Seizure Prediction by Graph Mining and Transfer Learning

Anonymous Author(s)
Affiliation
Address
email

Abstract

Abstract goes here.

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 [7]. 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 [8].

In [21] the authors posed the question whether characteristic features can be extracted from the continuous EEG signal that are predictive of an impending seizure. Many different approaches have been applied towards determining these features, such as frequency domain tools [4,24], wavelets [9,23], Markov processes [18], autoregressive models [2, 12,32], and artificial neural networks [17]. If it were possible to reliably predict seizure occurrence then preventive clinical strategies would be replaced by patient specific proactive therapy such as reseting brain by electrical or other 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.

Most general approaches to seizure prediction problem share several common steps including processing of multichannel EEG signals (i) discretize the time series into sliding windows with a constant number and overlapping epochs (ii) to extract sub-frequencies, to analyze the signal in frequency and/or time domain using (e.g., using wavelength transformation [1]), (iii) extract features either directly from the signal or from transformations. These featues can be univariate computed on each EEG channel separately or bivariate computed between two or more EEG channels they can be linear or nonlinear (e.g., [22]. A list of features used in characterization of epileptic seizure dynamics can be found in recent studies [22,25,34].

A particular bivariate EEG feature, which captures brainwave synchronization patterns, has been shown to be important in differentiating interictal, preictal and ictal states [13,14]. 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 seizure [20].

Recently there has been increasing focus in analyzing EEG recordings by building a *synchronization graph* that enable characterization of the pairwise correlations between electrodes using graph theoretical features over time [5,30]. In the spatio-temporal EEG graphs, *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 and autism, dementia, and epilepsy [30]. 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 [28], Kramer et. al. considered the evolution of local graph features including betweenness centrality to explain the coupling of brain signals at seizure onset [11],

and Douw et. al. recently showed epilepsy in glioma patients was attributed to the theta band activity in the brain [6]. In [19] authors independently suggest a similar approach that combines tensor decompositions with graph theory. In this paper, we continue studying synchronization graphs and introduce new features as the early indicators of a seizure onset.

Since the EEG signals are continuously recorded, they 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 predict future values, specifically in this case to predict the epileptic seizure, while minimizing the error between the actual value and the predicted value. *Autoregressive* models (AR) are commonly used tools for time-series prediction, and have been used to capture the spatio-temporal properties of EEG signals [2,32]. We further improve on the AR model by using *Transfer Learning* [26] 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 obtained from a noisy experiment, or partial data, can be supplemented by clean data from a different experiment.

The two main contributions of this work are as follows: (i) bridging the two concepts of AR modeling and synchronization graphs by constructing an AR(1) model on the features extracted from the time-evolving EEG synchronization graphs, and (ii) introducing the concept of transfer 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, procedure to construct EEG synchronization graphs and extract features from the graph, feature selection method based on quadratic programming, transfer learning on the features, and using anomaly detection as a means for seizure prediction. In Section 3 we present the results for seizure prediction. We discuss the results in Section 3, and provide an overview and list possible extensions to this study, in Section 4.

2 Methodology

2.1 Epileptic EEG Data Set

Our dataset consists of scalp EEG recordings of 26 seizures from 11 patients. All the patients were evaluated with scalp video-EEG monitoring in the international 10-20 system (as described in [10]), magnetic resonance imaging (MRI), fMRI for language localization, and position emission tomography (PET). All the patients had *Hippocampal Sclerosis* (HS) except one patient (IY) with *Cortical Dysplasia* (CD). After selective amygdalohippocampectomy, all the patients were seizure free. The patient information is provided in Table 1. For 4 patients, the seizure would onset from the left, whereas for 7 patients the seizure would onset from the right.

One patient has one 30 minute recording, two patients have two 30 min recordings, one patient has three 30 min recordings, one patient has single 60 minute recording, three patients have two 60 minute recordings, two patients have three 60 minute recordings, and one patient has five 60 minute recordings. 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 [29]. Seizures were 73.81 seconds long on average and their standard deviation was 52.42 seconds.

2.2 Initial Noise Removal – Reduced Rank Approximation

As an initial processing step, we remove noise from the EEG signal by creating a reduced rank approximation of the signal. We start by computing the Singular Value Decomposition (SVD) of the EEG signal. Let $\mathbf{X}[i, m]$ denote the recorded EEG signal, where $i \in \{1, \ldots, 18\}$ represents the index for the ith electrode and $m \in [1, \ldots, f_s \times M]$ represents the time index, f_s represents the sampling frequency, and M is the duration of the recording in seconds. Sampling frequency, f_s, is either 200 Hz or 400 Hz. We find the number of singular vectors required to represent 95% of the variance in the data. This reduced number of singular vectors is taken as the reduced rank r of the EEG signal. We then construct the reduced rank EEG signal X_r by post-multiplying with the product of the right singular vector V and its transpose: $X_r = X \times (V_r \times V_r^T)$.

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

Patient	Pathology	Lateralization
IY	CD	R
OB	HS	R
BMI	HS	R
FZE	HS	R
GSE	HS	L
IP	HS	L
MSO	HS	L
ABA	HS	L
DAK	HS	L
NT	HS	L
SUL	HS	L

2.3 Construction of EEG Synchronization Graphs

For the signal $\mathbf{X}[i, m]$, we constructed 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.

Epoch length selection is an important criterion - longer epoch lengths can lead to better frequency resolution while not being able to capture instantaneous changes to the graph [15,16,33]. Very short epoch lengths might lead to issues related to generation and mining of sparse graphs. Thus, a compromise is necessary. We use an epoch length of 5 seconds. These issues can be mitigated via use of wavelet-domain-based measure for constructing the time-evolving graphs. However, wavelet domain techniques fall short in exploring the entire range of the EEQ frequency spectrum as the range of frequencies must be specified prior to the analysis.

A sample time-evolving graph on an EEG recording is shown in Fig. 1. 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 $i, k \in \{1, \ldots, 18\}$, and $i \neq k$, for the nth epoch, given as $d^n_{i,k}$ 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.

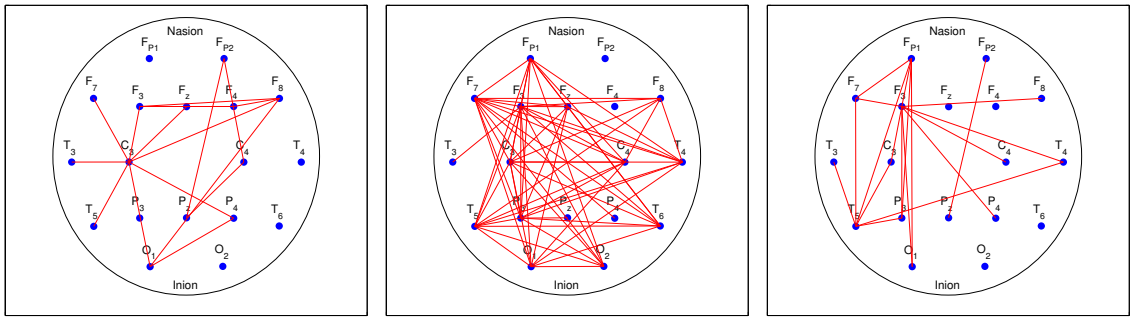
Several synchronization measures have been proposed as plausible options for $d^n_{i,k}$. Based on earlier results (under review), we chose Phase Lag Index [31]. PLI is defined as follows:

$$PLI_{i,k}(n) \quad = \quad \frac{1}{Lf_s} \left| \sum_{m=1}^{Lf_s} \text{sgn}\left(\boldsymbol{\phi}^n_i(m) - \boldsymbol{\phi}^n_k(m)\right) \right| \tag{1}$$

where $\boldsymbol{\phi}^n_i = \arctan(\frac{\hat{\mathbf{x}}^n_i}{\mathbf{x}^n_i})$ is the angle of the Hilbert transform $\hat{\mathbf{x}}^n_i$ of the signal $\mathbf{x}^n_i$.

2.4 Feature Extraction from EEG Synchronization Graphs

We extract 26 features from the EEG 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 eigenvector of the product of the adjacency matrix and the inverse of the degree matrix, and the second largest eigenvalue of the Laplacian matrix.



(a) Pre-ictal (b) Ictal (c) Post-Ictal

Figure 1: **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.

We also extracted *********NIVAS PLACE-HOLDER FOR CHANGE-POINT DETECTION FEATURES, REFER TABLE FOR DESCRIPTION OF FEATURES**

In subsequent text, we refer to the feature matrix as D. A complete list of the features used in this work and their definitions is listed in Table 2. For further information regarding the features, please refer to [3].

2.5 Transfer Learning on Feature Data

We apply an autoregressive model of order 1 (AR1) to our $epochs \times features$ data. For an AR(1) model the output at time t is only dependent on the values of the time-series at time $t-1$. Expressed differently, an AR(1) model is a Markov chain where the next state is dependent on the current state, and none of the states that preceded the current state. We define every input variable z_i at time t in terms of all the input variables at time $t-1$ as follows:

$$z_i^t = \rho_{0i} + \rho_{1i} z_1^{t-1} + \rho_{2i} z_2^{t-1} + \ldots + \rho_{mi} z_{mt} + \epsilon_t \tag{2}$$

where ρ_{ij} are the linear regression weights assigned to the lagged time-series input variables. In matrix format, we would write this as $[Z]_1^t = \mathrm{P} * [1, Z]_0^{t-1}$, where the notation $[A]_a^b$ is the range of time-steps for A. Thus, using an AR(1) model we can come up with the best estimate $\hat{\mathrm{P}}$ for P. We further improve our estimates for the future values predicted by an AR(1) model by incorporating concepts from transfer learning.

Since our data set consists of multiple EEG patient recordings, we can utilize this corpus to predict future values of the features for a particular recording. Given a feature data set $D = z_i^t$ and the feature transfer set, $D_1 = \hat{z}_i^t$, where z_i^t are the input variables, let the autoregressive parameters for D and D_1 be P and P_1 respectively. We can then solve an objective function of the form

$$argmin\{||[Z]_{t-1}^0 * \mathrm{P} - [Z]_t^1||_F^2 + \lambda ||\mathrm{P} - \mathrm{P}_1||_F^2\}\ , \tag{3}$$

where λ is the Tikhonov regularizer and decides the contribution of the transfer set in improving the estimates of future values predicted by D. The first term is the objective function for autoregression, and the second term is the additional objective function for the transfer learning. As more data is obtained for 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 D into training (TR) and testing (TE) sets. Within the training set, we create a further split thereby creating a training prime (TR') and validation (Val) sets. The idea behind this split is to train our AR(1) model on TR', and then use the transfer set to test on Val. Then, we retrain the model using the learned parameters on the entire training set TR and finally test on TE.

2.6 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, with respect to the phenomenon of epileptic seizure, certain features may be redundant, or low in

Table 2: Description of EEG global graph features.

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 *eccentricity* 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 *isolated node* has a degree 0
12	% of End Points	% of endpoints in the graph, where an *endpoint* 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 change 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

predictive importance. Therefore, we quantify the predictive significance of the features in a natural but effective way, and score the features to maximize predictive importance and minimize redundancy using the method in [27], which we summarize here.

2.6.1 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, D, spanning t epochs, is $t \times n$, with n features, $z_i, 1 \leq i \leq n, z_i \in \mathbb{R}^t$. We define, the correlation matrix, $Q \in \mathbb{R}^{n,n}$, element-wise, where $Q(i,j)$ is the pearson correlation coefficient between the feature vectors z_i and z_j:

$$Q(i,j) = \frac{z_i^T z_j}{\|z_i\| \, \|z_j\|}.$$

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

2.6.2 Measuring Predictive Importance

We first recall that the activity of the brain at time t is completely captured by d^t. We define the predictive importance, f_i, of a feature, z_i, as the r.m.s. influence of z_i^t on $z_j^{t+1}, 1 \leq j \leq n$, measured by the coefficients in the forecast

operator corresponding to z_i. Formally, let $\Psi \in \mathbb{R}^{n,n+1}$ be the forecast operator. Then our best prediction of d^{t+1} is $\tilde{d}^{t+1}$ where

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

The influence, $p_i(j)$, of feature z_i on z_j, contained in $p_i \in \mathbb{R}^n$, may be determined by predicting via Ψ using its indicator vector, e_i:

$$p_i = e_i \Psi^T,$$

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

$$f_i = \|p_i\| = \|\Psi_i\|,$$

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

2.6.3 Optimizing Redundancy and Predictive Importance

We note that we may indicate features that maximize predictive importance and minimize redundancy by solving

$$\bar{x} = \arg\min x^T Q x - f^T x; \quad x \geq 0, \quad \sum_i x_i = 1,$$

where the constraints arise from forcing the resulting vector to be a distribution, from which we may omit an appropriately sized tail, should we choose to do so. However, the magnitudes of entries in Q, f may be vastly different for very similar data - for example, a data matrix aD would have the same correlation matrix as D, but its predictive importance vector would be a times as large. To make the objective function scale invariant, we normalize f to obtain

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

Finally, to effect a meaningful trade-off between minimizing redundancy and maximizing predictive importance, we take a convex combination of the objective functions in both optimization problems:

$$q(x) = (1-\alpha) x^T Q x - \alpha f^T x,$$

where α is chosen, as in [27] as

$$\alpha = \frac{\sum_{i,j} Q(i,j)/n^2}{\sum_{i,j} Q(i,j)/n^2 + \sum_k f_k/n}.$$

We solve the resulting optimization problem:

$$x^* = \arg\min q(x); \quad s.t. x \geq 0, \quad \sum_i x_i = 1$$

to obtain an importance-distribution over the features.

2.7 Designing Ictal Discriminators

In order to separate ictal activity from regular activity, we obtain a parameter which maximally distinguishes between the two. To this end, we define a discriminator as follows.

Definition 2.1. Let Y be a domain with a region of interest, $R \subset X$, and let $f : Y \to \mathbb{R}$ be a function. We call f a δ-discriminator of R from X, iff

$$\Delta(f, R) = \frac{|\langle f(R) \rangle - \langle f(X \setminus R) \rangle|}{\langle f(X) \rangle} \geq \delta,$$

where $\langle f(S) \rangle$ is the average of the function restricted to S. We call f an (ρ, δ)-discriminator of R if f is a δ-discriminator of an ρ-expansion of R in X.

For predicting seizure, we use this notion of discrimination to construct a forecast operator, Ψ, for the feature vector, d_t, at time t of recorded brain activity, with the property that the prediction errors, ϵ_t, of the forecast form a δ-discriminator of the ictal period for an appreciably large δ. We achieve this by designing a forecast operator specifically to predict regular function with high fidelity, at the cost of predicting ictal activity accurately. Firstly, since approximately

$97 - 98\%$ of the data is non-ictal activity, AR(1) preferentially learns non-ictal activity over ictal activity, and the predictive importance vector may be expected to weight features that discriminate minimally between these activities. For the same reason that ictal activity resides in a small region of time, we may expect the correlation matrix to be minimally perturbed by ictal activity, resulting in the features selected by QPFS being good predictors preferential to normal function. We exploit this preference to predict normal function by restricting the predictor, Ψ, to just these features. Suppose that the features chosen in accordance with [QPFS] are $\{c_j\}_1^k \subset \{z_i\}_1^n$. We restrict the predictor, $\Psi \in \mathbb{R}^{n,n+1}$, to just these selected features to obtain $\hat{\Psi} \in \mathbb{R}^{n,k+1}$, the prediction using which is designed to bias *against* good prediction of ictal activity, and therefore incur larger errors therein. We refer the reader to the results for the discrimination parameters, ρ, δ, obtained via this manner.

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 learning.

3.1 Quadratic Programming Feature Selection Results

We ran the QPFS algorithm on the data sets, and selected all the features that were returned by the algorithm as discriminating features. These features are the following (i) Average Degree, (ii) Diameter of graph, (iii) Radius of graph, (iv) Average Path Length, (v) Giant Connected Component Ratio, (vi) Number of Connected Components, (vii) Percentage of Isolated Points, (viii) Laplacian Trace, (ix) Mean of change in EEG signal.

3.2 Best Predictor Operator for Seizure Prediction

3.3 Autoregression vs transfer learning

4 Conclusions and Future Work

Mention: advanced noise removal transfer learning based on similarity between patients

References

[1] E. Acar, C. A. Bingol, B. H., and Y. B. Computational analysis of epileptic focus localization. In *Fourth IASTED Int. Conf. on Biomedical Engineering*, pages 317–322, 2006.

[2] N. R. Anderson, K. Wisneski, L. Eisenman, D. W. Moran, E. C. Leuthardt, and D. J. Krusienski. An offline evaluation of the autoregressive spectrum for electrocorticography. *IEEE Trans Biomed Eng*, 56(3):913–916, 2009.

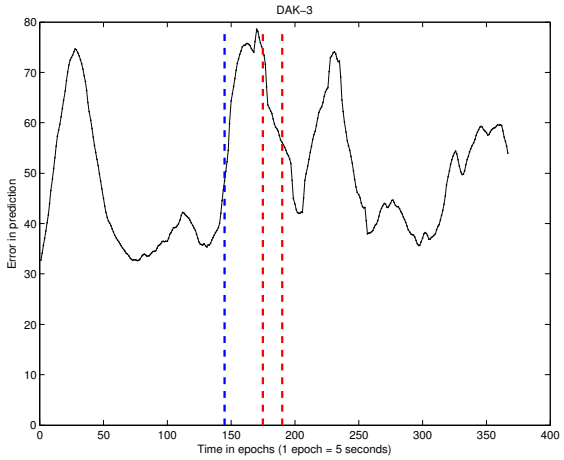
[3] C. C. Bilgin, S. Ray, B. Baydil, W. P. Daley, M. Larsen, and B. Yener. Multiscale feature analysis of salivary gland branching morphogenesis. *PLoS ONE*, 7(3):e32906, 2012.

[4] J. D. Bronzino. *Biomedical Engineering Handbook*, chapter Principles of electroencephalography. New York: Taylor and Francis, 3rd edition, 2006.

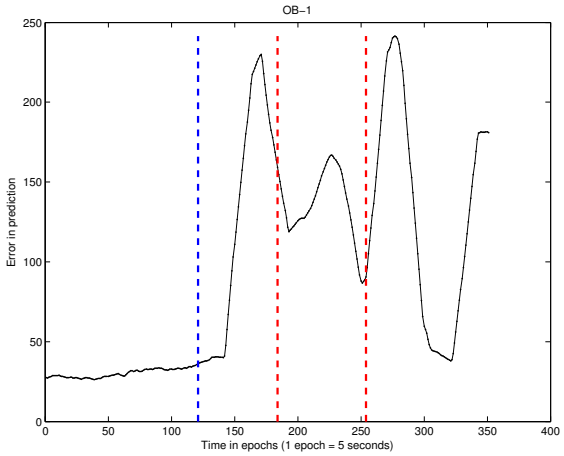
[5] E. Bullmore and O. Sporns. Complex brain networks: graph theoretical analysis of structural and functional systems. *Nature Reviews Neuroscience*, 10(3):186–198, 2009.

[6] L. Douw, E. van Dellen, M. de Groot, J. J. Heimans, M. Klein, C. J. Stam, and J. C. Reijneveld. Epilepsy is related to theta band brain connectivity and network topology in brain tumor patients. *BMC Neuroscience*, 11(1):103, 2010.

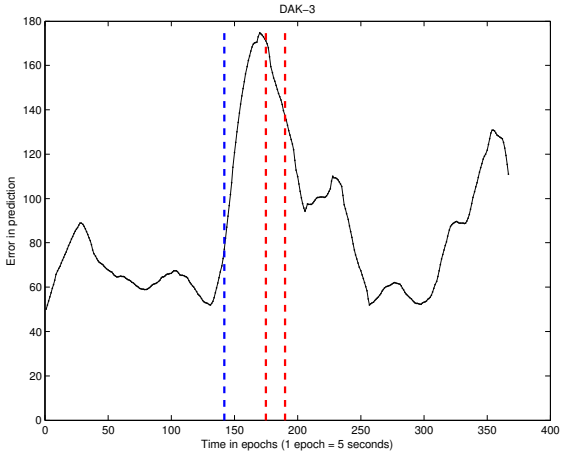
[7] R. S. Fisher, W. van Emde Boas, W. Blume, C. Elger, P. Genton, P. Lee, and J. J. Engel. Epileptic seizures and epilepsy: definitions proposed by the international league against epilepsy (ilae) and the international bureau for epilepsy (ibe). *Epilepsia*, 46(4):470–472, 2005.



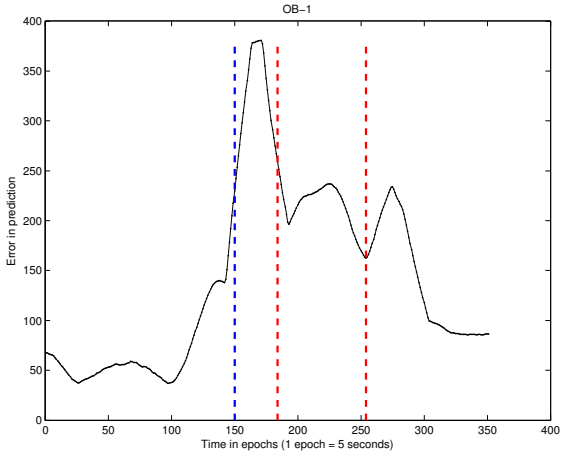
(a) $n \times (n+1)$ Forecast Operator



(b) $n \times (n+1)$ Forecast Operator

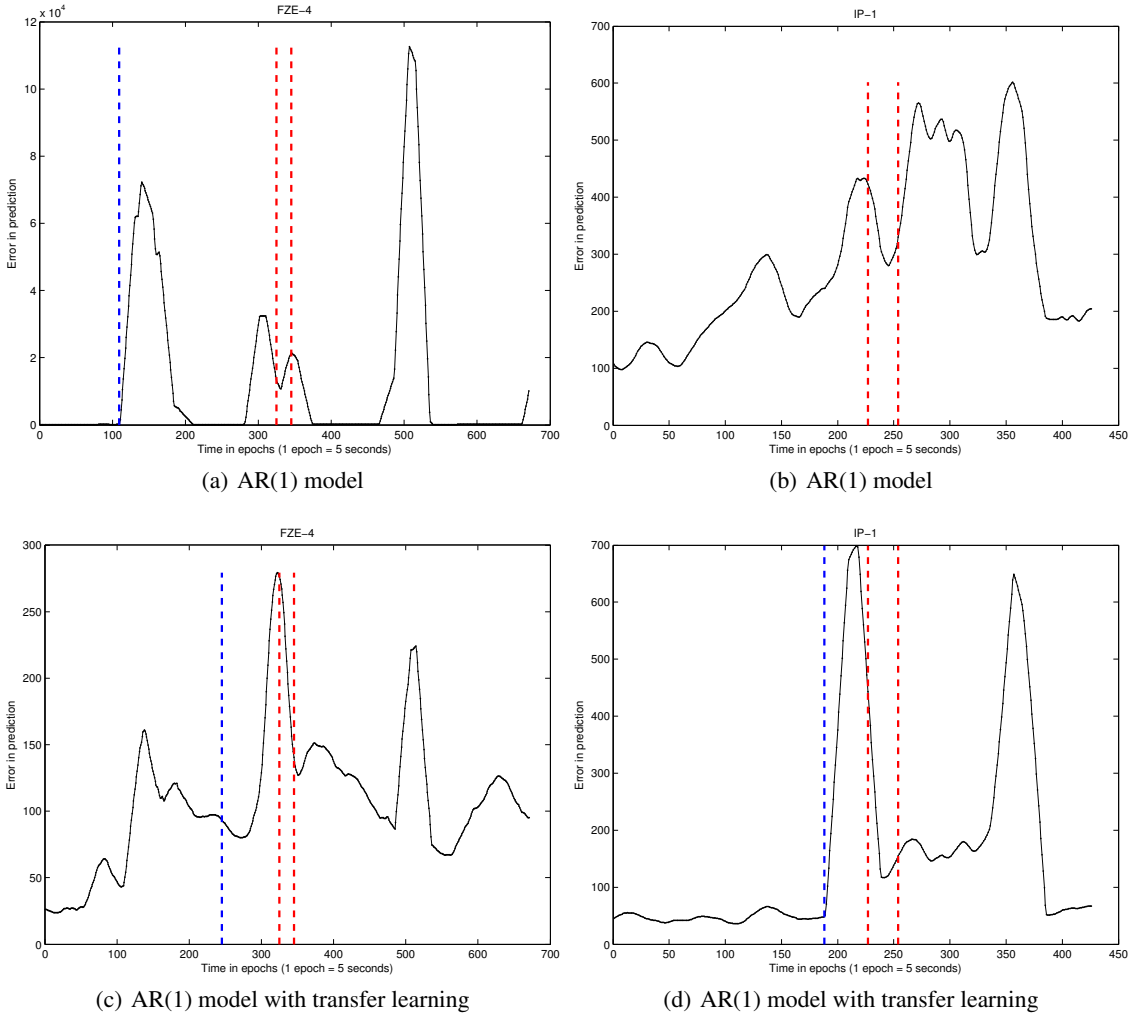


(c) $n \times 10$ Forecast Operator



(d) $n \times 10$ Forecast Operator

Figure 2: Comparison of the $n \times 10$ and $n \times (n+1)$ forecast operators. The epoch at which the seizure is detected is shown in blue, the extremities of the seizure region are marked in red.



(a) AR(1) model

(b) AR(1) model

(c) AR(1) model with transfer learning

(d) AR(1) model with transfer learning

Figure 3: Comparison of the AR(1) model with the AR(1) model enhanced by transfer learning. The epoch at which the seizure is detected is shown in blue, the extremities of the seizure region are marked in red.

[8] R. S. Fisher, B. G. Vickrey, P. Gibson, B. Hermann, P. Penovich, A. Scherer, and S. Walker. The impact of epilepsy from the patient's perspective I. descriptions and subjective perceptions. *Epilepsy Research*, 41(1):39–51, 2000.

[9] N. Hazarika, J. Z. Chen, A. C. Tsoi, and A. Sergejew. Classification of eeg signals using the wavelet transform. *Signal Process.*, 59:61–72, 1997.

[10] H. H. Jasper. The ten-twenty electrode system of the international federation. *Electroencephalogr Clin Neurophysiol Suppl.*, 10:371–375, 1958.

[11] M. A. Kramer, E. D. Kolaczyk, and H. E. Kirsch. Emergent network topology at seizure onset in humans. *Epilepsy Research*, 79(2):173–186, 2008.

[12] C. L., M. A., P. G., S. M., A. C., C. G., and F. Fuggetta. Real-time epileptic seizure prediction using ar models and support vector machines. *IEEE Trans Biomed Eng*, 57(5):1124–1132, May 2010.

[13] Q. M. Le Van, V. Navarro, J. Martinerie, M. Baulac, and V. F. J. Toward a neurodynamical understanding of ictogenesis. *Epilepsia*, 44(12):30–43, 2003.

[14] Q. M. Le Van, J. Soss, V. Navarro, R. Robertson, M. Chavez, M. Baulac, and J. Martinerie. Preictal state identification by synchronization changes in long-term intracranial EEG recordings. *Clinical Neurophysiology*, 116:559–568, 2005.

[15] W. J. Levy. The effect of epoch length on the identification of changes in EEG power spectra. *Anesthesiology*, 65(3A):A539, 1986.

[16] W. J. Levy. The effect of epoch length on the identification of changes in EEG power spectra. *Anesthesiology*, 66(4):489–495, 1987.

[17] H. S. Liu, T. Zhang, and F. S. Yang. A multistage, multimethod approach for automatic detection and classification of epileptiform EEG. *IEEE Trans Biomed Eng*, 49(12 Pt 2):1557–1566, 2002.

[18] W. W. Lytton. Computer modeling of Epilepsy. *Nat. Rev. Neurosci.*, 9(8):626–637, Aug. 2008.

[19] A. Mahyari and S. Aviyente. Identification of dynamic functional brain network states through tensor decomposition. In *39th IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2014)*, 2014.

[20] P. Mirowski, D. Madhavan, Y. LeCun, and R. Kuzniecky. Classification of patterns of EEG synchronization for seizure prediction. *Clinical Neurophysiology*, 120:1927–1940, 2009.

[21] F. Mormann, R. G. Andrzejak, C. E. Elger, and K. Lehnertz. Seizure prediction: the long and winding road. *Brain*, 130:314–333, 2007.

[22] F. Mormann, T. Kreuz, C. Rieke, R. Andrzejak, A. Kraskov, P. David, C. Elger, and K. Lehnertz. On the predictability of epileptic seizures. *Clinical neurophysiology*, 116(3):569–587, 2005.

[23] S. Murali and V. V. Kulish. Modeling of evoked potentials of electroencephalograms: An overview. *Digit. Signal Process.*, 17:665–674, 2007.

[24] J. Muthuswamy and N. V. Thakor. Spectral analysis methods for neurological signals. *J. Neurosci. Methods*, 83:1–14, 1998.

[25] N. Päivinen, S. Lammi, A. Pitkänen, J. Nissinen, M. Penttonen, and T. Grönfors. Epileptic seizure detection: A nonlinear viewpoint. *Computer methods and programs in biomedicine*, 79(2):151–159, 2005.

[26] S. J. Pan and Q. Yang. A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering*, 22(10):1345–1359, Oct. 2010.

[27] I. Rodriguez-Lujan, R. Huerta, C. Elkan, and C. S. Cruz. Quadratic programming feature selection. *J. Mach. Learn. Res.*, 11:1491–1516, 8 2010.

[28] K. A. Schindler, S. Bialonski, M.-T. Horstmann, C. E. Elger, and K. Lehnertz. Evolving functional network properties and synchronizability during human epileptic seizures. *CHAOS: An Interdisciplinary Journal of Nonlinear Science*, 18(3):033119, 2008.

[29] S. J. M. Smith. EEG in the diagnosis, classification, and management of patients with epilepsy. *J Neurol Neurosurg Psychiatry*, 76:ii2–i–i7, 2005.

[30] C. Stam and E. van Straaten. The organization of physiological brain networks. *Clinical Neurophysiology*, 2012.

[31] C. J. Stam, G. Nolte, and A. Daffertshofer. 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.

[32] A. Subasi, A. Alkan, E. Koklukaya, and M. K. Kiymik. Wavelet neural network classification of EEG signals by using ar models with mle processing. *Neural Networks*, 18(7):985–997, 2005.

[33] M. van de Velde, I. R. Ghosh, and P. J. M. Cluitmans. Context related artefact detection in prolonged EEG recordings. *Computer Methods and Programs in Biomedicine*, 60(3):183–196, 1999.

[34] M. J. A. M. van Putten, T. Kind, F. Visser, and V. Lagerburg. Detecting temporal lobe seizures from scalp EEG recordings: a comparison of various features. *Clinical Neurophysiology*, 116(10):2480–2489, 2005.