A Combined Linear & Nonlinear Approach for Classification of Epileptic EEG Signals

Tugce Balli and Ramaswamy Palaniappan
School of Computer Science and Electronic Engineering
University of Essex
Colchester, United Kingdom
<tballi, rpalan>@essex.ac.uk

Abstract—**The use of both linear autoregressive model coefficients and nonlinear measures for classification of EEG signals recorded from healthy subjects and epilepsy patients is investigated. A total of seven nonlinear measures namely the approximate entropy, largest lyapunov exponent, correlation dimension, nonlinear prediction error, hurst exponent, third order autocovariance, asymmetry due to time reversal, are used in this study. The class separability of individual and combined feature sets is measured using Linear Discriminant Analysis (LDA) algorithm where the multiple features are selected by sequential floating forward search (SFFS) algorithm. The results have shown that the use of combined feature sets provide a better characterization of EEG signals compared to individual features.**

Keywords-component; EEG; Linear Autoregressive Model, Nonlinear Complexity Measures; State Space Reconstruction

I. INTRODUCTION

Today the analysis of electroencephalogram (EEG) signals is an actively pursued research area for diagnosing certain pathological conditions like epilepsy, memory impairments and sleep disorders [1-4].

Traditionally, the analysis of EEG signals have been approached using linear methods (i.e. amplitude and frequency dependent measures). Although fairly good results have been obtained using such methods, they provide an approximation to the underlying nonlinear properties of these signals. Hence the use of nonlinear analysis methods for the analysis of EEG signals would be a logical step in obtaining an improved characterization of these signals.

During the last decade a variety of nonlinear measures like Largest Lyapunov Exponent, Correlation Dimension and Entropy measures have been repeatedly applied to different kinds of EEG signals, such as recordings from patients with diseases like epilepsy, Alzheimer's, Parkinson's, schizophrenia and in comparison against healthy control subjects [1-4]. The general aim of these studies is to investigate the applicability of the nonlinear measures for analyzing different physiological brain states for diagnostic purposes. The findings in these studies suggest that the nonlinear dynamics can provide new information for understanding the brain dynamics and further characterization of the EEG signals.

In this study we have investigated the use of both linear autoregressive model coefficients and nonlinear complexity measures for classification of the EEG signals recorded from healthy subjects and epilepsy patients. We have utilized a total of five nonlinear complexity measures namely the approximate entropy, correlation dimension, largest lyapunov exponent, nonlinear prediction error and hurst exponent. Apart from these two other nonlinear measures namely the third order autocovariance and asymmetry due to time reversal have also been included.

II. DATA SET

In this study we have used an EEG data set comprised of five classes (A-E) each containing 100 single channel EEG segments. The EEG segments were recorded for 23.6 seconds with a sampling rate of 173.61Hz.

The sets A and B consisted of EEG segments recorded extracranially from healthy subjects with eyes open and closed respectively. Set C, D and E recorded intracranially from epilepsy patients where C and D are recorded during seizure free interval and E is recorded during seizure interval. Segments in set C were recorded from the hippocampal formation of the opposite hemisphere of the brain while the segments in set D were recorded from the epileptogenic zone. This data set is publicly available at [5].

III. METHODOLGY

A. Nonlinear Measures

The first step in estimation of nonlinear complexity measures is the state space reconstruction. The state space reconstruction is used to transform the univariate data (i.e. single channel EEG) to its trajectory in a multidimensional state space using Takens method of time delay embeddings [6].

The dynamical properties of trajectories in the reconstructed state space can be quantified by nonlinear measures such as correlation dimension, largest lyapunov exponent, approximate entropy, nonlinear prediction error.

1) State Space Reconstruction

Suppose that a single scalar measure, $x(t)$, can be measured at a time from the system using an observation function g such that; $x(t) = g(\mathbf{s}(t))$, $g : M \rightarrow R$ and $t = 1, 2, \ldots, N$. The observation function cannot provide a complete representation of the underlying properties of the dynamical system. According to Takens theorem this can be achieved by representing time series as time lagged versions of itself such that;

$$f : M \to R^d \tag{1}$$

$$s(t) \to y(t) = f(s(t)) = [x(t), x(t-\tau), x(t-2\tau), ... , x(t-(m-1)\tau)] \tag{2}$$

$$S = \begin{bmatrix} x(1) & x(1+\tau) & \cdots & x(1+(m-1)\tau) \\ x(2) & x(2+\tau) & \cdots & x(2+(m-1)\tau) \\ \vdots & \vdots & \ddots & \vdots \\ x(N-(m-1)\tau) & x(N-(m-1)\tau+1) & \cdots & x(N) \end{bmatrix}$$

where τ is the time lag, m is the embedding dimension and S is the complete representation of the single scalar time series in state space.

The selection of embedding dimension, m, and time lag, τ, parameters is important to achieve a good reconstruction of time series in state space. In this study we have used false nearest neighbors method and first zero crossing of the autocorrelation function for selection of embedding parameters [7,8]. The results suggested that an embedding dimension, m, of 5 and time lag, τ, of 3 is appropriate for state space reconstruction.

2) Approximate Entropy

Approximate entropy (ApEn) is a method that measures the irregularity of time series where higher values of ApEn indicate more complexity and randomness [9].

Once the state space is reconstructed, the ApEn is calculated as follows;

$$C_i^m(r) = \sum_{j=1}^{Nv} \frac{\Theta(r - \|y_i - y_j\|)}{Nv} \tag{3}$$

where C_i^m is the correlation integral, Nv is the number of delay vectors which is set to $N-(m-1)*\tau$ and N is the number of data points in time series, y_i and y_j are embedding vectors in state space and r is the tolerance of comparison value.

Note that $\Theta(x)$ is Heaviside function such that $\Theta(x)=1$ for $x>0$ and $\Theta(x)=0$ for $x\le 0$. The ApEn is obtained by,

$$ApEn\ (m, \tau, r) = \Phi^m(r) - \Phi^{m+1}(r) \tag{4}$$

where

$$\Phi^m(r) = \frac{1}{Nv} \sum_{i=1}^{Nv} \ln\left[C_i^m(r)\right] \tag{5}$$

3) Largest Lyapunov Exponent

Largest Lyapunov exponent (LLE) is another nonlinear measure that quantifies the complexity of the time series. This measure estimates the mean exponential divergence or convergence of nearby trajectories in state space in which the sensitive dependence on initial conditions is obtained. There are several algorithms for calculation of LLE measure. Here we have used Rosenstein's algorithm [7,8].

$$L_k = \frac{1}{Nv - k} \sum_{j=1}^{Nv-k} \ln \cdot \|y_i - y_j\| \tag{6}$$

where Nv is the number of delay vectors in state space, m is the embedding dimension, y_i and y_j are embedding vectors in which y_i is the closest embedding vector to y_j.

This algorithm simply looks for nearest point y_i to each y_j in m-dimensional state space and the average logarithmic separation of two points is calculated for next k time steps. The LLE is given by;

$$\lambda_1 = \frac{dL_k}{dk} \tag{7}$$

4) Correlation Dimension

Correlation dimension (CD) measures the number of independent variables that is necessary to describe the dynamics of the system. The CD of EEG signals are generally considered to reflect the complexity of the underlying dynamics of these signals. Here we have calculated the CD of EEG signals using Grassberg and Proccacia algorithm [7,8]. In order to calculate CD, the correlation integral function, which is defined in Eq.3, must be estimated. A plot of log $C(r)$ versus $log\ r$ should give an approximately straight line whose slope, in the limit of small r and large N, is the correlation dimension [7,8].

$$CD = \lim_{r\to 0} \lim_{N\to\infty} \frac{d \ln C(r)}{d \ln r} \tag{8}$$

5) Nonlinear Prediction Error

The nonlinear prediction error (NLPE) exploits the deterministic structure in reconstructed state space. Higher NLPE values indicate more complexity and randomness of the time series while lower NLPE indicates less complexity [7,8].

The nonlinear prediction error is calculated such that for each embedding vector y_i in the state space, k nearest neighbors, y_j , $j=1,2,...,k$ is used to perform T step ahead prediction of y_{i+T} . In order to do the prediction a linear model is fitted to k nearest neighbors y_j and their T step ahead values y_{j+T}, j=1,2,…,k. Following this, the fitted model was used to estimate y_{i+T}. The difference between the actual y_{i+T} and predicted y_{i+T} is the nonlinear prediction error. In this study we have fixed the value of T to 1 and k to 10.

6) Hurst Exponent

The Hurst exponent (HURST) which is also called Rescaled Range statistics (R/S) is used for quantifying the correlation of points in a time series. HURST=0.5 for random data, HURST>0.5 for data with long range correlations and HURST<0.5 for data with long range anticorrelations [8].

In order to calculate the R/S statistic for time series $x(t)$ with t=1,2,…,N, one has to calculate the accumulated deviation from the mean of time series over time T such that;

$$X(t,T)=\sum_{i=1}^{t}x(i)-\bar{x} \quad \text{where} \quad \bar{x}=\frac{1}{T}\sum_{i=1}^{T}x(t) \tag{9}$$

Then *R(T)* is calculated as the difference between the maximum and minimum value of *X(t,T)* and *S(T)* is calculated as the standard deviation of time series over time *T*;

$$\frac{R(T)}{S(T)}=\frac{(\max(X(t,T))-\min(X(t,T)))}{\sqrt{\frac{1}{T}\sum_{t=1}^{T}(x(t)-\bar{x})^2}} \tag{10}$$

The Hurst exponent is obtained by calculating the slope of the line produced by *ln(R(n)/S(n))* versus *ln(n)*.

7) Third Order Autocovariance

The third order autocovariance is a higher extension of autocovariance method that measures the dependence of a data point on previous data points [4]. This measure is given by,

$$C3(\tau)=\frac{1}{N-2\tau}\sum_{k=2\tau+1}^{N}(x(k)*x(k-\tau)*x(k-2\tau)) \tag{11}$$

where N is the length of time series and τ is the time lag.

8) Assymetry Due To Time Reversal

A time series is said to be reversible if its properties are invariant with respect to time reversal [4]. This measure is given by;

$$REV(\tau)=\frac{1}{N-1-\tau}\sum_{k=1+\tau}^{N}(x(k)-x(k-\tau))^3 \tag{12}$$

where N is the length of time series and τ is the time lag.

B. Linear Measures

Here we have used the conventional linear autoregressive (AR) model for analysis of EEG signals. In the literature, this method has been widely used for classification of mental task EEG segments [10,11].

1) Linear Autoregressive Model

An AR model of order p is defined by;

$$x(t)=-\sum_{k=1}^{p}a_k x(t-k)+e_t \tag{13}$$

where *x(t)* is data at sample point *t*, a_k are coefficients of the AR model and e_t is Gaussian white noise with mean zero.

C. Feature Extraction And Classification

The EEG signals were divided into 8 non-overlapping segments with 512 data points for feature extraction. The lengths of segments were chosen according to weak stationarity criterion [3].

The linear AR modeling involves selection of model order, here the coefficients are extracted from EEG segments for order 2 to 10 and the optimal model order is selected according to classification results.

Apart from that the ApEn method involves selection of tolerance of comparison parameter, *r*. In previous studies it is reported that the *r* value of 0.2 times standard deviation of time series provides a good statistical validity, however in our studies we found out that feature separability might be improved using values other than 0.2 [12]. Therefore we have extracted ApEn features with *r* parameter coefficients ranging from 0.2 to 5 with increments of 0.2. The best *r* coefficient value is selected according to classification results. We have also investigated that the ApEn values estimated with different *r* values are complementary to each other, providing an improved characterization of EEG time series [12].

For classification of EEG segments, a total of three cases are considered: five-class case, simplified three class case where sets A and B, and sets C and D are treated as one class, and simplified two-class case where only sets A, B, C and D are considered in which sets A and B, and sets C and D are treated as one class.

The separability of individual features are assessed using linear discriminant analysis (LDA) method with 10 fold cross validation. Furthermore the separability of combined feature sets, where both linear and nonlinear methods are utilized, are also investigated using sequential floating forward search with linear discriminant analysis algorithms (SFFS-LDA) [13]. The SFFS algorithm works by combining individual features that are complementary to each other, thus providing an improved separability for different classes of data. The SFFS algorithm starts with an empty feature set and in each iteration the feature that provides highest improvement in the classification accuracy is added to the feature set (i.e. sequential forward selection). The algorithm also performs sequential backward selection and in each iteration tests if removal of some features from the created feature set will result in an increase in accuracy and eliminate the corresponding features from the feature set. The algorithm iterates until the desired number of features, which is set to 10 in this study, is reached in the combined feature set. During feature selection, each set of feature was trained and tested on the data with 10 fold cross validation and the mean accuracy from all folds was used for assessing the separability of combined feature sets.

IV. RESULTS AND DISCUSSIONS

The results of LDA classification of healthy and epileptic EEG segments for 5 Class, 3 Class and 2 Class cases are shown in Table 1. The results demonstrate that the best classification accuracies of 74.77% for 5 class case, 95.60 % for 3 class case and 98.28% for 2 class case, are obtained with 10^{th} order AR coefficients while the nonlinear measures didn't perform as good as AR coefficients. The poor performance of these measures can be attributed to the fact that with 10^{th} order AR model the feature vectors are created with 10 features (i.e. coefficients) where the feature vectors for nonlinear features are created with only one feature. Therefore, in order to do a fair comparison we have utilized SFFS-LDA algorithm for selecting 10 features from the nonlinear features. The results of nonlinear feature selections are demonstrated in Table 2 and the selected features are shown in Table 3.

TABLE I. THE RESULTS OF LDA CLASSIFICATION FOR INDIVIDUAL FEATURES

Classification Accuracy (%)	***5 Class***		***3 Class***		***2 Class***	
C3	29.82		46.20		51.63	
REV	33.17		49.08		52.50	
CD	35.92		56.63		72.84	
LLE	44.70		55.15		77.63	
NLPE	43.77		66.10		71.53	
HURST	40.72		58.72		66.16	
ApEn	**r Coef**	**Accuracy**	**r Coef**	**Accuracy**	**r Coef**	**Accuracy**
	1.2	40.20	1.2	65.80	1.0	90.75
AR	**Order**	**Accuracy**	**Order**	**Accuracy**	**Order**	**Accuracy**
	10	74.77	10	95.60	10	98.28

These results demonstrate that for all cases, the classification accuracies of 10^{th} order AR coefficients are still better but comparable to that of combined nonlinear features.

TABLE II. THE RESULTS OF FEATURE SELECTION AND LDA CLASSIFICATION FOR COMBINED NONLINEAR FEATURES

Classification Accuracy (%)	***5 Class***	***3 Class***	***2 Class***
	71.45	91.60	93.94

TABLE III. THE LIST OF SELECTED FEATURES FOR NONLINEAR FEATURE SELECTION

5 Class	***3 Class***	***2 Class***
r Coefficient 0.4 APEN	NLPE	r Coefficient 0.4 APEN
HURST	r Coefficient 0.4 APEN	LLE
NLPE	LLE	r Coefficient 3.0 APEN
LLE	r Coefficient 3.2 APEN	r Coefficient 1.8 APEN
r Coefficient 0.2 APEN	r Coefficient 0.6 APEN	r Coefficient 0.6 APEN
C3	HURST	C3
r Coefficient 4.8 APEN	CD	HURST
r Coefficient 5.0 APEN	r Coefficient 0.2 APEN	CD
REV	r Coefficient 3.8 APEN	REV
r Coefficient 4.4 APEN	r Coefficient 3.0 APEN	r Coefficient 0.2 APEN

The class separability of these measures is further assessed by using a combined approach where both linear and nonlinear measures are utilised. The results are shown in Tables 4 and 5. These results demonstrate that the performance is improved by 5.7% for 5 class case and 2.63% for 3 class case and 1% for 2 class case when compared to 10^{th} order AR coefficients. And the performance is improved by 9.02% for 5 class case, 6.63% for 3 class case and 5.34% for 2 class case when compared to combined nonlinear features.

TABLE IV. THE RESULTS OF FEATURE SELECTION AND LDA CLASSIFICATION FOR COMBINED LINEAR AND NONLINEAR FEATURES

Classification Accuracy (%)	***5 Class***	***3 Class***	***2 Class***
	80.47	98.23	99.28

TABLE V. THE LIST OF SELECTED FEATURES FOR LINEAR AND NONLINEAR FEATURE SELECTION

5 Class	***3 Class***	***2 Class***
10th order AR	10th order AR	9th Order AR
8th Order AR	LLE	r Coefficient 1.8 APEN
r Coefficient 0.4 APEN	NLPE	2nd Order AR
NLPE	2nd Order AR	6th Order AR
5th Order AR	r Coefficient 1.8 APEN	r Coefficient 0.6 APEN
9th Order AR	8th Order AR	r Coefficient 1.6 APEN
r Coefficient 1.0 APEN	7th Order AR	r Coefficient 2.4 APEN
r Coefficient 1.6 APEN	5th Order AR	r Coefficient 2.8 APEN
r Coefficient 2.0 APEN	C3	r Coefficient 2.2 APEN
C3	r Coefficient 4.8 APEN	r Coefficient 2.0 APEN

V. CONCLUSION

In this study we have investigated the characterization ability of linear AR model coefficients and nonlinear measures. They are tested by assessing their ability to differentiate between different classes of EEG data.

Considering the fact that the research on nonlinear EEG analysis is still at a preliminary level, the results obtained so far are promising. The results have shown that, on a fair basis, the utilized nonlinear measures can perform as good as linear AR coefficients. Moreover, it has also been shown that using a combined approach, where both linear and nonlinear measures are used, provided an improvement to the classification of EEG signals. This demonstrates that the utilised linear and nonlinear measures are independent and complementary features that quantify different properties of the EEG signals, thus providing a further characterization of these signals.

REFERENCES

[1] C.J. Stam, “Nonlinear Dynamical Analysis of EEG and MEG: Review of an Emerging Field”, Clinical Neurophysiology, vol. 116, pp. 2266-2301, 2005.

[2] I.A. Rezek and S.J. Roberts, “Stochastic Complexity Measures for Physiological Signal Analysis”, IEEE Transactions on Biomedical Enginerring, vol. 45, no.9, pp. 1186-1191, 1998.

[3] R.G. Andrzejak, K. Lehnertz, F. Mormann, C. Rieke, P. David, and C.E. Elger, “Indications of Nonlinear Deterministic and Finite-Dimensional Structures In Brain Electrical Activity: Dependence on Recording Region and Brain State”, Physical Review, E64 061907, 2001.

[4] T. Gautama, D.P. Mandic and M.M.V. Hulle, “Indications of Nonlinear Structures in Brain Electrical Activity”, Physical Review, E 67 046204, 2003.

[5] EEG time series are available under: http://www.meb.uni-bonn.de/epileptologie/science/physik/eegdata.htm

[6] F. Takens. “Detecting Strange Attractors In Turbulance”, Dynamical Systems and Turbulance, Lecture Notes in Mathematics, 898:366-381, 1981.

[7] H. Kantz and T. Schreiber. Nonlinear Time Series Analysis. Cambridge University Press, 2003.

[8] J.C. Sprott. Chaos and Time Series Analysis. Oxford University Press, 2003.

[9] S.M. Pincus, “Approximate Entropy As a Measure of System Complexity”, Proc. Natl. Acad. Sci., vol. 88, pp. 2297-2301, 1991.

[10] Z.A. Keiern, J.I. Aunon, “A New Mode of Communication Between Man and His Surroundings”, IEEE Trans. on Biomedical Engineering, vol.37, no.12, 1990.

[11] C.W. Anderson, E.A. Stolz, and S. Shamsunder, “Multivariate Autoregressive Models for Classification of Spontaneous Electroencephalogram During Mental Taks”, IEEE Trans. on Biomedical Engineering, vol.45, no.3, pp.277-268, 1998.

[12] M. Dyson, T. Balli, J.Q. Gan, F. Sepulveda and R. Palaniappan. “Approximate Entropy for EEG-Based Movement Detection”, Proceedings of the 4rd Int. Brain-Computer Interface Workshop and Training Course 2008. Verlag der Techniscen Universität Graz, 2008.

[13] P. Pudil, J. Novovicova, and J. Kittler. “Floating Search Methods in Feature Selection”, Pattern Recognition Letters, 15:1119-1125, 1994.