) = \begin{cases} 1 & \text{if } \epsilon(t + \Delta) - \mu(t, \Delta) > \tau^* \sigma(t, \Delta) \\ 0 & \text{otherwise} \end{cases}$$

where τ^* is a tolerance/sensitivity parameter. Clearly, κ is an indicator of the one-sided tail of the distribution from which $\epsilon(t)$ is drawn. Under our assumption that a recorded activity is either normal function or seizure, the measure of the tail of the distribution of errors during normal function is an appropriate estimate of the probability of seizure. We estimate the size of this *tail of normality* for an interval by the ensemble average of $\kappa(t)$ for the interval. When $\kappa(t)$ indicates seizure repeatedly in a manner highly unlikely to have arisen from random sampling from the tail of normality, we declare an imminent seizure. In practice, we choose $\tau^* = 3$, and declare a seizure when we see 3 consecutive alarms. We justify this choice as follows: in the case where $\kappa(t)$ indicates the result of the high-entropy fair coin toss (i.e. 1 and 0 with equal probability), the probability of obtaining 3 consecutive alarms is 12.5%. In practice our estimate of the size of the tail of normality is significantly below 1/5, resulting in our three-in-a-row rule to be an even rarer occurrence than once in 125 occurrences or 0.8%.

3 Results

We present results for the Quadratic Programming Feature Selection algorithm, determining the best forecast operator, and a comparison of the performance of our seizure prediction algorithm on basic autoregression vs. with the addition of transfer and transformation learning.

3.1 Quadratic Programming Feature Selection Results

The feature-significance vectors obtained from solving the QPFS problem in (2) were found to be highly sparse, and the features that were supported by these vectors were chosen without exception - 9 in all: (i) Average Degree, (ii) Diameter of graph, (iii) Average Path Length, (iv) Giant Connected Component Ratio, (v) Number of Connected Components, (vi) Percentage of Isolated Points, (vii) Number of eigenvalues with value 0 of the normalized Laplacian matrix, (viii) Number of eigenvalues with value 2 of the normalized Laplacian matrix, and (ix) Normalized Laplacian trace.

3.2 Baseline SVM results

To establish a baseline to validate the efficacy of our results, we compare our algorithm to the following benchmark algorithm:

Application of Support Vector Machine to feature matrix \mathbf{D} :

We provide as input the features identified by QPFS from \mathbf{D} to a two-class Support Vector Machine (SVM). We learn a model of the inter-ictal and ictal states based on their respective feature values. We then classify using the SVM the seizure onset in the pre-ictal region based on the feature values in that region. The intuition being that the initial part of the pre-ictal region will have features similar to the inter-ictal region, whereas the latter part will be more similar to the ictal region in the feature space. We consider the pre-ictal region to start 10 minutes prior to the onset of the seizure.

We found that the benchmark did not predict the seizure in 14 of the 41 analyzed recordings. The median prediction time for the recordings for which seizures were predicted was 1.25 minutes prior to the seizure.

3.3 Autoregression vs transfer and transformation learning

One of the objectives of this study was to improve the basic autoregressive model by the application of transfer learning and transformation learning. We now show that the additional functionality makes the prediction either at least as good as that by the AR(1) model or better for a significant percentage of the dataset. We found that in 60% of the analyzed EEG recordings, transformation learning was able to predict the seizure prior to AR(1), or transfer learning. In 52.5% of the analyzed EEG recordings, transfer learning performed better than the AR(1) model predicting the seizure earlier. Finally, in 67.5% of the cases, either transfer learning or transformation learning was better than the AR(1) model. The median prediction times prior to the occurrence of the seizure for the three methods are 10 min, 10.96 min, and 11.04 min for AR(1), transfer learning, and transformation learning, respectively. Considering only the recordings where transformation learning or transfer learning outperformed AR(1), the median prediction times change to 9.33 min for AR(1), 11.17 min for transfer learning, and 11.92 min for transformation learning.

In Fig. 3, the first row ((a)–(c)) consists of an analyzed EEG recording where the AR(1) model was able to predict the seizure before the other two techniques. The second row ((d)–(f)) consists of an analyzed EEG recording where the AR(1) model with transfer learning and transformation learning was able to predict the seizure before the other two techniques. Finally, the third row ((g)–(i)) consists of an anomalous result where the AR(1) model with transfer learning predicted the seizure before either AR(1) model or the AR(1) model with transfer and transformation learning.

4 Conclusions and Future Work

In this study, we outline a seizure prediction algorithm designed for EEG epileptic seizure data by constructing an autoregressive model improved by the addition of transfer learning and transformation learning on features extracted by building synchronization graphs on the independent components of the EEG signal. We use a quadratic programming algorithm called Quadratic Programming Feature Selection (QPFS) to select the features with the highest predictive importance and minimal redundancy.

One of the primary concerns with the seizure prediction area is the definition of a Seizure Prediction Horizon (SPH). In the literature prediction horizons have varied from several minutes to a few hours [55]. We would like to come up with a more rigorous theoretical basis for assigning prediction horizons. Another future direction is with respect to the various thresholds used in the study. Although, well-motivated and justified from the literature, we would like to obtain these thresholds from first principles. Examples of these thresholds include epoch lengths for the synchronization graphs, sensitivity parameters for raising an alarm, and number of columns to zero out from the mixing matrix in Independent Component Analysis (ICA). Yet another important future direction is analyzing the partial contribution of each module in the pipeline to determine the effect of individual modules in improving the basic prediction. Specifically, we would like to examine the influence of ICA vs. transformation learning to determine which of the two is better used for the initial surgery to suppress seizure - to ensure that we don't use a gas-engine for the short haul and a horse for the long one. Furthermore, we would like to qualify the use of transfer learning based on the similarity of the data sets being learned across: establish a metric of closeness of data sets/learnability across patients and recordings. Finally, the problem of seizure prediction is accompanied by the problem of localizing seizure, which, apart from requiring new methods, also sets a higher standard for understanding seizure. We hope to contribute to this problem in the future as well.

References

1. Acar, E., Aykut-Bingol, C., Bingol, H., Bro, R., Yener, B.: Multiway analysis of epilepsy tensors. *Bioinformatics* 23(13), i10–i18 (2007)
2. Alkan, A., Koklukaya, E., Subasi, A.: Automatic seizure detection in EEG using logistic regression and artificial neural network. *J Neurosci Methods* 148(2), 167–176 (2005)
3. Anderson, N.R., Wisneski, K., Eisenman, L., Moran, D.W., Leuthardt, E.C., Krusienski, D.J.: An offline evaluation of the autoregressive spectrum for electrocorticography. *IEEE Trans Biomed Eng* 56(3), 913–916 (2009)
4. Barrat, A., Barthelemy, M., Vespignani, A.: *Dynamical processes on complex networks*, vol. 1. Cambridge University Press Cambridge (2008)
5. Bertsekas, D.P.: *Nonlinear Programming*. Athena Scientific, 2nd edn. (Sep 1999)

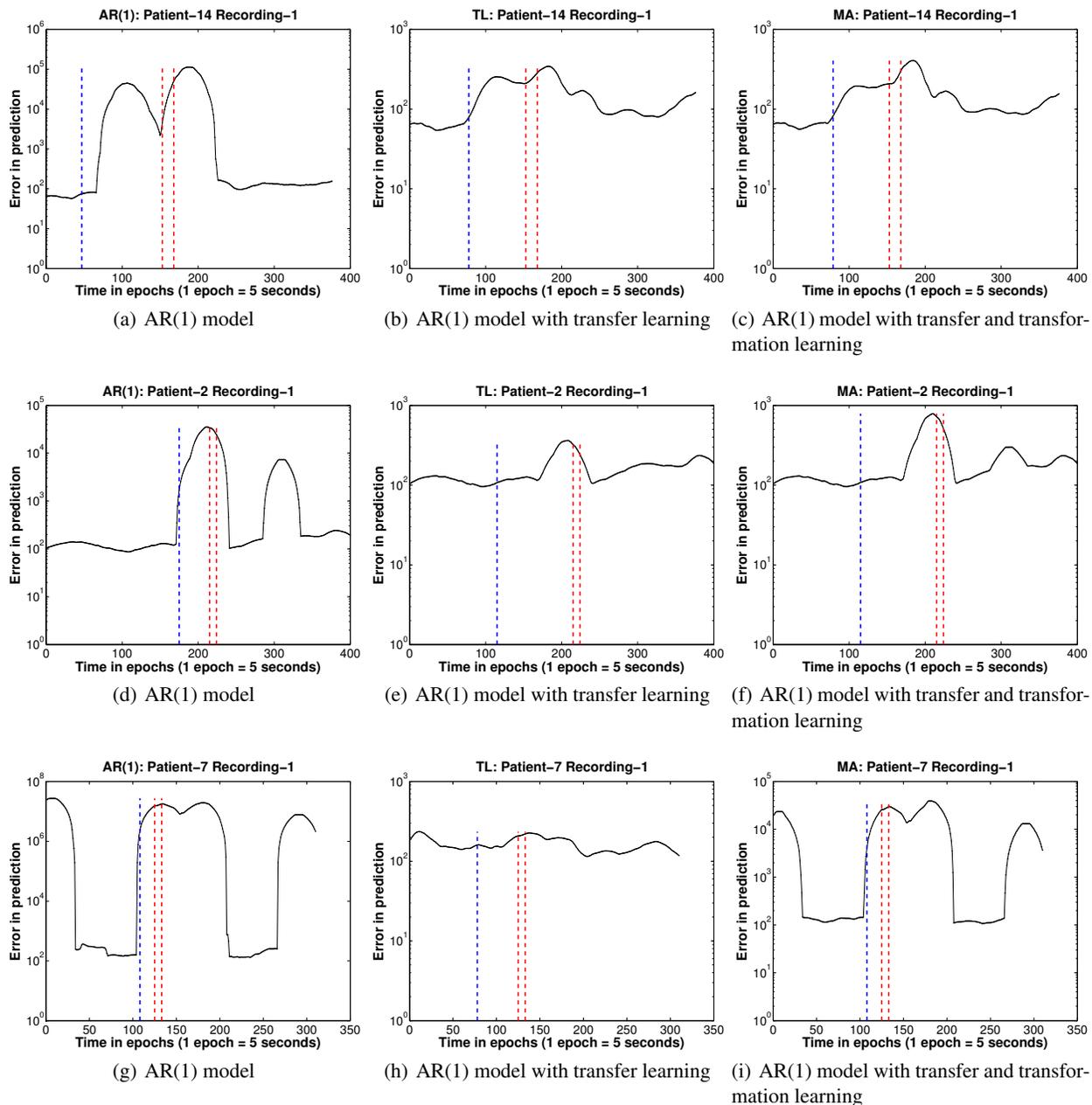


Fig. 3. Comparison of the AR(1) model ((a), (d), and (g)) with the AR(1) model improved by transfer learning ((b), (e), and (h)) and transformation learning ((c), (f), and (i)). The epoch at which the seizure is detected is shown in blue, the start and end of the seizure region are marked in red. The first row is an example of where the AR model does better than both transfer learning and transformation learning. The second row is a typical example of the AR(1) model with transfer and transformation learning outperforming both the AR(1) model and the AR(1) model enhanced by transfer learning. The third row is an example of an anomaly where the AR(1) model with transfer learning performs much better than the AR(1) model and the AR(1) model with transfer and transformation learning.

6. Bilgin, C.C., Ray, S., Baydil, B., Daley, W.P., Larsen, M., Yener, B.: Multiscale feature analysis of salivary gland branching morphogenesis. *PLoS ONE* 7(3), e32906 (2012)
7. Boccaletti, S., Latora, V., Moreno, Y., Chavez, M., Hwang, D.U.: Complex networks: Structure and dynamics. *Physics Reports* 424(4–5), 175 – 308 (2006), <http://www.sciencedirect.com/science/article/pii/S037015730500462X>
8. Bronzino, J.D.: *Biomedical Engineering Handbook*, chap. Principles of electroencephalography. New York: Taylor and Francis, 3rd edn. (2006)
9. Bullmore, E., Sporns, O.: Complex brain networks: graph theoretical analysis of structural and functional systems. *Nature Reviews Neuroscience* 10(3), 186–198 (2009)
10. Chandaka, S., Chatterjee, A., Munshi, S.: Cross-correlation aided support vector machine classifier for classification of EEG signals. *Expert Syst Appl* 36(2, Part 1), 1329–1336 (2009)
11. Chisci, L., Mavino, A., Perferi, G., Sciandrone, M., Anile, C., Colicchio, G., Fuggetta, F.: Real-time epileptic seizure prediction using AR models and support vector machines. *IEEE Trans Biomed Eng* 57(5), 1124–1132 (May 2010)
12. Comon, P.: Independent component analysis – A new concept. *Signal Process.* 36, 287–314 (1994)
13. Comon, P., Jutten, C.: *Handbook of Blind Source Separation: Independent Component Analysis and Applications*. Academic Press, 1 edn. (2010)
14. Corsini, J., Shoker, L., Sanei, S., Alarcon, G.: Epileptic seizure predictability from scalp EEG incorporating constrained blind source separation. *IEEE Trans Biomed Eng* 53, 790–799 (2006)
15. Cranstoun, S.D., Ombao, H.C., von Sachs, R., Guo, W., Litt, B., et al.: Time-frequency spectral estimation of multichannel EEG using the auto-slex method. *IEEE Trans Biomed Eng* 49(988–996) (2002)
16. D’Alessandro, M., Vachtsevanos, G., Esteller, R., Echauz, J., Cranstoun, S., Worrell, G., et al.: A multi-feature and multi-channel univariate selection process for seizure prediction. *Clin Neurophysiol* 116, 506–516 (2005)
17. Delorme, A., Sejnowski, T.J., Makeig, S.: Enhanced detection of artifacts in EEG data using higher-order statistics and independent component analysis. *Neuroimage* 4(1443–1449) (34)
18. Demir, C., Gultekin, S.H., Yener, B.: Augmented cell-graphs for automated cancer diagnosis. *Bioinformatics* 21(suppl 2), ii7–ii12 (2005)
19. Dhulekar, N., Oztan, B., Yener, B., Bingol, H.O., Irim, G., Aktekin, B., Aykut-Bingol, C.: Graph-theoretic analysis of epileptic seizures on scalp eeg recordings. In: *Proceedings of the 5th ACM Conference on Bioinformatics, Computational Biology, and Health Informatics*. pp. 155–163. BCB ’14, ACM, New York, NY, USA (2014), <http://doi.acm.org/10.1145/2649387.2649423>
20. Douw, L., van Dellen, E., de Groot, M., Heimans, J.J., Klein, M., Stam, C.J., Reijneveld, J.C.: Epilepsy is related to theta band brain connectivity and network topology in brain tumor patients. *BMC Neuroscience* 11(1), 103 (2010)
21. Elger, C.E.: Future trends in epileptology. *Curr Opin Neurol* 14, 185–186. (2001)
22. Esteller, R., Echauz, J., D’Alessandro, M., Worrell, G., Cranstoun, S., Vachtsevanos, G., et al.: Continuous energy variation during the seizure cycle: Towards an on-line accumulated energy. *Clin Neurophysiol* 116, 517–526 (2005)
23. Fisher, N., Talathi, S.S., Carney, P.R., Ditto, W.L.: Epilepsy detection and monitoring. In: Tong, S., Thankor, N.V. (eds.) *Quantitative EEG Analysis Methods and Applications*, pp. 157–183. Artech House (2008)
24. Fisher, R.S., van Emde Boas, W., Blume, W., Elger, C., Genton, P., Lee, P., Engel, J.J.: Epileptic seizures and epilepsy: definitions proposed by the international league against epilepsy (ILAE) and the international bureau for epilepsy (IBE). *Epilepsia* 46(4), 470–472 (2005)
25. Giannakakis, G., Sakkalis, V., Pediaditis, M., Tsiknakis, M.: *Methods for Seizure Detection and Prediction: An Overview*. pp. 1–27. *Neuromethods* (2014)
26. Güler, N.F., Übeyli, E.D., Güler, I.: Recurrent neural networks employing lyapunov exponents for EEG signals classification. *Expert Syst Appl* 29(3), 506–514 (2005)
27. Harrison, M.A., Frei, M.G., Osorio, I.: Accumulated energy revisited. *Clin Neurophysiol* 116, 527–531 (2005a)
28. Hazarika, N., Chen, J.Z., Tsoi, A.C., Sergejew, A.: Classification of eeg signals using the wavelet transform. *Signal Process.* 59, 61–72 (1997)
29. Iasemidis, L.D., Shiau, D.S., Pardalos, P.M., Chaovalitwongse, W., Narayanan, K., Prasad, A., et al.: Long-term prospective on-line real-time seizure-prediction. *Clin Neurophysiol* 116, 532–544 (2005)
30. Jasper, H.H.: The ten-twenty electrode system of the international federation. *Electroencephalogr Clin Neurophysiol Suppl.* 10, 371–375 (1958)
31. Jouny, C.C., Franzaszczuk, P.J., Bergey, G.K.: Signal complexity and synchrony of epileptic seizures: Is there an identifiable preictal period? *Clin Neurophysiol* 116, 552–558 (2005)
32. Jung, T.P., Makeig, S., Humphries, C., Lee, T.W., McKeown, M.J., Iragui, V., Sejnowski, T.J.: Removing electroencephalographic artifacts by blind source separation. *Psychophysiology* 37, 163–178 (2000)
33. Jutten, C., Herault, J.: Blind separation of sources, part I: An adaptive algorithm based on neuromimetic architecture. *Signal Process.* 24, 1–10 (1991)

34. Kannathal, N., Choo, M.L., R., A.U., Sadasivan, P.K.: Entropies for detection of epilepsy in EEG. *Comput Methods Programs Biomed* 80(3), 187–194 (2005)
35. Kramer, M.A., Kolaczyk, E.D., Kirsch, H.E.: Emergent network topology at seizure onset in humans. *Epilepsy Research* 79(2), 173–186 (2008)
36. Kuhnert, M.T., Elger, C.E., Lehnertz, K.: Long-term variability of global statistical properties of epileptic brain networks. *Chaos: An Interdisciplinary Journal of Nonlinear Science* 20(4), 043126 (2010), <http://scitation.aip.org/content/aip/journal/chaos/20/4/10.1063/1.3504998>
37. Le Van, Q.M., Navarro, V., Martinerie, J., Baulac, M., Varela, F.J.: Toward a neurodynamical understanding of ictogenesis. *Epilepsia* 44(12), 30–43 (2003)
38. Le Van, Q.M., Soss, J., Navarro, V., Robertson, R., Chavez, M., Baulac, M., Martinerie, J.: Preictal state identification by synchronization changes in long-term intracranial EEG recordings. *Clinical Neurophysiology* 116, 559–568 (2005)
39. Le Van Quyen, M., Soss, J., Navarro, V., Robertson, R., Chavez, M., Baulac, M., et al.: Preictal state identification by synchronization changes in long-term intracranial EEG recordings. *Clin Neurophysiol* 116, 559–568 (2005)
40. Lee, T.W., Girolami, M., Sejnowski, T.J.: Independent component analysis using an extended infomax algorithm for mixed sub-gaussian and super-gaussian sources. *Neural Computation* 11(2), 417–441 (1999)
41. Lehnertz, K., Litt, B.: The first international collaborative workshop on seizure prediction: Summary and data description. *Clin Neurophysiol* 116, 493–505 (2005)
42. Li, G., Semerci, M., Yener, B., Zaki, M.J.: Effective graph classification based on topological and label attributes. *Statistical Analysis and Data Mining: The ASA Data Science Journal* 5(4), 265–283 (2012)
43. Litt, B., Esteller, R., Echauz, J., D’Alessandro, M., Shor, R., et al.: Epileptic seizures may begin hours in advance of clinical onset: A report of five patients. *Neuron* 30, 51–64 (2001)
44. Liu, H.S., Zhang, T., Yang, F.S.: A multistage, multimethod approach for automatic detection and classification of epileptiform EEG. *IEEE Trans Biomed Eng* 49(12 Pt 2), 1557–1566 (2002)
45. Lytton, W.W.: Computer modeling of Epilepsy. *Nat. Rev. Neurosci.* 9(8), 626–637 (Aug 2008)
46. Mahyari, A., Aviyente, S.: Identification of dynamic functional brain network states through tensor decomposition. In: 39th IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2014) (2014)
47. Mirowski, P., Madhavan, D., LeCun, Y., Kuzniecky, R.: Classification of patterns of EEG synchronization for seizure prediction. *Clinical Neurophysiology* 120, 1927–1940 (2009)
48. Mormann, F., Andrzejak, R.G., Elger, C.E., Lehnertz, K.: Seizure prediction: the long and winding road. *Brain* 130, 314–333 (2007)
49. Mormann, F., Kreuz, T., Andrzejak, R., David, P., Lehnertz, K., et al.: Epileptic seizures are preceded by a decrease in synchronization. *Epilepsy Research* 53, 173–185 (2003)
50. Mormann, F., Kreuz, T., Rieke, C., Andrzejak, R.G., Kraskov, A., David, P., et al.: On the predictability of epileptic seizures. *Clin Neurophysiol* 116, 569–587 (2005)
51. Mormann, F., Lehnertz, K., David, P., Elger, C.E.: Mean phase coherence as measure for phase synchronization and its application to the EEG of epilepsy patients. *Physica D* 144, 358–369 (2000)
52. Murali, S., Kulish, V.V.: Modeling of evoked potentials of electroencephalograms: An overview. *Digit. Signal Process.* 17, 665–674 (2007)
53. Muthuswamy, J., Thakor, N.V.: Spectral analysis methods for neurological signals. *J. Neurosci. Methods* 83, 1–14 (1998)
54. Newman, M.E.J.: The structure and function of complex networks. *SIAM Review* 45, 167–256 (2003)
55. Osorio, I., Zaveri, H., Frei, M., Arthurs, S.: *Epilepsy: The Intersection of Neurosciences, Biology, Mathematics, Engineering, and Physics*. Taylor & Francis (2011), <http://books.google.com/books?id=097hKvvyYgsC>
56. Pan, S.J., Yang, Q.: A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering* 22(10), 1345–1359 (Oct 2010)
57. van Putten, M.J.A.M., Kind, T., Visser, F., Lagerburg, V.: Detecting temporal lobe seizures from scalp EEG recordings: a comparison of various features. *Clinical Neurophysiology* 116(10), 2480–2489 (2005)
58. Rodriguez-Lujan, I., Huerta, R., Elkan, C., Cruz, C.S.: Quadratic programming feature selection. *J. Mach. Learn. Res.* 11, 1491–1516 (8 2010)
59. Rogowski, Z., G.I., Bental, E.: On the prediction of epileptic seizures. *Biol Cybern* 42, 9–15 (1981)
60. Salant, Y., Gath, I., Henriksen, O.: Prediction of epileptic seizures from two-channel EEG. *Med Biol Eng Comput* 36, 549–56 (1998)
61. Schindler, K.A., Bialonski, S., Horstmann, M.T., Elger, C.E., Lehnertz, K.: Evolving functional network properties and synchronizability during human epileptic seizures. *CHAOS: An Interdisciplinary Journal of Nonlinear Science* 18(3), 033119 (2008)
62. Siegel, A., Grady, C.L., Mirsky, A.F.: Prediction of spike-wave bursts in absence epilepsy by EEG power-spectrum signals. *Epilepsia* 116, 2266–301 (1982)

63. Smith, S.J.M.: EEG in the diagnosis, classification, and management of patients with epilepsy. *J Neurol Neurosurg Psychiatry* 76, ii2–i–i7 (2005)
64. Srinivasan, V., Eswaran, C., Sriraam, N.: Artificial neural network based epileptic detection using time-domain and frequency-domain features. *J Med Syst* 29(6), 647–660 (2005)
65. Stam, C.J., Nolte, G., Daffertshofer, A.: Phase lag index: Assessment of functional connectivity from multi channel EEG and MEG with diminished bias from common sources. *Human Brain Mapping* 28, 1178–1193 (2007)
66. Stam, C., van Straaten, E.: The organization of physiological brain networks. *Clinical Neurophysiology* (2012)
67. Strogatz, S.H.: Exploring complex networks. *Nature* 410(6825), 268–276 (03 2001), <http://dx.doi.org/10.1038/35065725>
68. Subasi, A., Alkan, A., Koklukaya, E., Kiymik, M.K.: Wavelet neural network classification of EEG signals by using ar models with mle processing. *Neural Networks* 18(7), 985–997 (2005)
69. Tzallas, A.T., Tsipouras, M.G., Fotiadis, D.I.: The use of time-frequency distributions for epileptic seizure detection in EEG recordings. 29th Annual international conference of the IEEE engineering in medicine and biology society pp. 1265–1268 (2007)
70. Tzallas, A.T., Tsipouras, M.G., Fotiadis, D.I.: Epileptic seizure detection in EEGs using time-frequency analysis. *IEEE Trans Inf Technol Biomed* 13(5), 703–710 (2009)
71. Viglione, S.S., Walsh, G.O.: Epileptic seizure prediction. *Electroencephalogr Clin Neurophysiol* 39, 435–436 (1975)
72. Wang, C., Mahadevan, S.: Manifold alignment using procrustes analysis. In: Proceedings of the 25th International Conference on Machine Learning. pp. 1120–1127. ICML '08, ACM, New York, NY, USA (2008), <http://doi.acm.org/10.1145/1390156.1390297>
73. Wu, H., Li, X., Guan, X.: Networking property during epileptic seizure with multi-channel eeg recordings. In: Advances in Neural Networks - ISNN 2006, Lecture Notes in Computer Science, vol. 3973, pp. 573–578. Springer Berlin Heidelberg (2006), http://dx.doi.org/10.1007/11760191_84