Mixed-Band Wavelet-Chaos-Neural Network Methodology for Epilepsy and Epileptic Seizure Detection

Samanwoy Ghosh-Dastidar, Hojjat Adeli*, *Member, IEEE*, and Nahid Dadmehr

***Abstract*—A novel wavelet-chaos-neural network methodology is presented for classification of electroencephalograms (EEGs) into healthy, ictal, and interictal EEGs. Wavelet analysis is used to decompose the EEG into *delta*, *theta*, *alpha*, *beta*, and *gamma* sub-bands. Three parameters are employed for EEG representation: standard deviation (quantifying the signal variance), correlation dimension, and largest Lyapunov exponent (quantifying the non-linear chaotic dynamics of the signal). The classification accuracies of the following techniques are compared: 1) unsupervised k-means clustering; 2) linear and quadratic discriminant analysis; 3) radial basis function neural network; 4) Levenberg–Marquardt backpropagation neural network (LMBPNN). To reduce the computing time and output analysis, the research was performed in two phases: band-specific analysis and mixed-band analysis. In phase two, over 500 different combinations of mixed-band feature spaces consisting of promising parameters from phase one of the research were investigated. It is concluded that all three key components of the wavelet-chaos-neural network methodology are important for improving the EEG classification accuracy. Judicious combinations of parameters and classifiers are needed to accurately discriminate between the three types of EEGs. It was discovered that a particular mixed-band feature space consisting of nine parameters and LMBPNN result in the highest classification accuracy, a high value of 96.7%.**

***Index Terms*—Chaos, EEG sub-bands, epilepsy, neural network classification, wavelet.**

I. Introduction

EPILEPSY is a common brain disorder characterized by intermittent abnormal neuronal firing in the brain which can lead to seizures. Ictal brain activity (during a seizure) differs significantly from the activity in the normal state with respect to both frequency as well as pattern of neuronal firing. Further, the spatio-temporal pattern of neuronal firing gradually evolves from a normal state, first to a preictal (interictal) state and then, to an ictal state [1], [2]. Normally, neurons in different parts of the brain fire independent of each other. In contrast, in the interictal state, neurons start firing in multiple parts of the brain in synchronization with the epileptogenic focus which leads to the ictal state. Despite these differences, detection of seizures can be challenging even from a visual inspection of the electroencephalogram (EEG) by a trained neurologist for a variety of reasons such as excessive presence of myogenic artifacts, interference, and overlapping symptomatology of various mental states. Overcoming these obstacles by means of effective and accurate algorithms for automatic seizure detection and prediction can have a far reaching impact on diagnosis and treatment of epilepsy.

Recent attempts to detect seizures from EEG analysis employ two different approaches: 1) examination of the waveforms in the interictal EEG to find events (markers) or changes in neuronal activity such as spikes [3], [4] which may be precursors to seizures and 2) analysis of the nonlinear spatio-temporal evolution of the EEG signals to find a governing rule [1], [2], [5]–[8]. Some work has also been reported using artificial neural networks [9] with wavelet preprocessing [10]. Primarily due to a relatively low understanding of the mechanisms underlying the problem, most existing methods suffer from the drawback of low accuracy which leads to higher false alarms and missed detections [11]. Moreover, due to the lack of reliable standardized data, most of the EEG analysis reported in the literature is performed on a small number of datasets which reduces the statistical significance of the conclusions. Such algorithms often demonstrate good accuracy for selected EEG segments but are not robust enough to adjust to EEG variations commonly encountered in a hospital setting.

In order to improve the statistical significance, in this research, a large number of EEG data sets belonging to three subject groups are used: 1) healthy subjects (normal EEG); 2) epileptic subjects during a seizure-free interval (interictal EEG); 3) epileptic subjects during a seizure (ictal EEG). The epilepsy and seizure detection problem is modeled as a three-group classification problem which is of great clinical significance. An automated computer model that can accurately differentiate between normal EEGs from interictal EEGs can be used to diagnose epilepsy in a clinical setting. A model that can accurately differentiate between interictal and ictal EEGs can be used to detect seizures in the environment of epilepsy monitoring units. In a clinical setting, the distinctions between these different EEG groups are often not very well-defined. The ultimate objective of this research is real-time monitoring of EEGs and seizure prediction which requires the computer model to be robust with respect to EEG variations across various mental states and subjects. Towards this end, the classification model must be able to classify all three groups accurately.

Manuscript received July 22, 2006; revised December 14, 2006. *Asterisk indicates corresponding author.*

S. Ghosh-Dastidar is with the Department of Biomedical Engineering, The Ohio State University, Columbus, OH 43210 USA.

*H. Adeli is with the Department of Civil and Environmental Engineering, The Ohio State University, 470 Hitchcock Hall, 2070 Neil Ave., Columbus, OH 43210 USA (e-mail: adeli.1@osu.edu).

N. Dadmehr is practicing in Westerville, OH 43081 USA.

Digital Object Identifier 10.1109/TBME.2007.891945

In this paper, the problem of improving the classification accuracy is attacked from two different angles: 1) designing an appropriate feature space by identifying combinations of parameters that increase the interclass separation and 2) designing a classifier that can accurately model the classification problem based on the selected feature space. To the best of the authors' knowledge no research has been published on an integrated and simultaneous application of wavelets, chaos theory, and neural networks for analysis of EEGs. Moreover, this research challenges the assumption that the EEG represents the dynamics of the entire brain as a unified system and needs to be treated as a whole. On the contrary, it is postulated that an EEG is a signal that represents the effect of the superimposition of diverse processes in the brain. Very little research has been done to select parameters from various physiological EEG sub-bands and use them simultaneously as input to classifiers.

II. Wavelet-Chaos Analysis: EEG Sub-Bands and Feature Space Design

Recently, the authors developed a wavelet-chaos methodology for analysis of EEGs and EEG sub-bands to identify potential parameters to be used in seizure and epilepsy detection [8]. This was the first time both chaos theory and wavelets were used simultaneously for analysis of EEG records. The methodology is shown succinctly in Fig. 1. It was observed that when the values of the correlation dimension (CD) and largest Lyapunov exponent (LLE) are computed from specific EEG sub-bands, the resulting differences in parameter values among the three groups are statistically significant (at the 99% confidence level). In this work, in addition to CD and LLE, the standard deviation (STD) of the EEGs and EEG sub-bands is selected to characterize the signal variance. Although STD, by itself, cannot differentiate all three groups, it is expected to increase the classification accuracy in combination with CD and LLE.

The purpose of the wavelet-chaos analysis is twofold. First, based on chaotic nonlinear dynamics of the brain it yields parameters with significant differences across the three groups of EEGs. This plays the role of feature enhancement and improves the subsequent classification accuracy. Second, it reduces a huge input space into a more manageable feature space. Prior to the wavelet-chaos analysis the dimensions of the input space is $N \times 6 \times 3K$ (N data points per EEG or EEG sub-band, six EEGs and EEG sub-bands and three groups of K EEGs). After the wavelet-chaos analysis, each EEG is represented as a P-dimensional data point in the reduced input space, $\mathbf{F}$, which is dubbed *feature space*. The dimensions of the feature space are $P \times 3K$, where $P \in \{1, 2, \ldots, 18\}$ is selected out of the three parameters (STD, CD, and LLE represented by the vectors (of size K) $\mathbf{S_D}$, $\mathbf{C_D}$, and $\mathbf{L_{LE}}$) obtained from the band-limited EEG as well as its five sub-bands (total of $6 \times 3 = 18$). Various combinations of these parameters are investigated as feature spaces for input to the classifiers with the goal of finding the most effective combination of parameters as well as the most effective classifier.

In this paper, four types of classifiers are investigated for classifying the EEGs into the aforementioned three groups: 1) unsupervised k-means clustering; 2) statistical discriminant analysis; 3) radial basis function (RBF) neural network (RBFNN); 4) Levenberg–Marquardt backpropagation (BP) neural network (LMBPNN). Three different variations of the discriminant analysis are compared: linear discriminant analysis (LDA) using the Euclidean distance (ELDA) and the Mahalanobis distance (MLDA), and quadratic discriminant analysis (QDA). Based on these studies, an innovative multi-paradigm wavelet-chaos-neural network methodology is presented for accurate classification of the EEGs. An overview of the methodology for the 3-class EEG classification problem is presented schematically in Fig. 1.

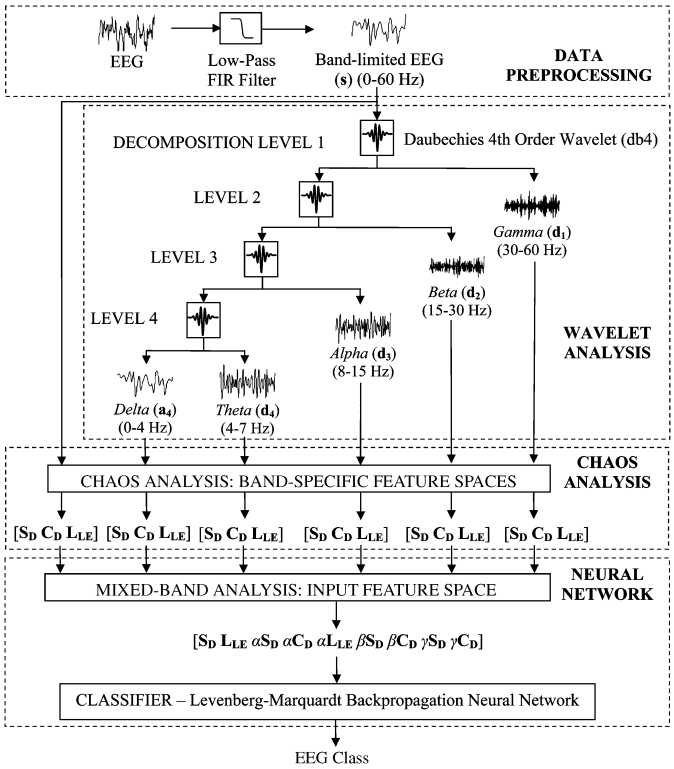

Fig. 1. Overview of the wavelet-chaos-neural network methodology for the 3-class EEG classification problem.

III. Classification Methods

A. k-Means Clustering

k-means clustering is an unsupervised clustering method which requires no prior information about the assignments of data points to groups. No training data is required and the data points are clustered based on the Euclidean distance into k groups [12]. Since all the data points are to be divided into three clusters, the process is initiated by an arbitrary selection of three points.

B. Discriminant Analysis

The aim of discriminant analysis is to minimize the intra-group variance and maximize the inter-group variance. If the mean of the entire $P \times n$ training set input, $\mathbf{F_R}$, (consisting of n data points), is represented by the vector $\boldsymbol{\mu}$ and the group mean of the input data points belonging to the jth group, $\mathbf{F_{R}}_j$ (consisting of n_j data points) is represented by the vector $\boldsymbol{\mu}_j$,

then the $P \times P$ intragroup and intergroup variance matrices are $\mathbf{S_W} = \sum_j \{n_j[\mathbf{F_R}j - \boldsymbol{\mu}_j][\mathbf{F_R}j - \boldsymbol{\mu}_j]^T/n\}$ and $\mathbf{S_B} = \sum_j \{n_j[\boldsymbol{\mu}_j - \boldsymbol{\mu}]^T[\boldsymbol{\mu}_j - \boldsymbol{\mu}]/n\}$, respectively, where the superscript T denotes the transpose of the matrix. For the EEG classification problem, all three methods, ELDA, MLDA, and QDA, are explored because of the lack of knowledge about 1) the underlying distribution and 2) the required number of training datasets to achieve accurate classification results.

C. LMBPNN

The input to each node in the hidden or output layer l is the same as the output from the preceding layer $l+1$ represented by the $N_{l+1} \times 1$ vector $\mathbf{Y}_{l+1}$ where N_{l+1} is the number of nodes in layer $l+1$. The $N_l \times 1$ weighted input vector to the hidden layer l is computed as $\mathbf{I}_l = \mathbf{W}_l^T \mathbf{Y_{l+1}}$ where $\mathbf{W_l}$ is the $N_{l+1} \times N_l$ weight matrix for layer l. The output of node j in layer l is computed using the tan-sigmoid activation function as $\mathbf{Y_l}(\mathrm{j}) = 2/[1 + \exp\{-\mathbf{I}_l(\mathrm{j})\}] - 1$ [13]. The classifier is trained using the input matrix $\mathbf{F_R}$ and the desired output vector $\mathbf{O_R}$. The least-mean squares method is used to minimize the error. The simple BP algorithm suffers from a very slow rate of convergence [9]. The Levenberg–Marquardt BP algorithm overcomes this shortcoming by defining an $m \times 1$ updated weight vector in the form $\mathbf{w_n} = \mathbf{w_o} - (\mathbf{J}^T\mathbf{J} + \mu_n\mathbf{I})^{-1}\mathbf{J}^T\mathbf{r}$ where $\mathbf{w_o}$ is the $m \times 1$ weight vector from the previous iteration, $\mathbf{r}$ is the $n \times 1$ error vector, $\mathbf{I}$ is the $m \times m$ identity matrix, m is the number of weights, $\mathbf{J}$ is the $n \times m$ Jacobian matrix, and μ_n is a coefficient used to insure invertibility [14].

D. RBFNN

The weighted input to the hidden layer is computed as the Euclidean distance between the $P \times 1$ input vector for the kth training instance, $\mathbf{F_R}(k)$ (a row vector), and the $N_{l+1} \times N_l$ weight matrix $\mathbf{W_l}$ of links connecting the input nodes to the nodes in the hidden layer. Since there is only one hidden layer, N_{l+1} is equal to n and N_l is equal to P. Therefore, the weighted input to the jth hidden node for the kth training instance is expressed as $\mathbf{I_l}(j) = [\mathbf{W_l}(j) - \mathbf{F_R}(k)]^T[\mathbf{W_l}(j) - \mathbf{F_R}(k)]$. The activation function for the hidden layer is a Gaussian function of the form $\mathbf{Y_l}(\mathrm{j}) = \exp\{-\mathbf{I}_l(\mathrm{j}).\ln(0.5)/p^2\}$ where p is the *spread* of the RBF which affects the shape of the Gaussian function [13], [14]. If the weight vector of the jth hidden node, $\mathbf{W_l}(j)$, is equal to the input vector for the kth training instance, $\mathbf{F_R}(k)$, the weighted input is 0 which results in an output of 1. The factor ln(0.5) in the exponent term is used to scale the output to 0.5 (the average of the limits of 0 and 1) when the weighted input is equal to the spread, p. RBFNN training also employs the least mean square error method. In this research, each iteration involves the addition of a hidden layer node, j. The input weight vector for this node, $\mathbf{W_l}(j)$, is selected to be equal to the input vector of the training instance k, $\mathbf{F_R}(k)$, that produces the minimum mean square error [14].

IV. Data Description and Analysis

The EEG data used in this research are from three different groups: normal, interictal, and ictal, made available online by Dr. Ralph Andrzejak of the Epilepsy Center at the University of Bonn, Germany (www.meb.uni-bonn.de/epileptologie/science/physik/eegdata.html). The diagnosis was temporal lobe epilepsy (epileptogenic focus: hippocampal formation). Each group contains $K = 100$ single channel EEG segments of 23.6 s duration each sampled at 173.61 Hz [16]. Each segment contains $N = 4097$ data points recorded at 173.61 Hz. Each EEG segment is considered as a separate EEG signal resulting in a total of 300 EEGs.

Testing the accuracy of all four classifiers with all different combinations of the eighteen parameters (2^{18} possible combinations) is a daunting task requiring months of computing time on a workstation. To reduce the computing time and output analysis to a more manageable one, the research is performed in two phases: 1) band-specific analysis and 2) mixed-band analysis. In the first phase, the six types of signals (EEG and the five EEG sub-bands) are considered one by one. For each type of signal, the EEGs are classified based on STD, CD, and LLE. Consequently, each signal is represented by a point in a 1-D to 3-D feature space. This is dubbed *band-specific* analysis in this article and the corresponding feature space is referred to as the *band-specific* feature space. One objective of band-specific analysis is to identify the classifiers that yield accurate classification and to eliminate the less accurate ones. Another objective is to identify specific combinations of parameters that may increase the classification accuracy of the selected classifiers.

In the second phase, the research is continued with the more promising classifiers and combinations of eighteen parameters selected from the six types of EEG and EEG sub-bands in the first step. As a result, each EEG is represented by a point in a 2-D to 18-D feature space. This is dubbed *mixed-band* analysis in this research and the corresponding feature space is referred to as the *mixed-band* feature space.

V. Band-Specific Analysis: Selecting Classifiers and Feature Spaces

A. k-Means Clustering

The centroid of a cluster representing any of the three EEG groups is initially selected randomly from the same group. Beyond this *group-biased* initialization step, no prior information about group assignments of data points is utilized. The sensitivity of k-means clustering to changes in starting points often leads to incorrect conclusions about the clustering accuracy. To overcome this shortcoming, the clustering process is repeated $N_R = 100$ times, with new points from each group selected randomly as initial points for every repetition.

The average classification accuracy percentages obtained in 100 repetitions are tabulated in Table I (standard deviations are noted in parentheses). A maximum clustering accuracy of 59.3% is observed using LLE obtained from the band-limited EEG (identified with boldface fonts in Table I).

B. Discriminant Analysis

In supervised classifications, for a given number of EEGs available, the dataset has to be divided carefully between training and testing sets. If the training set is selected too large, then the testing set becomes too small to yield meaningful classification accuracy results. In this paper, considering 100 EEGs are available for each group, the size of the training dataset is

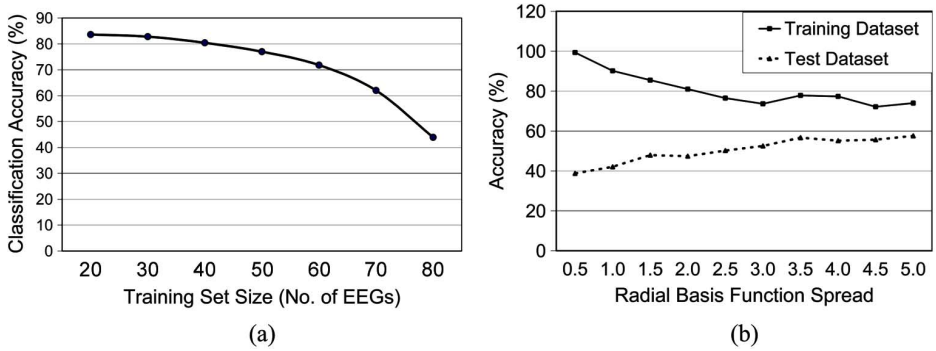

Fig. 2. (a) Classification accuracy of QDA using the feature space $[\mathbf{S_D\ C_D\ L_{LE}}]$ for various training set sizes and (b) variations of the RBFNN training and testing classification accuracies versus the RBF spread.

TABLE I
AVERAGE CLASSIFICATION ACCURACY PERCENTAGES USING k-MEANS CLUSTERING ($N_R = 100$; STANDARD DEVIATIONS IN PARENTHESES)

	Parameter Combination						
Signal	**S_D**	**C_D**	**L_{LE}**	**[S_D C_D]**	**[S_D L_{LE}]**	**[C_D L_{LE}]**	**[S_D C_D L_{LE}]**
Band-limited EEG	48.3 (0.0)	48.0 (0.7)	**59.3 (0.1)**	48.3 (0.0)	48.3 (0.0)	48.1 (0.9)	48.3 (0.0)
Delta (0-4 Hz)	52.3 (0.0)	40.3 (0.0)	45.9 (0.3)	52.3 (0.0)	52.3 (0.0)	40.3 (0.0)	52.3 (0.0)
Theta (4-7 Hz)	50.0 (0.0)	39.0 (0.4)	42.7 (0.6)	50.0 (0.0)	50.0 (0.0)	39.0 (0.4)	50.0 (0.0)
Alpha (8-12 Hz)	48.3 (0.0)	48.7 (0.5)	48.5 (0.3)	48.3 (0.0)	48.3 (0.0)	48.7 (0.5)	48.3 (0.0)
Beta (13-30 Hz)	49.3 (0.0)	44.0 (0.0)	36.3 (0.6)	49.3 (0.0)	49.3 (0.0)	44.0 (0.0)	49.3 (0.0)
Gamma (30-60 Hz)	48.0 (0.0)	49.2 (0.3)	37.1 (1.3)	48.0 (0.0)	48.0 (0.0)	48.9 (0.2)	48.0 (0.0)

varied from 20 to 50 EEGs (in increments of 10). Maximum classification accuracy was found when the training size is 20 EEGs for all three classifiers (ELDA, MLDA, and QDA) and input feature spaces. It was also found that the classification accuracy decreases with an increase in the size of the training dataset. As an example, the classification accuracy of QDA using the feature space $[\mathbf{S_D\ C_D\ L_{LE}}]$ is plotted for training sizes of 20–80 EEGs in Fig. 2(a). The results for all three classifiers and other feature spaces display a similar trend.

To account for the effect of the training data on classification accuracy, the classification is repeated $N_R = 100$ times, each time with a new randomly selected training dataset. The average classification accuracy percentages are tabulated in Table II (standard deviations are noted in parentheses). The accuracies of three types of classifiers based on discriminant analysis are compared using various input feature spaces. Results obtained from the classifiers trained by a dataset of 20 EEGs only are presented in this article (Table II) because this training size yields the highest classification accuracy, as mentioned previously. It is observed that, in general, the classification accuracy of QDA is the highest, followed by MLDA and, then, ELDA, with a few exceptions. The exceptions are for the combinations that yield low accuracies and consequently of no interest in the rest of this research.

It is observed that feature spaces that are based on certain combinations of parameters computed from the EEG or specific EEG sub-bands yield higher accuracies such as $[\mathbf{S_D\ C_D}]$ when computed from the *beta* and *gamma* sub-bands, $[\mathbf{S_D\ L_{LE}}]$ when computed from the band-limited EEG, and $[\mathbf{S_D\ C_D\ L_{LE}}]$ when computed from the *alpha* sub-band. These values are bold-faced in Table II. Much lower classification accuracies are obtained from the *delta* and *theta* sub-bands. The two highest values of the classification accuracy percentages are obtained with QDA when the $[\mathbf{S_D\ C_D}]$ feature space is computed from the *gamma* sub-band (85.5%) and the $[\mathbf{S_D\ L_{LE}}]$ feature space is computed from the band-limited EEG (84.8%). These results validate the assertions made in Adeli *et al.* [8] that CD computed from the higher frequency *alpha*, *beta*, and *gamma* sub-bands and the LLE computed from the band-limited EEG and *alpha* sub-band can be instrumental in differentiating between the three groups.

TABLE II
AVERAGE CLASSIFICATION ACCURACY PERCENTAGES USING A TRAINING DATASET SIZE OF 20 EEGS

		Parameter Combination (Training Size = 20 EEGs)						
Signal	**Classifier**	**S_D**	**C_D**	**L_{LE}**	**[S_D C_D]**	**[S_D L_{LE}]**	**[C_D L_{LE}]**	**[S_D C_D L_{LE}]**
Band-limited EEG	ELDA	47.3 (2.3)	36.1 (1.3)	52.9 (2.3)	47.6 (2.0)	79.6 (2.3)	68.6 (1.9)	81.1 (2.1)
	MLDA	53.1 (2.4)	35.3 (2.0)	49.8 (4.1)	54.9 (2.8)	84.0 (2.2)	66.1 (2.6)	81.6 (2.7)
	QDA	57.8 (3.1)	45.6 (1.4)	52.6 (2.7)	56.9 (3.0)	**84.8 (3.1)**	66.4 (2.6)	83.5 (2.7)
Delta (0-4 Hz)	ELDA	60.4 (2.0)	21.6 (2.9)	31.0 (2.7)	53.4 (3.5)	57.1 (3.2)	29.4 (1.9)	54.0 (3.4)
	MLDA	58.9 (4.5)	21.0 (3.0)	28.8 (3.0)	55.5 (4.1)	60.1 (3.3)	27.8 (2.6)	57.4 (3.4)
	QDA	62.3 (2.7)	24.5 (2.8)	30.3 (2.8)	60.5 (3.4)	64.1 (2.9)	29.1 (2.1)	63.9 (2.7)
Theta (4-7 Hz)	ELDA	53.6 (2.8)	21.8 (4.3)	31.0 (3.7)	53.8 (3.7)	59.1 (2.9)	34.4 (3.6)	56.6 (2.7)
	MLDA	57.6 (2.2)	22.3 (2.8)	30.9 (3.7)	57.8 (3.0)	62.0 (4.2)	31.8 (3.7)	61.0 (4.0)
	QDA	60.5 (2.6)	23.9 (2.2)	27.8 (3.5)	65.3 (3.3)	63.0 (4.0)	33.3 (3.2)	65.1 (3.0)
Alpha (8-12 Hz)	ELDA	60.1 (1.9)	34.1 (2.1)	33.4 (1.4)	67.3 (1.8)	66.4 (2.2)	38.0 (1.9)	68.4 (2.1)
	MLDA	63.4 (3.5)	32.3 (2.8)	31.1 (1.9)	64.0 (4.4)	73.1 (2.8)	38.4 (3.2)	69.9 (3.3)
	QDA	65.4 (2.4)	38.3 (2.0)	35.4 (2.0)	74.5 (2.9)	75.4 (1.8)	42.6 (2.4)	**75.9 (2.4)**
Beta (13-30 Hz)	ELDA	72.3 (2.0)	30.1 (2.1)	16.4 (1.9)	58.6 (7.1)	56.8 (7.7)	30.9 (2.0)	55.3 (5.2)
	MLDA	74.6 (2.1)	27.9 (3.0)	17.3 (2.5)	73.0 (3.6)	69.5 (4.4)	31.0 (2.7)	72.0 (4.5)
	QDA	75.4 (1.6)	31.8 (2.0)	17.8 (2.4)	**80.9 (2.8)**	74.6 (2.1)	33.3 (1.6)	80.1 (2.9)
Gamma (30-60 Hz)	ELDA	71.8 (2.7)	35.6 (1.0)	17.1 (2.9)	59.0 (4.1)	51.9 (9.8)	35.1 (2.4)	55.6 (4.6)
	MLDA	74.5 (2.9)	32.0 (3.0)	15.9 (4.1)	71.1 (5.0)	68.8 (5.4)	29.6 (4.5)	68.4 (5.6)
	QDA	81.0 (1.7)	36.6 (1.7)	17.1 (3.2)	**85.5 (1.2)**	78.8 (2.6)	37.6 (2.4)	83.6 (1.9)

C. RBFNN

Two parameters of RBFNN architecture and formulation are investigated: 1) the number of nodes in the hidden layer and 2) the spread of the RBF, p, with the goal of achieving optimum network performance. The number of nodes in the hidden layer affects the training performance. Employing a large number of nodes, for example, equal to the number of training instances, is computationally expensive but enables the network to be trained with 100% accuracy. In many cases, employing a smaller number of nodes yields sufficiently accurate results. Therefore, training is started with a small number of nodes and is repeated with an increasingly larger number of nodes. Training is terminated when any one of the two conditions holds: 1) training error decreases to 0.001 or 2) number of nodes is equal to the number of training instances.

The spread, p, has to be selected very carefully. In RBFNN, a specific input is supposed to excite only a limited number of nodes in the hidden layer. When the spread is too large, all hidden nodes respond to a given input which results in loss of classification accuracy. On the other hand, when the spread is too small, each node responds only to a very specific input. Consequently, the node will be unable to classify any new input accurately. The optimum spread for maximum classification accuracy is found by varying it from 0.5 to 5 in increments of 0.5. The training and classification accuracies versus the spread values are presented in Fig. 2(b). It is observed that for the most part the classification training accuracy decreases whereas the classification (testing) accuracy increases as the spread is increased. Both curves reach approximately a plateau for larger values of the spread, say, equal or greater than 4. The reciprocal relationship may be attributed to the overtraining of the RBFNN and its consequent inability to respond to new inputs appropriately.

The classification training-testing process is repeated $N_R = 10$ times, each time with a new randomly selected training dataset. Since, typically, neural network training requires significantly larger computational efforts compared with k-means clustering or discriminant analysis, the number of repetitions is reduced to ten. The average classification accuracy percentages for all band-specific feature spaces using a training size of 40 EEGs and RBF spread of 4 are tabulated in Table III (standard deviations are noted in parentheses). It is observed that RBFNN requires a larger training size than the classifiers based on discriminant analysis but yield less accurate results. Also, the standard deviations of the results are much larger for RBFNN than for discriminant analysis which implies greater dependence of RBFNN on the training data. The two highest values of classification accuracy are obtained when the $[\mathbf{S_D}]$ feature space is computed from the *gamma* sub-band using a spread of 4 and training sizes of 40 (76.5%) (shown by boldface in Table III) and 50 (76.2%) (not shown in Table III). It should be noted that results obtained with other spread values in the range of 0.5 to 5 are not shown in this paper for the sake of brevity.

D. LMBPNN

LMBPNN is investigated in a manner similar to RBFNN in terms of the training dataset sizes and the number of classification training-testing repetitions. Overall, a training size of 40 EEGs yields the best classification accuracy and is therefore selected for all EEGs and EEG sub-bands. To accurately model the complex dynamics underlying EEGs, the effect of the network architecture, that is, the number of hidden layers and the number of nodes in the hidden layers on the classification accuracy is investigated. This was performed in a manner different from that of RBFNN due to basic differences in the architecture of the two neural networks. The number of nodes in the first and second hidden layers is increased in increments of 5 from 5 to 20 and 0 to 15, respectively. Training is terminated when any one of the three conditions holds: 1) training error decreases to 0.001; 2) training error gradient decreases to 0.01; or 3) number of training epochs reaches 100.

TABLE III
AVERAGE CLASSIFICATION ACCURACY PERCENTAGES USING A TRAINING DATASET SIZE OF 40 EEGS AND RBFNN

	Parameter Combination (Training Size = 40 EEGs, RBF Spread = 4)						
Signal	S_D	C_D	L_{LE}	$[S_D\ C_D]$	$[S_D\ L_{LE}]$	$[C_D\ L_{LE}]$	$[S_D\ C_D\ L_{LE}]$
Band-limited EEG	41.9 (12.3)	43.4 (3.1)	44.3 (2.9)	47.0 (11.0)	52.1 (6.7)	61.9 (2.7)	51.9 (11.8)
Delta (0-4 Hz)	53.8 (7.7)	38.9 (1.8)	40.8 (1.6)	41.0 (7.7)	47.7 (7.4)	41.3 (3.6)	48.1 (4.7)
Theta (4-7 Hz)	50.2 (11.5)	37.1 (1.8)	34.1 (0.9)	45.8 (6.8)	45.8 (9.8)	42.6 (2.8)	51.0 (4.1)
Alpha (8-12 Hz)	49.7 (9.5)	40.5 (3.2)	33.6 (1.6)	52.9 (5.7)	38.2 (6.9)	42.2 (2.5)	58.8 (7.9)
Beta (13-30 Hz)	58.5 (6.7)	37.1 (3.4)	33.1 (1.0)	59.1 (6.8)	49.5 (9.0)	39.8 (3.4)	64.8 (6.1)
Gamma (30-60 Hz)	**76.5 (3.0)**	39.5 (4.6)	34.1 (0.8)	55.3 (4.9)	66.8 (5.8)	42.2 (3.4)	51.2 (9.0)

Based on this parametric study two hidden layers each with 10 or 15 nodes appeared to yield the best classification results for all combinations of the parameters computed from the six types of EEGs and EEG sub-bands. No other discriminatory patterns were observed. The average classification accuracy percentages for all band-specific feature spaces for a training size of 40 are summarized in Table IV (standard deviations are noted in parentheses). In general, it is observed that the feature spaces that show higher classification accuracy are $[\mathbf{S_D\ C_D}]$ when computed from the *beta* and *gamma* sub-bands and $[\mathbf{S_D\ L_{LE}}]$ when computed from the band-limited EEG and the *alpha* sub-band. These values are identified by boldface in Table IV. Much lower classification accuracy is obtained from the *delta* and *theta* sub-bands. This observation is similar to that observed for the case of QDA. The two highest values of classification accuracy are obtained when the $[\mathbf{S_D\ L_{LE}}]$ feature space is computed from the band-limited EEG (89.9%) and the $[\mathbf{S_D\ C_D}]$ feature space is computed from the *gamma* sub-band (87.3%).

VI. MIXED-BAND ANALYSIS: WAVELET-CHAOS-NEURAL NETWORK

Fig. 3(a) summarizes the maximum classification accuracy percentages of all band-specific feature spaces for six different types of classifiers obtained in phase one of this research. The results of the extensive band-specific analysis lead to the conclusion that QDA and LMBPNN yield the two highest classification accuracies (85.5% and 89.9%, respectively). Moreover, both classifiers have small standard deviations using various

TABLE IV
AVERAGE CLASSIFICATION ACCURACY PERCENTAGES USING A TRAINING DATASET SIZE OF 40 EEGS AND LMBPNN

Signal	Parameter Combination (Training Size = 40 EEGs)						
	S_D	C_D	L_{LE}	[S_D C_D]	[S_D L_{LE}]	[C_D L_{LE}]	[S_D C_D L_{LE}]
Band-limited EEG	64.8 (2.7)	47.2 (3.3)	56.7 (6.8)	63.8 (3.2)	**89.9 (1.8)**	72.1 (3.4)	86.9 (2.9)
Delta (0-4 Hz)	68.2 (8.5)	39.9 (4.2)	41.1 (2.1)	63.9 (4.0)	66.1 (6.3)	40.2 (3.3)	65.1 (8.0)
Theta (4-7 Hz)	69.4 (4.5)	39.2 (2.5)	48.2 (4.6)	72.8 (4.4)	71.6 (4.5)	43.8 (3.9)	73.8 (4.1)
Alpha (8-12 Hz)	74.3 (4.3)	42.8 (2.9)	42.1 (3.9)	79.9 (3.3)	**80.5 (2.7)**	48.7 (5.0)	79.4 (6.7)
Beta (13-30 Hz)	81.1 (2.0)	43.1 (5.1)	38.1 (3.9)	**85.5 (2.1)**	79.7 (2.7)	45.2 (4.6)	83.4 (1.8)
Gamma (30-60 Hz)	86.4 (2.0)	44.9 (3.8)	46.9 (3.4)	**87.3 (1.7)**	83.1 (2.6)	50.0 (4.1)	85.7 (2.7)

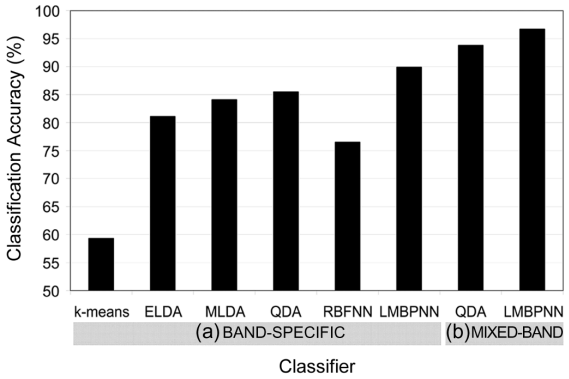

Fig. 3. Maximum classification accuracy percentages obtained from (a) six different types of classifiers for all band-specific feature spaces and (b) QDA and LMBPNN for the mixed-band feature space $[\mathbf{S_D}\ \mathbf{L_{LE}}\ \alpha\mathbf{S_D}\ \alpha\mathbf{C_D}\ \alpha\mathbf{L_{LE}}\ \beta\mathbf{S_D}\ \beta\mathbf{C_D}\ \gamma\mathbf{S_D}\ \gamma\mathbf{C_D}]$.

combinations of training/testing datasets. Consequently, they are robust with respect to changes in training data. In general, LMBPNN yields higher classification accuracies than QDA, but requires a larger training dataset size (40 EEGs versus 20 EEGs in the parametric studies performed in this research). In phase two of the research, QDA and LMBPNN are selected for further investigation using the 2- to 18-dimensional mixed-band feature spaces described earlier.

Based on the band-specific analysis of phase one, two other important conclusions are made. For LMBPNN, the following eight parameters yield higher classification accuracies: STD computed from the band-limited EEG and *alpha*, *beta*, and *gamma* sub-bands; CD computed from *beta* and *gamma* sub-bands; and LLE computed from the band-limited EEG and *alpha* sub-band. The same parameters yield higher classification accuracies for QDA as well, plus, a ninth parameter, that is, CD computed from the *alpha* sub-band.

In phase two of this research, over five hundred different combinations of mixed-band feature spaces consisting of promising parameters from phase one of the research were investigated. This effort took weeks of workstation computing time. For statistical consistency, the same number of training-testing repetitions of $N_R = 10$ was used for both QDA and LMBPNN. A training size of 40 EEGs yielded the highest classification accuracies for the mixed-band feature spaces and was employed for both classifiers.

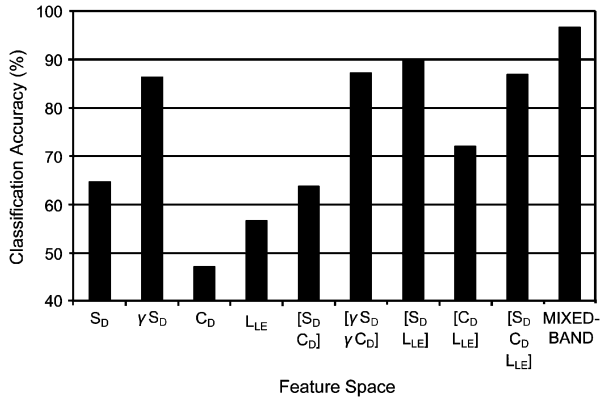

Fig. 4. Comparison of maximum classification accuracy percentages of LMBPNN for different feature spaces. The mixed-band feature space is $[\mathbf{S_D}\ \mathbf{L_{LE}}\ \alpha\mathbf{S_D}\ \alpha\mathbf{C_D}\ \alpha\mathbf{L_{LE}}\ \beta\mathbf{S_D}\ \beta\mathbf{C_D}\ \gamma\mathbf{S_D}\ \gamma\mathbf{C_D}]$.

It was discovered that a mixed-band feature space consisting of all nine aforementioned parameters $[\mathbf{S_D}\ \mathbf{L_{LE}}\ \alpha\mathbf{S_D}\ \alpha\mathbf{C_D}\ \alpha\mathbf{L_{LE}}\ \beta\mathbf{S_D}\ \beta\mathbf{C_D}\ \gamma\mathbf{S_D}\ \gamma\mathbf{C_D}]$ result in the highest classification accuracy for both QDA and LMBPNN. In this notation, the parameter prefix denotes the EEG sub-band from which the parameter is computed. Absence of a prefix indicates that the parameter is computed from the band-limited EEG. The classification accuracies obtained using the 9-parameter mixed-band feature space for both QDA and LMBPNN are presented in Fig. 3(b).

The average classification accuracy percentages using QDA and LMBPNN are 93.8% (with a standard deviation of 1.0) and 96.7% (with a standard deviation of 2.9), respectively. The classification accuracy of LMBPNN using the 9-parameter mixed-band feature space is significantly improved compared with using various band-specific feature spaces as shown in the sample results of Fig. 4. Band-specific feature spaces based on single parameters yield low to very low classification accuracies ($<$65.0%) except for STD computed from the *gamma* sub-band which yields a classification accuracy of 86.4%. Band-specific feature spaces based on combinations of two and three parameters lead to more accurate classification (in the range of 65%–90%). The 9-parameter mixed-band feature space yields the highest classification accuracy (96.7%). To the best of the authors' knowledge such a high level of accuracy has not been reported previously in the literature.

The training of the LMBPNN classifier with the 9-parameter mixed-band feature space presented in the paper is done quickly. On an Intel Pentium IV workstation (1700 MHz, 512 MB RAM), the classifier training for ten repetitions is completed in approximately 30 s for 120 training instances (40 from each of the three subject groups). Using the trained network to classify a new 9-parameter mixed-band input takes only a fraction of a second on the aforementioned workstation.

VII. CONCLUDING REMARKS

Based on this research and previous work by the authors [8], the 9-parameter mixed-band feature space $[\mathbf{S_D}\ \mathbf{L_{LE}}\ \alpha\mathbf{S_D}\ \alpha\mathbf{C_D}\ \alpha\mathbf{L_{LE}}\ \beta\mathbf{S_D}\ \beta\mathbf{C_D}\ \gamma\mathbf{S_D}\ \gamma\mathbf{C_D}]$ yields the most accurate classification results. When this feature space is

used as input to the LMBPNN classifier, a maximum classification accuracy of 96.7% is achieved. The results of this research lead to the conclusion that all three key components of the wavelet-chaos-neural network methodology are important for improving the accuracy of EEG classification. Wavelet analysis decomposes the EEG into sub-bands and is instrumental in creating the mixed-band features spaces with an improved accuracy over band-specific feature spaces. The parameters used in these feature spaces are obtained by statistical analysis (STD) and chaos analysis (CD and LLE). Although, these parameters are unable to classify the EEGs individually, their combinations improve the classification accuracy significantly especially when multiple-parameter mixed-band feature spaces are employed. Finally, the LMBPNN classifier classifies the EEGs into the three groups more accurately than any of the other classification methods investigated. Judicious combinations of parameters and classifiers can accurately discriminate between the three types of EEGs and result in increased accuracy for real-time epilepsy and seizure detection systems to be explored in future work.

REFERENCES

[1] L. D. Iasemidis, L. D. Olson, J. C. Sackellares, and R. S. Savit, "Time dependencies in the occurrences of epileptic seizures: A nonlinear approach," *Epilepsy Res.*, vol. 17, pp. 81–94, 1994.

[2] F. H. Lopes da Silva, J. P. Pijn, and W. J. Wadman, "Dynamics of local neuronal networks: Control parameters and state bifurcations in epileptogenesis," *Progress Brain Res.*, vol. 102, pp. 359–370, 1994.

[3] H. Adeli, Z. Zhou, and N. Dadmehr, "Analysis of EEG records in an epileptic patient using wavelet transform," *J. Neurosci. Meth.*, vol. 123, pp. 69–87, 2003.

[4] P. J. Durka, "From wavelets to adaptive approximations: Time-frequency parameterization of EEG," *Biomed. Eng. Online*, vol. 2, p. 1(1–30), 2003.

[5] L. D. Iasemidis and J. C. Sackellares, "The temporal evolution of the largest Lyapunov exponent on the human epileptic cortex," in *Measuring Chaos in the Human Brain*, D. W. Duke and W. S. Pritchard, Eds. Singapore: World Scientific, 1991, pp. 49–82.

[6] C. E. Elger and K. Lehnertz, "Seizure prediction by non-linear time series analysis of brain electrical activity," *Eur. J. Neurosci.*, vol. 10, pp. 786–789, 1998.

[7] I. D. Iasemidis, D. S. Shiau, W. Chaovalitwongse, J. C. Sackellares, P. M. Pardalos, J. C. Principe, P. R. Carney, A. Prasad, B. Veeramani, and K. Tsakalis, "Adaptive epileptic seizure prediction system," *IEEE Trans. Biomed. Eng.*, vol. 50, no. 5, pp. 616–627, May 2003.

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

[9] H. Adeli and S. L. Hung, *Machine Learning—Neural Networks, Genetic Algorithms, and Fuzzy Sets*. New York: Wiley, 1995.

[10] A. Petrosian, D. Prokhorov, R. Homan, R. Dascheiff, and D. Wunsch, II, "Recurrent neural network based prediction of epileptic seizures in intracranial and extracranial EEG," *Neurocomputing*, vol. 30, pp. 201–218, 2000.

[11] L. D. Iasemidis, "Epileptic seizure prediction and control," *IEEE Trans. Biomed. Eng.*, vol. 50, no. 5, pp. 549–556, May 2003.

[12] W. R. Dillon and M. Goldstein, *Multivariate Analysis: Methods and Applications*. New York: Wiley, 1984.

[13] N. K. Bose and P. Liang, *Neural Network Fundamentals With Graphs, Algorithms and Applications*. New York: McGraw Hill, 1996.

[14] M. T. Hagan, H. B. Demuth, and M. Beale, *Neural Network Design*. Boston, MA: PWS Publishing, 1996.

[15] S. Ghosh-Dastidar and H. Adeli, "Wavelet-clustering-neural network model for freeway incident detection," *Comput.-Aided Civil and Infrastructure Eng.*, vol. 18, pp. 325–338, 2003.

[16] R. G. Andrzejak, K. Lehnertz, C. Rieke, F. Mormann, P. David, and C. E. Elger, "Indications of non-linear deterministic and finite dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state," *Phys. Rev. E*, vol. 64, p. (061907)1–8, 2001.

Samanwoy Ghosh-Dastidar was born in New Delhi, India, in 1976. He received the B.E. degree in civil engineering from the University of Roorkee, Roorkee, India, in 1997 and the M.S. degree in civil engineering and the Ph.D. degree in biomedical engineering both from The Ohio State University, Columbus, in 2002 and 2007, respectively.

His research interests include mathematical modeling of dynamic systems, signal and image processing, data mining, artificial neural networks, and nonlinear (chaos) analysis. His current areas of application include epilepsy diagnosis, detection and prediction of epileptic seizures, and early detection of Alzheimer's disease. Previous areas of interest include intelligent transportation systems.

Hojjat Adeli (M'95) received the B.S.-M.S. degree from the University of Tehran, Tehran, Iran, in 1973 and the Ph.D. degree from Stanford University, Stanford, CA, in 1976.

He is currently a Professor in the departments of Civil and Environmental Engineering and Geodetic Science, Aerospace Engineering, Biomedical Engineering, Biomedical Informatics, Electrical and Computer Engineering, and Neuroscience at The Ohio State University, Columbus. He is also the holder of the Abba G. Lichtenstein Professorship. Since 1976, he has authored over 400 research and scientific publications in various fields of computer science, engineering, and applied mathematics. His research has been published in 72 different journals. He has authored 11 books including *Machine Learning—Neural Networks, Genetic Algorithms*, and *Fuzzy Systems* (Wiley, 1995) and *Wavelets in Intelligent Transportation Systems* (Wiley, 2005). He has also edited 12 books. He is the Founder and Editor-in-Chief of the international research journals *Computer-Aided Civil and Infrastructure Engineering* and *Integrated Computer-Aided Engineering*. He is also the Editor-in-Chief of the *International Journal of Neural Systems*.

Dr. Adeli received the Distinguished Scholar Award "*in recognition of extraordinary accomplishment in research and scholarship*" from The Ohio State University in 1998. He was elected an Honorary Member of American Society of Civil Engineers "*for wide-ranging, exceptional, and pioneering contributions to computing in civil engineering and extraordinary leadership in advancing the use of computing and information technologies in many engineering disciplines throughout the world*" in 2005.

Nahid Dadmehr received the M.D. degree from the University of Tehran, Tehran, Iran, in 1979. She did her medical residency and fellowship at The Ohio State University, Columbus.

She has been a practicing board-certified Neurologist in central Ohio since 1991. Her research has been published in several journals including *Clinical Electroencephalography, Electroencephalography and Clinical Neurophysiology, Neurology*, and the *Journal of Neuroscience Methods*.