Cybernetics and Systems Analysis, Vol. 44, No. 1, 2008

ON THE TIME SERIES SUPPORT VECTOR MACHINE USING DYNAMIC TIME WARPING KERNEL FOR BRAIN ACTIVITY CLASSIFICATION[1]

W. A. Chaovalitwongse[a] and P. M. Pardalos[b]

UDC 612.821:51+519.7+519.8

A new data mining technique used to classify normal and pre-seizure electroencephalograms is proposed. The technique is based on a dynamic time warping kernel combined with support vector machines (SVMs). The experimental results show that the technique is superior to the standard SVM and improves the brain activity classification.

Keywords: *time series, classification, EEG, brain dynamics, optimization, dynamic time warping, epilepsy, support vector machines.*

1. INTRODUCTION

The human brain is among the most complex systems known to man. In neuroscience research, a countless number of studies attempted to comprehend the mechanism of brain functions through detailed analysis of neuronal excitability and synaptic transmission. Many theories of brain functions were proposed over the last century. Only in the last few years, it became feasible to capture simultaneous responses from sufficiently large numbers of neurons to empirically test those long-standing hypotheses about brain function. However, most neuro-scientific experiments resulted in massive datasets in the form of multi-dimensional time series data. These data contain both spatial and temporal properties of brain functions. Making sense of such massive data requires very efficient and sophisticated techniques that are capable of capturing both spatial and temporal properties simultaneously. Current research studies in data mining and classification are mostly focused on data with only spatial or temporal properties. In addition, very few studies in quantitative neuroscience are not tailored to exploit both spatial and temporal properties of this relentless flood of information.

In this study, epilepsy will be a case in point. Epilepsy is the second most common brain disorder after stroke, yet the most devastating one. The most disabling aspect of epilepsy is the uncertainty of recurrent seizures that can be characterized by a chronic medical condition produced by temporary changes in the electrical function of the brain. Most epilepsy studies employ electroencephalograms (EEGs) as a tool for capturing electrical changes and evaluating physiological states (normal and abnormal) of the brain. Although EEGs offer excellent spatial and temporal resolution of brain activity, EEG data are so enormous and are represented in the form of such long-term multi-dimensional time series that neuroscientists understand very little about the dynamical transitions to neurological dysfunctions of seizures. A necessary first step to advance epilepsy research is to develop a seizure prediction/warning system. Therefore, the main goal of this study is to employ techniques in data mining and optimization to discover seizure-precursor patterns encrypted in enormous EEGs. In order to validate the reliability of a seizure prediction/warning system, one has to test the hypothesis that the EEGs during the normal period differ from the EEGs during a seizure-precursor. In turn, this will lead to a classical classification problem. However, data used in this classification problem have spatio-temporal properties. We herein propose an efficient and effective spatio-temporal data mining/classification method for multi-dimensional time series classification of brain activity.

[1]This research was partially supported by Rutgers Research Council grant 202018, NSF grants CCF-0546574, DBI-980821, EIA-9872509, and CCF 0546574, and NIH grant R01-NS-39687-01A1.

[a]Rutgers University, Piscataway, NJ, USA, *wchaoval@rci.rutgers.edu*. [b]University of Florida, Gainesville, FL, USA, *pardalos@ufl.edu.* Translated from Kibernetika i Sistemnyi Analiz, No. 1, pp. 159–173, January–February 2008. Original article submitted November 24, 2006.

1060-0396/08/4401-0125

The organization of the succeeding sections of this paper is as follows. The background of this work including classification techniques in the literature and previous studies in seizure prediction and classification will be discussed in Sec. 2. Then the basic concepts and standard classification procedure of support vector machines are discussed in Sec. 3. The methods employed in this study including the quantification of brain dynamics, EEG data acquisition, and support vector machines with dynamic time warping kernel are given in Sec. 4. The design of experiments and details of an empirical study are described in Sec. 5. The results on the statistical evaluation and performance characteristics of the proposed classification method are presented in Sec. 6. In Sec. 7, concluding remarks are given and future works are discussed.

2. BACKGROUND

In this section, we give an overview of classification techniques described in the literature. Then we will focus on optimization-based classification techniques like support vector machines. Later in this section, we give a brief background about epilepsy and the significance of this work to epilepsy research.

2.1. Classification Techniques. Generally, classification techniques are combinatorial in nature as they are involved with discrete decisions. Thus, classification problems can be naturally considered as discrete optimization problems [3, 6, 15, 18, 19, 22, 34]. An enormous number of optimization techniques for classification problems were developed during the past few decades, including classification trees, support vector machines (SVMs), linear discriminant analysis, logic regression, least squares, nearest neighbors, etc. Most optimization methods in classification were applied to SVMs. A number of linear programming formulations for SVMs were used to explore the properties of the structure of optimization problems and to solve large-scale problems [5, 35]. The SVM technique proposed in [35] was also demonstrated to be applicable to the generation of complex space partitions similar to those obtained by C4.5 [39] and CART [7]. Current SVM research mainly focuses on extending SVMs to solve multi-class problems [20, 31, 45]. Nevertheless, there were very few studies in the literature that address the use of SVMs for time series classification. Moreover, most of existing studies are focused only on pattern recognition and similarity search [12, 13, 17, 29, 27, 28] or on the use of kernel functions for single time series transformation like speech recognition [47, 43, 50, 4, 52].

Multi-dimensional time series classification (MDTSC) has to deal with massive time series data with both spatial and temporal properties (e.g., EEGs). However, to our knowledge, almost all of SVM studies address only a spatial or a temporal classification problem. None of current SVM studies attempts to simultaneously consider both problems. In this study, a novel SVM approach to MDTSC is developed based on an improved kernel of finding a separating plane of SVMs in a time series sample and also the idea of projecting data from temporal properties. This technique represents a bridge between parametric techniques that require a priori knowledge of the distributions underlying data and nonparametric techniques that presuppose the functional form of the discriminant surfaces separating different pattern classes.

2.2. Epilepsy Research. Epilepsy will be a case in point in this proposal. Epilepsy is the second most common brain disorder currently afflicting at least 2 million Americans. The diagnosis and treatment of epilepsy are complicated by the disabling aspect that seizures occur spontaneously and unpredictably. The major epilepsy research lies in studying how neuronal circuitries of the brain support these electrical changes. Most epilepsy studies use EEGs as a tool for capturing electrical changes and evaluating physiological states (normal and abnormal) of the brain. Although EEGs were widely used over the past few decades, neuro-scientists understand very little about the dynamical transitions to neurological dysfunctions of seizures [21]. In some types of epilepsy (e.g., focal or partial epilepsy), there is a localized structural change in the neuronal circuitry within the cerebrum that produces organized quasi-rhythmic discharges that spread from the region of origin (epileptogenic zone) and activate other areas of the cerebral hemispheres. The development of the epileptic state can be considered as changes in network circuitry of neurons in the brain that produce changes in voltage potential that can be captured by EEG recordings. These changes are reflected by wriggling lines along the time axis in a typical EEG recording. A typical electrode montage for intracranial EEG recordings in our study is shown in Fig. 1. The 10-second EEG profiles during the normal and pre-seizure periods of patient 1 are illustrated in Figs. 2a and 2b. The EEG onset of a typical epileptic seizure is illustrated in Fig. 2c. Figure 2d shows the post-seizure state of a typical epileptic seizure.

2.3. Seizure Prediction and Brain Activity Classification. If seizures could be predicted, this would lead to the development of completely novel diagnostic and therapeutic advances in controlling epileptic seizures. This will tremendously improve the quality of life for patients who currently suffer from epilepsy. There is growing evidence that human epileptic seizures are preceded by physiological changes that are reflected in the dynamical characteristics of EEG signals. Our group reported the pre-seizure convergence of $STL_{\max}$ values (calculated from intracranial or scalp electrode

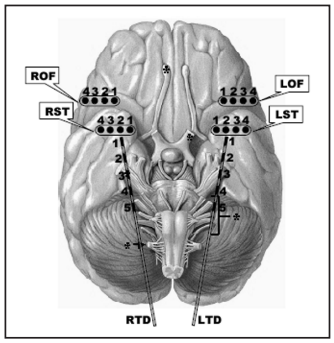

Fig. 1. Inferior transverse views of the brain are depicted that illustrate approximate depths and subdural electrode placements for EEG recordings. Subdural electrode strips are placed over the left orbitofrontal (LOF), right orbitofrontal (ROF), left subtemporal (LST), and right subtemporal (RST) cortices. Depth electrodes are placed in the left temporal depth (LDT) and right subtemporal depth (RTD) to record hippocampal activity.

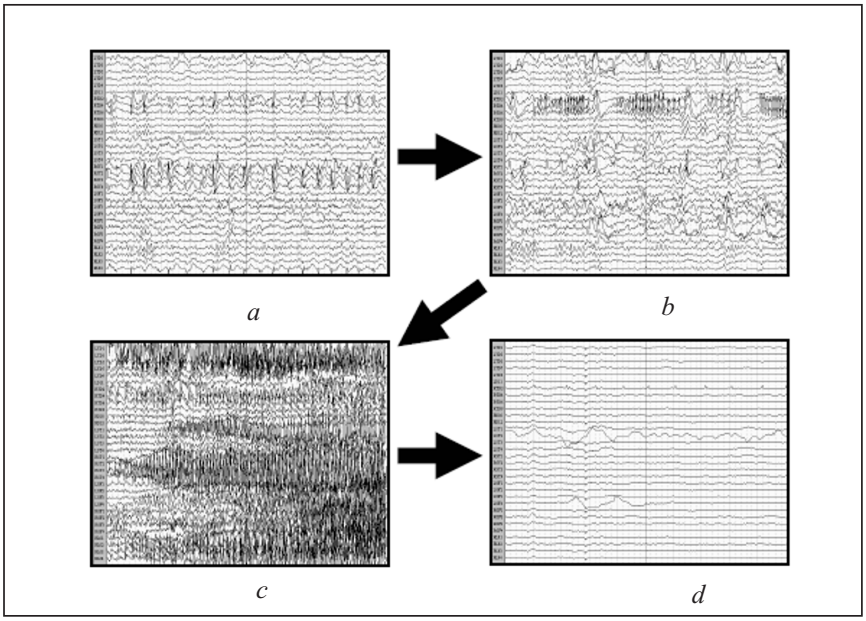

Fig. 2. Twenty-second EEG recordings of normal activity (a), pre-seizure activity (b), seizure onset activity (c), and post-seizure activity obtained from 32 electrodes for Patient 1 (d). Each horizontal trace represents the voltage recorded from one of the electrode sites whose anatomical location is presented in Fig. 1.

EEG recordings) that occurred tens of minutes prior to epileptic seizures [24]. Later on, Elger and Lehnertz [14, 32] reported reductions in the effective correlation dimension (D_2^{eff}, a measure of the complexity of EEG signals) that were more prominent in pre-seizure EEG samples than at times more distant from a seizure. They estimated that a detectable change in dynamics could be observed at least 2 minutes before a seizure in most cases [14]. Martinerie and coworkers [36] also reported significant differences between dimension measures obtained in pre-seizure versus normal EEG samples. They found an abrupt decrease in dimension during the pre-seizure transition. This study also employed relatively brief (40 minutes) samples of pre-seizure and normal data. More recently, this group reported changes in brain dynamics obtained from scalp electrode recordings ofEEGs. By comparing pre-seizure EEG samples to a reference sample selected from normal data, they found evidence for dynamical changes that anticipated temporal lobe seizures by periods of up to 15 minutes [40]. Recently, Litt and coworkers [33] reported sustained bursts of energy in some EEG channels visually selected by one of the investigators.

During the past decade, seizure predictability was demonstrated through the above-mentioned studies including our previous studies in [9,10,25,38]. These studies were motivated by mathematical models used to analyze multidimensional complex systems (e.g., a neuronal network in the brain) based on the chaos theory and optimization techniques. The results of these studies demonstrated that a seizure is essentially a reflecting transition of progressive changes of hidden dynamical patterns in EEGs. Such transitions have been shown to be detectable through the quantitative analysis of brain dynamics [9,10,38]. However, in order to validate seizure predictability, one would have to qualitatively and quantitatively demonstrate that normal EEGs differ from pre-seizure (abnormal) EEGs. The discriminant ability to differentiate and classify a pre-seizure EEG signal is logically a prerequisite and a necessary first step of seizure prediction/warning development. Thus far, to our knowledge, none of current epilepsy studies in the literature is undertaken to test this hypothesis. Our group attempted to apply data mining techniques using hidden dynamical characteristics to differentiate normal and pre-seizure EEGs [11].

3. SUPPORT VECTOR MACHINES (SVMs)

In this section, we discuss some basic concepts of SVMs. Then we explain a general classification procedure of SVMs. Later, we discuss the use of kernel functions, the most widely used trick of SVMs.

3.1. Basic Concepts. SVMs is one of classification techniques widely used in practice. The essence of support vector machines is to construct separating surfaces that will minimize the upper bound on the out-of-sample error. In the case of one linear surface (plane) separating elements from two classes, this approach will choose the plane that maximizes the sum of

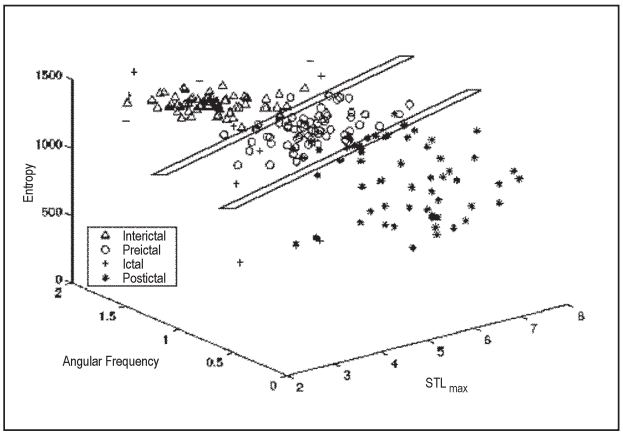

Fig. 3. Example of hyperplanes separating different brain's states.

the distances between the plane and the closest elements from each class (i.e., the "gap" between the elements from different classes). The mathematical definition of support vector machines can be described as follows. Let all the data elements be represented as *n*-dimensional vectors (or points in an *n*-dimensional space). Then these elements can be separated geometrically by constructing the surfaces that serve as "borders" between different groups of points. One of common approaches is the use of linear surfaces (planes) for this purpose, but different types of nonlinear (e.g., quadratic) separating surfaces can be considered in certain applications. Note that, in practice, it is impossible to find a surface that would "perfectly" separate the points according to the value of some attribute. In other words, data points with different values of a given attribute may not necessarily lie at the different sides of the surface. However, in general, the number of these errors should be small enough. The classification problem of support vector machines can be represented as the problem of finding geometrical parameters of a separating surface (or surfaces). As will be described below, these parameters can be found by solving the optimization problem of minimizing the misclassification error for the elements in the training dataset (the so-called "in-sample error"). After determining these parameters, every new data element will be automatically assigned to a certain class according to its geometrical location in the elements space. The procedure of using the existing dataset for classifying new elements is often called "training a classifier" (and the corresponding dataset is referred to as the "training dataset"). This means that the parameters of separating surfaces are "tuned" (or "trained") to fit the attributes of the existing elements to minimize the number of errors in their classification. However, a crucial issue for this procedure is not to "overtrain" the model so that it would have enough flexibility to classify new elements, which is the primal purpose of constructing the classifier. An example of hyperplanes separating the brain's pre-seizure, normal, and post-seizure states is presented in Fig. 3.

3.2. SVMs Mathematical Formulation. The main idea of applying SVMs to the classification of EEG time series is to embed EEG data (both normal and pre-seizure) into a higher-dimensional space and try to find a hyperplane that separates the data. The problem can be formally defined as follows. Let all EEG data samples be represented as *n*-dimensional vectors (or points in an *n*-dimensional space). A very common SVM approach consists of finding a plane that would separate all the vectors (points) in the *n*-dimensional space defined in A from the vectors in B. If such a plane is defined by the standard expression $x^T \omega = \gamma$, where $\omega = (\omega_1, \ldots, \omega_n)^T$ is an *n*-dimensional vector of real numbers and γ is a scalar, then this plane will separate all the elements from A and B. Thus, the discrimination rules can be formulated as an optimization problem of determining vectors ω and γ such that the separating hyperplane defines two open half spaces,

$$\{x \mid x \in \Re^n,\ x^T \omega < \gamma\}$$

and

$$\{x \mid x \in \Re^n,\ x^T \omega > \gamma\}$$

that contain most data points in A and B, respectively. However, in practice, it is usually impossible to perfectly separate two sets of elements by a plane. For this reason, one should try to minimize the average measure of misclassifications. Violations of these constraints are modeled by introducing nonnegative variables u and v. The most common mathematical model for SVMs that minimizes the total average measure of misclassification errors is written as follows:

$$\min_{\omega,\gamma,u,v} \frac{1}{m}\sum_{i=1}^{m} u_i + \frac{1}{k}\sum_{j=1}^{k} v_j, \tag{1}$$

$$\text{s.t. } A\omega + u \geq e\gamma + e, \tag{2}$$

$$A\omega - v \leq e\gamma - e, \tag{3}$$

$$u \geq 0, \; v \geq 0. \tag{4}$$

As one can see, this is a linear programming problem, and the decision variables here are the geometrical parameters ω and γ of the separating plane and also the variables u and v representing the misclassification error. Although this type of problems may involve high dimensionality of data in many cases, they can be efficiently solved by available LP solvers, for instance, Matlab, Xpress-MP, or CPLEX.

3.3. Time Series Kernel Functions. In this section, we will discuss the use of kernel functions, i.e., one of the most widely used technique in SVM learning, for time series classification. Generally, kernel functions are used to extend decision functions of SVMs to the nonlinear separation case. The main idea of kernel functions is to map data from the input space X into a high-dimensional feature space $\mathcal{X}$ by a function

$$\Phi : X \to \mathcal{X}$$

and solving the linear learning problem to find a separating hyperplane in $\mathcal{X}$. The actual kernel function Φ does not need to be known since it suffices to have a kernel function k that calculates the inner product in the feature space,

$$k(x, y) = \Phi(x) \cdot \Phi(y) \; .$$

A kernel function can be considered as a similarity (distance) measure in the input space [46]. The similarity between the samples x and y can be represented by the kernel function $k(x, y)$ as follows:

$$d^2(x, y) = (\Phi(x) - \Phi(y))^2 = k(x,x) - 2k(x, y) + k(y, y).$$

3.3.1. Linear Kernel. The most simple kernel function is the linear kernel $k(x, y) = x \cdot y$. A decision function can be specified by the formula $f(x) = wx + b$. In time series prediction, a linear kernel can be interpreted as a statistical autoregressive model of order k (AR[k]). This can be shown as follows: $x_T = f(x_{T-1}, \ldots, x_{T-k}) = \sum_{t=1}^{k} w_t x_{T-t} + b$. The interpretation of this kernel function is that time series are considered to be similar if they are generated by the same AR-model.

3.3.2. Radial Basis Function Kernel. Another commonly used kernel function is the radial basis function (RBF) kernel $k_\gamma(x, y) = \exp(-\gamma \| x - y \|^2)$. The similarity of two samples in terms of the RBF kernel can be interpreted as their Euclidian distance. In time series prediction, the RBF kernel, in turn, has a parallel in the phase space representation. This can be explained as follows. Assume that a time series is generated by a function f of the form $x_T = f(x_{T-1}, \ldots, x_{T-k})$. If one takes such a time series $x_1, \ldots, x_k, \ldots, x_N$ and plots it in a $(k+1)$-dimensional phase space, then it can be easily observed that the resulting plot is a part of the graph of f and, hence, the function f can be estimated from the time series. This is especially true if the function is assumed to be linear and the time series is generated by $x_T = f(x_{T-1}, \ldots, x_{T-k}) + \eta$, where η is a Gaussian noise. Clearly, the time series model is AR [1] and it can be shown that most of time series data points lie in an ellipsoid determined by the mean of the time series and the variance of η. The interpretation of this kernel function is that time series are considered to be similar in terms of means of the Euclidian distance in the phase space.

3.3.3. Fourier Kernel. The Fourier transform is among the most common transformations in time series analysis. The Fourier kernel function is advantageous when the information or pattern of a time series lies not in the individual values at each time point but in the frequency of some events. The inner product of the Fourier expansion of two time series can be directly calculated by the regularized kernel function $k_F(x, y) = \dfrac{1 - q^2}{2(1 - 2q\cos(x - y)) + q^2}$, where $0 < q < 1$ and $X \subset [0, 2\pi]^n$ [49].

4. METHODS

In this section, we describe methods used at each step of multi-dimensional EEG classification starting from the quantification of brain dynamics from EEG signals and ending with the implementation of the dynamic time warping kernel with SVMs to analyze multidimensional time series of brain dynamics.

4.1. Quantification of Brain Dynamics. The quantification of brain dynamics from EEGs that is used in this study is suitable for the investigation of a nonstationary system such as the brain because it is capable of automatically identifying and appropriately weighing existing transients in data. This technique is motivated by mathematical models from chaos theory used to characterize multi-dimensional complex systems and reduce the dimensionality of EEGs [1, 26, 37, 42, 48]. To quantify the brain dynamics, we divide EEG signals into sequential 10.24-second epochs (non-overlapping windows) to properly account for possible nonstationarities in the epileptic EEG. For each epoch of each channel of EEG signals, we estimate the measures of chaos to quantify the chaoticity of an attractor. These measures include Short-Term Maximum Lyapunov Exponent and Angular Frequency. A chaotic system like the human brain is a system in which orbits that originate from similar initial conditions or nearby points in the phase space diverge exponentially in an expansion process. The rate of divergence is an important aspect of a dynamical system and is reflected in the value of Lyapunov exponents and Angular Frequency. In other words, Lyapunov exponents and Angular Frequency measure the average uncertainty along the local eigenvectors and phase differences of an attractor in the phase space, respectively. Next, we will give a short overview of mathematical models used in the estimation of the Short-Term Maximum Lyapunov Exponent and Angular Frequency from EEG signals.

4.1.1. EEG Time Series Embedding. In the study of brain dynamics, the initial step in analyzing the dynamical properties of EEG signals is to embed it in a higher-dimensional space of dimension p, which enables us to capture the time behavior of the p variables that are primarily responsible for the dynamics of EEGs. We can now construct p-dimensional vectors $X(t)$ whose components consist of values of the recorded EEG signal $x(t)$ at p points in time separated by a time delay. The embedding phase space is constructed from a data segment $x(t)$ of duration T by the method of delays. Vectors X_i are constructed in the phase space as follows:

$$X_i = (x(t_i), x(t_i + \tau) \ldots x(t_i + (p-1)*\tau)),$$

where τ is the selected time lag between the components of each vector in the phase space, p is the selected dimension of the embedding phase space, and $t_i \in [1, T-(p-1)\tau]$. The vectors X_i in the phase space are illustrated in Fig. 4.

4.1.2. Estimation of the Short-Term Maximum Lyapunov Exponent ($STL_{\max}$). A method for estimation of $STL_{\max}$ for nonstationary data (e.g., EEG time series) is previously explained in [23, 51]. In this section, we will only give a short description and basic notation of our mathematical models used to estimate $STL_{\max}$. We first define the following notation:

— Δt is the evolution time for $\delta X_{i,j}$, i.e., the time during which $\delta X_{i,j}$ is allowed to evolve in the phase space. If the evolution time Δt is given in seconds, then L is expressed in bits per second.

— t_0 is the initial time point of the fiducial trajectory and coincides with the time point of the first data item in the data segment being analyzed. In estimating $STL_{\max}$, for a complete scan of the attractor, t_0 should move within $[0, \Delta t]$.

— N_a is the number of local $STL_{\max}$'s that will be estimated within a duration T of a data segment. Therefore, if D_t is the sampling period of the time-domain data, then we have $T = (N-1)D_t = N_a \Delta t + (p-1)\tau$.

— $X(t_i)$ is the point of the fiducial trajectory $\phi_t(X(t_0))$ with $t = t_i$, $X(t_0) = (x(t_0), \ldots x(t_0 + (p-1)*\tau))$, and $X(t_j)$ is a properly chosen vector adjacent to $X(t_i)$ in the phase space.

— $\delta X_{i,j}(0) = X(t_i) - X(t_j)$ is the displacement vector at t_i, i.e., a perturbation of the fiducial orbit at t_i, and $\delta X_{i,j}(\Delta t) = X(t_i + \Delta t) - X(t_j + \Delta t)$ is the evolution of this perturbation after a time increment Δt.

— $t_i = t_0 + (i-1)*\Delta t$ and $t_j = t_0 + (j-1)*\Delta t$, where $i \in [1, N_a]$ and $j \in [1, N]$, $j \neq i$.

$STL_{\max}$ is defined as the average of local Lyapunov exponents in the state space and can be calculated using the following equation:

$$STL_{\max} = \frac{1}{N_a \Delta t} \sum_{i=1}^{N_a} \log_2 \frac{|\delta X_{i,j}(\Delta t)|}{|\delta X_{i,j}(0)|}.$$

4.1.3. Estimation of Angular Frequency ($\overline{\Omega}$). As well as the estimation of $STL_{\max}$, the estimation of the Angular Frequency $\overline{\Omega}$ is motivated by the representation of a state as a vector in the state space. $\overline{\Omega}$ is merely an average uncertainty along the phase differences of an attractor in the phase space. We first define the difference in phase between two evolved states $X(t_i)$ and $X(t_i + \Delta t)$ as $\Delta\Phi_i$. Then, denoting by $(\Delta\Phi)$ the average of local phase differences $\Delta\Phi_i$ between vectors in

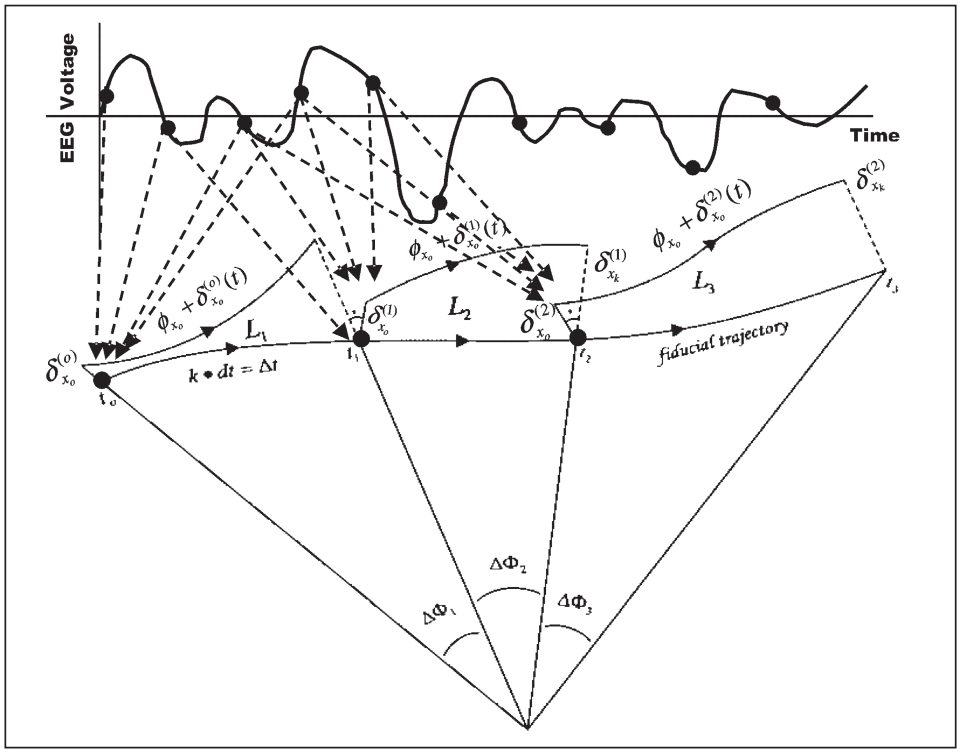

Fig. 4. Diagram illustrating an EEG epoch embedded in the phase space for quantification of brain dynamics for $p = 4$. The fiducial trajectory and the first three Lyapunov exponents ($L1$, $L2$, and $L3$) are shown.

the state space, we have

$$\Delta\Phi = \frac{1}{N_a}\sum_{i=1}^{N_a} \Delta\Phi_i,$$

where N_a is the total number of phase differences estimated from the evolution of $X(t_i)$ to $X(t_i + \Delta t)$ in the state space according to

$$\Delta\Phi_i = \left| \arccos \frac{X(t_i)\cdot X(t_i + \Delta t)}{\|X(t_i)\|\cdot\|X(t_i + \Delta t)\|} \right|.$$

Then, the average angular frequency $\overline{\Omega}$ is defined as

$$\overline{\Omega} = \frac{1}{\Delta t}\cdot\Delta\Phi.$$

If Δt is given in seconds, then $\overline{\Omega}$ is expressed in rad/sec. Thus, while $STL_{\max}$ measures the local stability of the state of the system on average, $\overline{\Omega}$ measures how fast a local state of the system changes on average (e.g., after dividing $\overline{\Omega}$ by 2π, the rate of the change of the state of the system is expressed in $\sec^{-1} = Hz$).

4.2. Dynamic Time Warping Kernel. Given two time series (or vector sequences) X and Y of equal length $|X| = |Y| = n$, pattern similarity is determined by aligning the time series X with the time series Y with the distortion of alignment $D_{\text{align}}(X, Y)$. Dynamic time warping (DTW) is used to compute the best possible alignment warp between two time series by selecting the one with the minimum distortion. In other words, the DTW distance is a distance measure (or a similarity measure) between two time series obtained by computing the best possible alignment or the minimum mapping (aligning) distance between two time series. In this study, all our EEG data samples are equal in length; however, the DTW can be extended to the case where the lengths of two time series are unequal. DTW was widely used in many contexts including data mining [30, 2], gesture recognition [16], robotics [44], speech processing [41, 47, 50], and medicine [8].

The problem of calculating the DTW distance can be solved using a dynamic programming approach. The basic concept can be described as follows. First, construct an alignment in an $n \times n$ matrix so that each vector (data points) in X is matched with the corresponding vector in Y. Typically, the Euclidean distance $d(x_i, y_j) = (x_i - y_j)^2$ is used as the local

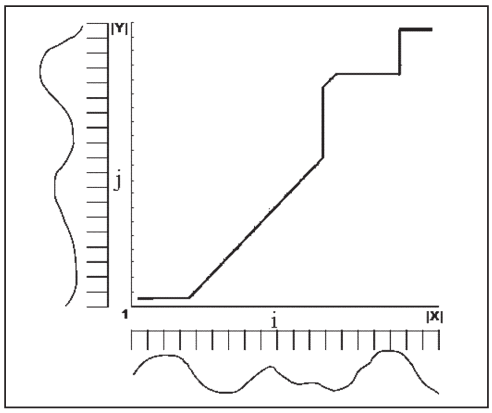

Fig. 5. A warping matrix with the minimum-distance warp path of two time series X and Y.

distance between two vectors, where the (i, j)th element of the matrix is the distance $d(x_i, y_j)$ between the ith point of the time series X and the jth point of the time series Y. Then we construct a warp path $W = w_1, \ldots, w_K$, where K is the length of the warp path and $\max(|X|,|Y|) \le K < |X|+|Y|$. The kth element of the warp path represents a matching point of two time series, $w_k = (i, j)$, where (i, j) corresponds to the index i of the time series X and the index j of the time series Y (see Fig. 5). The warp path must start at the beginning of each time series and finish at the end of both time series. In other words, the path starts from the beginning of each time series, $w_1 = (1,1)$ and finishes at the end of both time series, $w_K = (n, n)$. The warp path can actually be calculated in reverse order starting at the end of both time series. There is also a constraint on the warp path that forces the indices i and j to be monotonically increasing in the warp path, i.e., $w_k = (i, j)$ and $w_{k+1} = (i', j')$, where $i \le i' \le i+1$ and $j \le j' \le j+1$. Note that there can be an exponential number of warping paths that satisfy the above conditions. However, the optimal warp path is the one with a minimum warping (distortion) cost defined by the formula

$$D_{\text{align}}(X, Y) = \min \frac{1}{K} \sum_{k=1}^{K} d(w_{ki}, w_{kj}) \ .$$

In a dynamic programming approach, the warp path must either be incremented by one unit (adjacent) or stay at the same i or j axis. Therefore, we only need to evaluate the recurrence of the cumulative distance found in the adjacent elements,

$$D(i, j) = d(x_i, y_j) + \min \begin{cases} D(i, j-1), \\ D(i-1, j), \\ D(i-1, j-1). \end{cases}$$

4.3. SVMs with DTW Kernel. The essence of the SVMs framework with the DTW kernel in this application can be described as follows. Based on the DTW concept, the Euclidean distance measure is used to find the optimal path that minimizes the accumulated distance of the warping path. The SVMs with DTW uses the inner product or kernel function to find the optimal path that maximize the accumulated similarity (or minimize the distance) as follows:

$$k_{DTW}(x, y) = \max_{\psi_I, \psi_J} \frac{1}{M_\psi} \sum_{k=1}^{L} m(k) x_{\psi_I(k)} y_{\psi_J(k)}, \tag{5}$$

$$\text{s.t. } 1 \le \psi_I(k) \le \psi_I(k+1) \le L, \tag{6}$$

$$1 \le \psi_J(k) \le \psi_J(k+1) \le L, \tag{7}$$

where $L = |X| = |Y|$, $\psi_I(k)$ and $\psi_J(k)$ are linear warping functions, $m(k)$ is a nonnegative (path) weighting coefficient, and M_ψ is a (path) normalizing factor [47]. The linear discriminant function of SVMs with DTW kernel for time series classification can be then expressed in the same way as the original linear SVM function except using the DTW kernel. It is important to note that, in our case, we have multiple time series; therefore, the similarity of DTW kernels will be determined in a pair-wise manner. In other words, if we let N be the total number of electrodes, then we have to calculate a total of $\frac{N(N-1)}{2}$ kernel functions or similarity indices. These similarity indices can in turn be considered as attributes of input data. Finally, note that the same algorithms used to solve standard SVMs can be used to solve SVMs with DTW kernel as well.

TABLE 1. EEG Dataset Characteristics

Patient ID	Gender	Age	Seizure Types	Duration of EEG, days	Number of Seizure
1	F	41	CP	9.06	24
2	M	45	CP, SC	3.63	9
3	M	29	CP, SP	6.07	19
Total				18.76	59

CP — Complex Partial; SC — Subclinical

5. EMPIRICAL STUDY

The underlying hypothesis in this empirical study is that the proposed SVMs with DTW kernel are capable of discriminating/classifying different physiological stages (normal and pre-seizure) of the brain. The features of input data are represented in the form of time series of the brain dynamics measure (i.e., $STL_{\max}$, and $\overline{\Omega}$). In this section, we discuss in detail each step of our empirical study.

5.1. EEG Data Acquisition. In this study, datasets consisted of continuous long-term (3 to 9 days) multichannel intracranial EEG recordings from macroelectrodes bilaterally surgically implanted in the hippocampus and temporal and frontal lobe cortices of 3 epileptic patients with medically intractable temporal lobe epilepsy (outlined in Table 1). These recordings were obtained as a part of a pre-surgical clinical evaluation, using a Nicolet BMSI 4000 recording system with amplifiers of an input range of 0.6 mV, sampling rate of 200 Hz, and filters with the frequency range 0.5-70Hz. Each recording included a total of 28 to 32 intracranial electrodes (8 subdural and 6 hippocampal depth electrodes for each cerebral hemisphere, and a strip of 4 additional electrodes if they deemed necessary by the neurologist). Note that, in this study, we use EEG recordings only from 26 electrodes since such electrodes are most commonly used. The recorded EEG signals were digitized and stored on magnetic media for subsequent off-line analysis. These EEG recordings were viewed by two independent electroencephalographers to determine the number and type of recorded seizures, seizure onset and end times, and seizure onset zones.

5.2. Data Sampling and Pre-Processing. In this study, the classification will be performed separately for each subject. We use the Monte-Carlo sampling technique to randomly select EEG data from 2 groups (normal and pre-seizure states) per individual from continuous recordings. Each data sample contains a 5-minute epoch of EEG data from 26 electrodes. Note that, in this analysis, we consider only clinical seizures and un-clustered seizures. In the data set, we consider 22, 7, and 15 seizures in the EEG data for Patients 1, 2, and 3, respectively. Since the lengths and total numbers of seizures of the data sets of the patients are very different, we randomly select three epochs of pre-seizure EEG data per seizure for each patient. In other words, 66, 21, and 45 epochs of the EEG data are selected from the group of pre-seizure EEG data for Patients 1, 2 and 3, respectively. 200 epochs of EEG data per patient for the normal state are randomly and uniformly sampled. The criteria used in determining normal and pre-seizure states of EEG data are as follows. Normal EEG samples are selected from EEG recordings that are more than 8 hours apart from a seizure. Pre-seizure EEG samples are selected from EEG recordings during the 30-minute interval before a seizure. For instance, we analyze 22 seizures from Patient 1's data; therefore, 266 EEG epochs (200 normal and 66 pre-seizure) are sampled. After sampling EEG data, we first calculate measures of brain dynamics (i.e., $STL_{\max}$ and $\overline{\Omega}$) from EEG data using the methods described in Sec. 4. Each measure is calculated continuously for each non-overlapping 10.24-second segment of EEG data; therefore, each of EEG epochs contains 30 data points of the brain dynamical time series.

5.3. Classification Procedure. The input data consist of 66, 21, and 45 $N * m$-dimensional time series, where N is the number of electrodes and m is the length of each EEG epoch times two (2 dynamical measures). We then calculate a pair-wise kernel function for every electrode pair of multi-dimensional EEG time series. Therefore, the total number of feature vectors corresponding to Patients 1, 2, and 3 is $\frac{N(N-1)}{2} * 2 = 650$. The method used to calculate the kernel function is described in Sec. 4. Subsequently, we employ SVMs to classify these EEG data. We then use Matlab to solve the SVMs model constructed to find the plane that would separate all the vectors of normal and pre-seizure EEGs.

		Predict	
		Abnormal	Normal
Actual	Abnormal	True Positive	False Negative
	Normal	False Positive	True Negative

Fig. 6. The concept of evaluation of classification results. Note that we use "pre-seizure" and "abnormal" interchangeably.

TABLE 2. Performance Characteristics of the Support Vector Machine Classifier for an Individual Patient

Patient	SVMs [11]			SVMs-DTW		
	Sensitivity, %	Specificity, %	Overall, %	Sensitivity, %	Specificity, %	Overall, %
1	81.21	87.46	84.34	92.15	93.84	92.99
2	71.18	76.85	74.02	79.45	78.07	78.76
3	74.13	70.60	72.37	80.13	84.66	82.39
Average	75.51	78.30	76.91	83.91	85.52	84.72

5.4. Training and Testing. In this section, we describe how SVMs are trained and tested. There are many choices in dividing data into training and test sets. In order to reduce the bias of training and test data, we propose to implement a leave-one-out cross validation scheme extensively used as a method for estimating the generalization error based on "resampling." It is important to note that the classification techniques are trained and tested individually for each patient. To train SVMs, it is important to note that, in general, the training of support vectors machines is optimized when the numbers of pre-seizure and normal samples are comparable. Otherwise, SVMs will be biased to classify most samples to the physiological state with larger size samples. To adequately implement SVMs, we train the classifier with the same number of pre-seizure and normal samples by implementing Monte Carlo sampling simulation. First, we shuffle (in a random manner) the pre-seizure and normal samples individually. Since the size of pre-seizure samples is much larger than the size of normal samples, the number of pre-seizure samples will be used to determine the size of the training and testing sets. Then we divide the first of pre-seizure samples for the training and the other half for the testing. After that, we randomly select training data (with the same size) from normal samples. For an individual patient, we run the simulation 100 times.

6. RESULTS

To evaluate the performance characteristics of the proposed classification technique, we calculate the sensitivity and specificity of the proposed classification technique. These results will be discussed in this section.

6.1. Performance Evaluation of Classification Schemes. In general, to evaluate the classifier, we categorize the classification into two classes: positive (pre-seizure) and negative (normal). Then we consider the following four subsets of classification results: 1. True positives (TP) denote correct classifications of positive cases. 2. True negatives (TN) denote correct classifications of negative cases. 3. False positives (FP) denote incorrect classifications of negative cases into the class of true positives. 4. False negatives (FN) denote incorrect classifications of positive cases into the class of true negatives.

To better explain the concept of evaluation of classifiers, let us consider the case of detection of pre-seizure EEG data (see Fig. 6). A classification result was considered to be true positive if we classify a pre-seizure EEG sample as a pre-seizure sample. A classification result was considered to be true negative if we classify a normal EEG sample as a normal sample. A classification result was considered to be false positive when we classify a normal EEG sample as a pre-seizure sample. A classification result was considered to be false negative when we classify a pre-seizure EEG sample as a normal sample.

Sensitivity and specificity are widely used in the medical domain as classification performance measures Sensitivity measures the fraction of positive cases that are classified as positive. Specificity measures the fraction of negative cases classified as negative. Sensitivity and specificity are defined as follows:

$$\text{Sensitivity} = \frac{TP}{TP + FN}, \quad \text{Specificity} = \frac{TN}{TN + FP}.$$

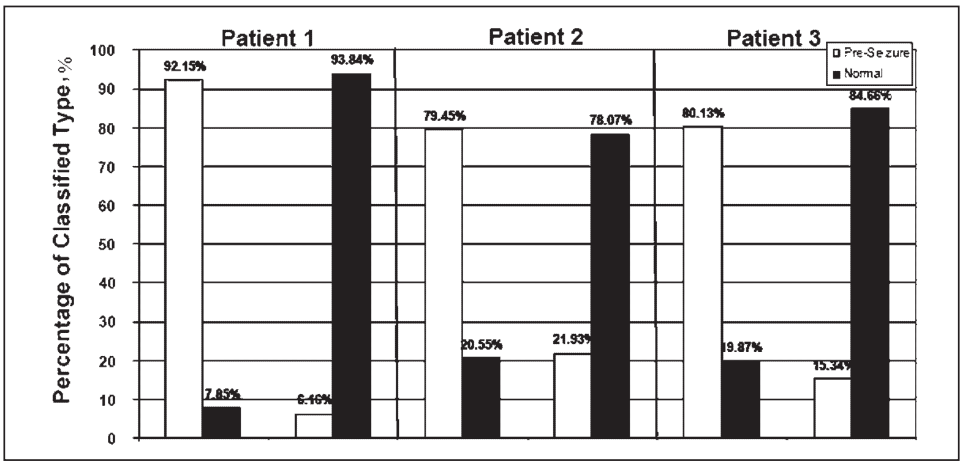

Fig. 7. Average classification results for all three patients obtained using leave-one-out cross validation to train and test the algorithm over 100 simulations.

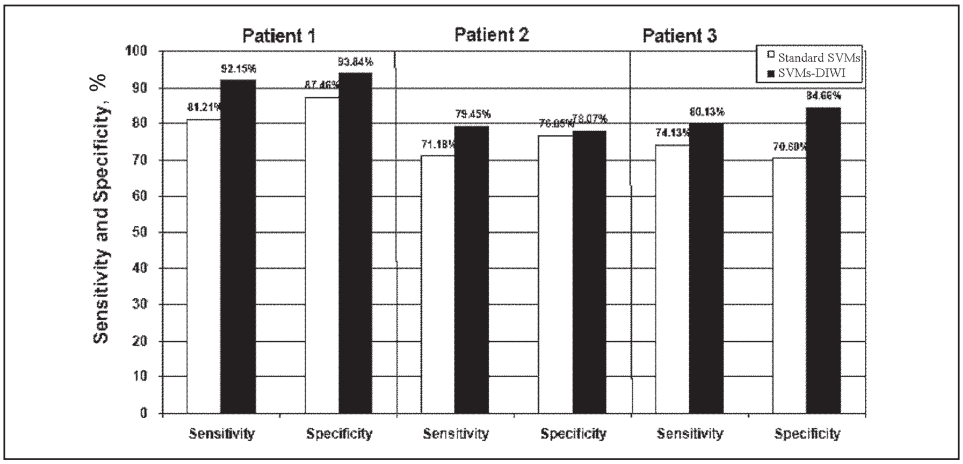

Fig. 8. Performance comparison of the standard SVMs for EEG classification proposed in [11] versus the proposed SVMs-DTV algorithm for all three patients obtained using leave-one-out cross validation to train and test the algorithm over 100 simulations.

In fact, sensitivity can be considered as the probability of accurately classifying EEG samples in the pre-seizure case. The specificity can be considered as the probability of accurately classifying EEG samples in the normal case. In general, a good classifier is that with high sensitivity and specificity.

6.2. Performance Characteristics of the Proposed Classification Methods. In this section, we report the average classification performance obtained after running 100 simulations. Figure 7 illustrates the overall classification results of the proposed SVMs-DTW algorithm. Table 2 and Fig. 8 illustrate a performance comparison of the standard SVMs for the EEG classification proposed in [11] versus the results obtained from the SVMs with DTW proposed in this paper and tested on 3 patients. In all cases, the incorporation of the DTW kernel function allows us to achieve substantially better classification results (about 8% better on average). For Patient 1, the proposed algorithm achieved about 92% sensitivity and over 93% specificity on average. This result demonstrates the improvement in classification performance of almost 8% on average. For Patient 2, the proposed algorithm achieved sensitivity of about 80% and specificity of about 78% on average. This result demonstrates the improvement in classification performance of almost 5% on average. For Patient 3, the proposed algorithm achieved sensitivity of about 84% and specificity of over 85% on average. This result demonstrates the improvement in classification performance of over 10% on average. Overall, the proposed SVMs-DTW can achieve the sensitivity of 83.91% for correctly classifying pre-seizures and the specificity of 85.52% for correctly classifying normal EEGs. This reflects an improvement of almost 8% in the classification accuracy. In Fig. 8, we observe that the incorporation of the DTW kernel function can improve the classification performance of SVMs in every case.

It is very interesting to note that the classification performance of our algorithm for each patient is consistent with the standard SVMs [11]. Specifically, the EEG data from Patient 1 tend to be more classifiable than those of Patients 2 and 3.

We speculate that the number of seizures in an EEG data set could play a very important role in terms of providing more training data for abnormal (pre-seizure) EEGs. Since our algorithm yields the worst classification results for Patient 2 among all 3 cases, it is very intuitive to claim that there is so much the classifier can learn from 7 seizure samples as apposed to 22 and 15 samples. Nonetheless, these results confirm our hypothesis that the brain's states are classifiable based on the brain dynamics measures and data mining techniques applied to EEG signals. The framework of classifiers proposed in this study can be extended to the development of an abnormal brain activity classifier or an online brain activity monitoring.

7. CONCLUSIONS

This study addresses the open question of the classifiability of the brain's pre-seizure and normal EEGs. The results of this study are another proof of the concept of application of quantification of brain dynamics and data mining techniques. This framework was proved successful in providing insights and characterizing different states of brain activities reflected from pathological dynamical interactions of brain network. In addition, these results also confirm our hypothesis that it is possible to differentiate and classify the brain's pre-seizure and normal activities based on optimization, data mining, and dynamical system approaches in multichannel intracranial EEG recordings. Moreover, the incorporation of DTW kernel function with SVMs is very straightforward and can be easily implemented. Optimization problems in the framework of support vector machines can be solved in a reasonable time. All the programming efforts were made in the Matlab environment on a desktop computer Pentium IV 2.4 GHz with 1 GB of RAM. The proposed technique is very fast and scalable. The running time for the statistical cross-validation technique is less than 5 minutes on average. In the future, more cases (patients and seizures) will be studied to validate the observation across patients and also the development of a multi-class classifier based on the support vector machines framework. In addition, the feature selection study will be possible in the future. This study will help us to select electrodes that show prominent changes, which might lead us to the solution to the epileptogenic zone localization problem.

REFERENCES

1. A. Babloyantz and A. Destexhe, "Low dimensional chaos in an instance of epilepsy," Proc. Nat. Acad. Sci. USA, **83**, 3513–3517 (1986).
2. D. Berndt and J. Clifford, "Using dynamic time warping to find patterns in time series," in: Proc. of the AAAI-94 Workshop on Knowledge Discovery in Databases (KDD-94) (1994).
3. D. Bertsimas, C. R. Darnell, and R. Soucy, "Portfolio construction through mixed-integer programming at Grantham, Mayo, Van Otterloo and Company," Interfaces, **29**, No. 1, 49–66 (1999).
4. M. K. Borgwardt, S. V. N. Vishwanathan, and H-P. Kriegel, "Class prediction from time series gene expression profiles using dynamical systems kernels," in: Pacific Symp. on Biocomput (2006), pp. 547–558.
5. P. S. Bradley, U. Fayyad, and O. L. Mangasarian, "Mathematical programming for data mining: Formulations and challenges," INFORMS J. Computing, **11**, 217–238 (1999).
6. P. S. Bradley, O. L. Mangasarian, and W. N. Street, "Clustering via concave minimization," in: M. C. Mozer, M. I. Jordan, and T. Petsche (eds.), Adv. in Neural Inform. Proces. Systems, MIT Press, Cambridge (1997), pp. 368–374.
7. L. Breiman, J. Friedman, R. Olsen, and C. Stone, Classification and Regression Tree, Wadsworth Inc., Belmont (1993).
8. E. G. Caiani, A. Porta, G. Baselli, et al., "Warped-average template technique to track on a cycle-by-cycle basis the cardiac filling phases on left ventricular volume," IEEE Comput. in Cardiology, **25**, No. 98, CH36292 (1998).
9. W. A. Chaovalitwongse, P. M. Pardalos, L. D. Iasemidis, et al., "Applications of global optimization and dynamical systems to prediction of epileptic seizures," in: P.M. Pardalos, J.C. Sackellares, L. D. Iasemidis, and P. R. Carney (eds.), Quantitative Neuroscience, Kluwer, Dordrecht (2003), pp. 1–36.
10. W. A. Chaovalitwongse, P. M. Pardalos, and O. A. Prokoyev, "Reduction of multi-quadratic 0-1 programming problems to linear mixed 0-1 programming problems," Oper. Res. Letters, **32**, No. 6, 517–522 (2004).
11. W. A. Chaovalitwongse, P. M. Pardalos, and O. A. Prokoyev, "Electroencephalogram (EEG) time series classification: Applications in epilepsy," Ann. Oper. Res., **148**, No. 1, 227–250 (2006).
12. D. Dasgupta and S. Forrest, "Novelty detection in time series data using ideas from immunology," Intern. Conf. on Intell. Systems (1999).

13. J. J. R. Diez and C. A. Gonzalez, "Applying boosting to similarity literals for time series classification," in: Intern. Workshop on Multiple Classifier Systems (2000), pp. 210–219.
14. C. E. Elger and K. Lehnertz, "Seizure prediction by non-linear time series analysis of brain electrical activity," Eur. J. Neurosci., **10**, 786–789 (1998).
15. G. M. Fung and O. L. Mangasarian, "Proximal support vector machines," in: 7th ACM SIGKDD Intern. Conf. on Knowledge Discovery and Data Mining (2001).
16. D. M. Gavrila and L. S. Davis, "Towards 3-d model-based tracking and recognition of human movement: A multi-view approach," in: Proc. Intern. Workshop on Autom. Face- and Gesture-Recognition (1995).
17. P. Geurts, "Pattern extraction for time series classification," in: 5th Eur. Conf. on Principles of Data Mining and Knowledge Discovery (2001), pp. 115–127.
18. R. L. Grossman, C. Kamath, P. Kegelmeyer, et al., Data Mining for Scientific and Engineering Applications, Kluwer, Dordrecht (2001).
19. D. J. Hand, H. Mannila, and P. Smyth, Principle of Data Mining, Bradford Books, Concord (2001).
20. C.-W. Hsu and C.-J. Lin, "A comparison of methods multi-class support vector machines," IEEE Trans. on Neural Networks, **13**, 415–425 (2002).
21. A. S. Hurn, K. A. Lindsay, and C. A. Michie, "Modelling the lifespan of human *t*-lymphocite subsets," Math. Biosciences, **143**, 91–102 (1997).
22. F. J. Iannatilli and P. A. Rubin, "Feature selection for multiclass discrimination via mixed-integer linear programming," IEEE Trans. on Pattern Analysis and Machine Learning, **25**, 779–783 (2003).
23. L. D. Iasemidis, "On the dynamics of the human brain in temporal lobe epilepsy," PhD Thesis, Univ. of Michigan, Ann Arbor (1991).
24. L. D. Iasemidis, P. M. Pardalos, J. C. Sackellares, and D.-S. Shiau, "Quadratic binary programming and dynamical system approach to determine the predictability of epileptic seizures," J. Comb. Optimiz., **5**, 9–26 (2001).
25. L. D. Iasemidis, D.-S. Shiau, W. A. Chaovalitwongse, et al., "Adaptive epileptic seizure prediction system," IEEE Trans. Biomed. Eng., **5**, No. 5, 616–627 (2003).
26. L. D. Iasemidis, H. P. Zaveri, J. C. Sackellares, and W. J. Williams, "Phase space analysis of eeg in temporal lobe epilepsy," in: 10th Ann. Intern. Conf. "IEEE Eng. in Medicine and Biology Soc." (1988), pp. 1201–1203.
27. E. Keogh, K. Chakrabarti, M. Pazzani, and S. Mehrotra, "Dimensionality reduction for fast similarity search in large time series databases," Knowledge and Inform. Systems, **3**, No. 3, 263–286 (2000).
28. E. Keogh and S. Kasetty, "On the need for time series data mining benchmarks: A survey and empirical demonstration," in: 8th ACM SIGKDD Intern. Conf. on Knowledge Discovery and Data Mining (2002), pp. 102–111.
29. E. Keogh and M. Pazzani, "An enhanced representation of time series which allows fast and accurate classification, clustering and relevance feedback," in: 4th Intern. Conf. on Knowledge Discovery and Data Mining (1998), pp. 239–241.
30. E. Keogh and M. Pazzani, "Scaling up dynamic time warping for datamining applications," in: Proc. 6th ACM SIGKDD Intern. Conf. on Knowledge Discovery and Data Mining (2000), pp. 285–289.
31. U. Krebel, "Pairwise classification and support vector machines," in: Adv. in Kernel Methods–Support Vector Learning, MIT Press, Cambridge (1999), pp. 255-268.
32. K. Lehnertz and C. E. Elger, "Can epileptic seizures be predicted? Evidence from nonlinear time series analysis of brain electrical activity," Phys. Rev. Lett., **80**, 5019–5022 (1998).
33. B. Litt, R. Esteller, J. Echauz, et al., "Epileptic seizures may begin hours in advance of clinical onset: A report of five patients," Neuron, **30**, 51–64 (2001).
34. O. L. Mangasarian, "Linear and nonlinear separation of pattern by linear programming," Oper. Res., **31**, 445–453 (1965).
35. O. L. Mangasarian, W. N. Street, and W. H. Wolberg, "Breast cancer diagnosis and prognosis via linear programming," Oper. Res., **43**, No. 4, 570–577 (1995).
36. J. Martinerie, C. Van Adam, and M. L. Van Quyen, "Epileptic seizures can be anticipated by non-linear analysis," Nature Medicine, **4**, 1173–1176 (1998).
37. N. H. Packard, J. P. Crutchfield, and J. D. Farmer, "Geometry from time series," Phys. Rev. Lett., **45**, 712–716 (1980).
38. P. M. Pardalos, W. A. Chaovalitwongse, L. D. Iasemidis, et al., "Seizure warning algorithm based on spatiotemporal dynamics of intracranial EEG," Math. Program., **101**, No. 2, 365–385 (2004).

39. J. R. Quinlan, C4.5: Programs for Machine Learning, Morgan Kaufmann, Orlando (1993).
40. M. L. Van Quyen, J. Martinerie, M. Baulac, and F. Varela, "Anticipating epileptic seizures in real time by non-linear analysis of similarity between eeg recordings," Neuro Rep., **10**, 2149–2155 (1999).
41. L. Rabiner and B. Juang, Fundamentals of Speech Recognition, Prentice Hall, Upper Saddle River (1993).
42. P. E. Rapp, I. D. Zimmerman, and A. M. Albano, "Experimental studies of chaotic neural behavior: Cellular activity and electroencephalographic signals," in: H. G. Othmer (ed.), Nonlinear Oscillations in Biology and Chemistry, Springer, New York (1986), pp. 175–205.
43. S. Rüping, "SVM kernels for time series analysis," in: R. Klinkenberg, S. Rüping, A. Fick, N. Henze, C. Herzog, R. Molitor, and O. Schröder (ed.), LLWA 01-Tagungsband der GI-Workshop-Woche Lernen-Lehren-Wissen-Adaptivitet (2001), pp. 43–50.
44. M. Schmill, T. Oates, and P. Cohen, "Learned models for continuous planning," in: Proc. 7th Intern. Workshop on Artif. Intell. and Statist. (1999), pp. 278–282.
45. B. Scholkopf, C. Burges, and V. Vapnik, "Extracting support data for a given task," in: Proc. 1st Intern. Conf. on Knowledge Discovery and Data Mining, AAAI Press, Menlo Park (1995).
46. B. Schölkopf, "The kernel trick for distances," in: Techn. Rep., Microsoft Research (2000).
47. H. Shimodaira, K. Noma, M. Naka, and S. Sagayama, "Support vector machine with dynamic time-alignment kernel for speech recognition," in: Proc. of Eurospeech (2001), pp. 1841–1844.
48. F. Takens, "Detecting strange attractors in turbulence," D. A. Rand and L. S. Young (eds.), Dynamical Systems and Turbulence, Lecture Notes in Mathematics, Springer-Verlag, Berlin (1981).
49. V. N. Vapnik, The Nature of Statistical Learning, Springer, Berlin (1995).
50. V. Wan and J. Carmichael, "Polynomial dynamic time warping kernel support vector machines for dysarthric speech recognition with sparse training data," in: Proc. of Interspeech (2005), pp. 3321–3324.
51. A. Wolf, J. B. Swift, H. L. Swinney, and J. A. Vastano, "Determining Lyapunov exponents from a time series," Physica D, **16**, 285–317 (1985).
52. K. Yang and C. Shahabi, "A pca-based kernel for kernel pca on multivariate time series," in: Proc. of ICDM 2005 Workshop on Temporal Data Mining: Algorithms, Theory and Applications held in conjunction with the 5th IEEE Intern. Conf. on Data Mining (ICDM'05) (2005).