Pattern- and Network-Based Classification Techniques for Multichannel Medical Data Signals to Improve Brain Diagnosis

Wanpracha Art Chaovalitwongse, *Senior Member, IEEE*, Rebecca S. Pottenger, Shouyi Wang, Ya-Ju Fan, and Leon D. Iasemidis, *Member, IEEE*

***Abstract*—There is an urgent need for a quick screening process that could help neurologists diagnose and determine whether a patient is epileptic versus simply demonstrating symptoms linked to epilepsy but actually stemming from a different illness. An inaccurate diagnosis could have fatal consequences, particularly in operating rooms and intensive care units. Electroencephalogram (EEG) has been traditionally used, as a gold standard, to diagnose patients by evaluating those brain functions that might correspond to epilepsy and other brain disorders. This research therefore focuses on developing new classification techniques for multichannel EEG recordings. Two time-series classification techniques, namely, Support Feature Machine (SFM) and Network-Based Support Vector Machine (SVM) (NSVM), are proposed in this paper to predict from EEG readings whether a person is epileptic or nonepileptic. The SFM approach is an optimization model that maximizes classification accuracy by selecting a group of electrodes (features) that has strong class separability based on time-series similarity measures and correctly classifies EEG samples in the training phase. The NSVM approach integrates a new network-based model for multidimensional time-series data with traditional SVMs to exploit both the spatial and temporal characteristics of EEG data. The proposed techniques are tested on two EEG data sets acquired from ten and five patients, respectively. Compared with other commonly used classification techniques such as SVM and decision trees, the proposed SFM and NSVM techniques provide very promising and practical results and require much less time and memory resources than traditional techniques. This study is a necessary application of data mining to advance the diagnosis and treatment of human epilepsy.**

***Index Terms*—Electroencephalogram (EEG) classification, epilepsy diagnosis, multidimensional time series, optimization, pattern recognition.**

Manuscript received September 30, 2009; revised April 1, 2010 and September 12, 2010; accepted October 30, 2010; date of current version August 23, 2011. The work of W. Chaovalitwongse, S. Wang, Y.-J. Fan, and R. S. Pottenger was supported in part by the National Science Foundation under a CAREER Grant and in part by the Center for Discrete Mathematics and Theoretical Computer Science under the Research Experiences for Undergraduates Program. The work of L. D. Iasemidis was supported by the National Institutes of Health, Veterans Affairs, Defense Advanced Research Projects Agency, and the Whitaker Foun-dation. This paper was recommended by Associate Editor J. M. Carmena.

W. A. Chaovalitwongse, S. Wang, and Y.-J. Fan are with the Department of Industrial and Systems Engineering, Rutgers University, Piscataway, NJ 08855-0909 USA (e-mail: wchaoval@rci.rutgers.edu; shouyi@eden.rutgers.edu; yjfan@eden.rutgers.edu).

R. S. Pottenger is with the Department of Computer Science, Princeton University, Princeton, NJ 08540-5233 USA (e-mail: rpotteng@princeton.edu).

L. D. Iasemidis is with the Department of Bioengineering, Arizona State University, Tempe, AZ 85287 USA (e-mail: Leon.Iasemidis@asu.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier

I. Introduction

EPILEPSY, a disease characterized by a tendency for recurrent seizures, is one of the most common brain disorders in the world, coming second only to strokes. Currently, about 3 million Americans and 40 million people worldwide (about 1% of human population) suffer from epilepsy [1]–[3]. As common as epilepsy is, the accuracy of epilepsy diagnosis varies greatly, from a misdiagnosis rate of 5% in a prospective childhood epilepsy study to at least 23% in a British population-based study [4]. In fact, the rate may be even higher in everyday practice. For instance, temporal lobe epilepsy is a less common form of epilepsy that does not result in the typical physical seizures. Rather, patients suffer from symptoms such as depression, moodiness, anger, or irritability. Misdiagnosis of this condition as depression is extremely common [3]. In today's brain diagnosis studies, particularly in epilepsy studies, the electroencephalogram (EEG) recordings are the most commonly used neurophysiological signal employed to evaluate brain functions that might be related to brain disorders and abnormal cognitive functions. Neurologists are trained to recognize certain prominent patterns in EEG signals that reflect the brain's activity. For diagnosis purposes, pattern recognition is a natural method that neurologists employ to identify the presence of a disease such as epilepsy. However, neurologists have to "eye ball" EEG signals, spatially and temporally, in an attempt to recognize abnormal patterns (e.g., epileptiforms) in the brain activity. "Eye balling" these massive signals for hours or even days can be very tedious and challenging. For neurologists, there is an urgent need for new automated signal-processing and pattern-recognition techniques that help these physicians diagnose, with more accuracy and speed, patients who have epilepsy or other related brain disorders.

The main application of this study is to improve current *epilepsy diagnosis* by developing a new medical signal-pattern-recognition framework to identify (or distinguish) abnormal spatiotemporal patterns from multichannel EEG recordings. In this computational framework, we develop two new classification techniques for multidimensional time-series classification, namely, support feature machine (SFM) and network-based support vector machine (SVM) (NSVM). SFM is a pattern-based classification technique employing the nearest neighbor rule and time-series similarity measures, whereas its optimization model selects the features with strong class separability so that the classification accuracy is maximized. NSVM is a

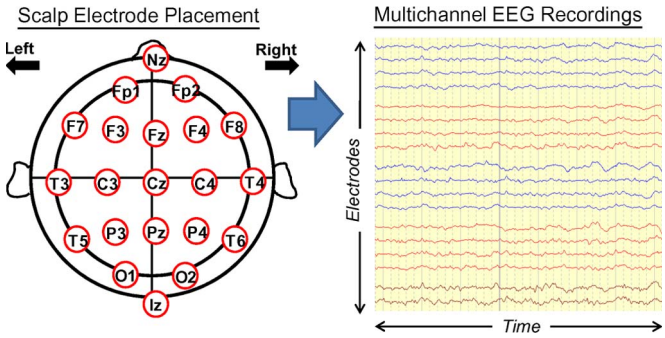

Fig. 1. Simplified scalp EEG recording placement and an example of 10-s EEG recordings.

network-based version of SVM that incorporates statistical and correlation measures of temporal synchronization among time-series profiles. It can overcome the drawback of SVM regarding time-series classification as SVM generally treats each time stamp of a time-series epoch as an *independent* attribute although data in the epoch are actually highly correlated. We evaluate and assess the performances of the proposed techniques on an EEG data set acquired from ten patients, five epileptic and five nonepileptic, during their routine EEG checks. The proposed framework can be applied as a medical decision-support system to improve the current medical diagnosis and prognosis by assisting physicians in recognizing via data mining abnormal patterns in complex medical data, particularly EEG recordings. Such a system could be used in a quick screening process that could determine whether a patient is epileptic or nonepileptic.

The organization of the succeeding sections of this paper is as follows. In Section II, the research background and previous work are discussed. In Section III, the SFM and NSVM frameworks including time-series similarity measures and optimization models are described. Section IV describes the acquisition and cleansing procedures of the EEG data set and the design of experiments. The computational results and the performance characteristics of our classification techniques are provided in Section V. The concluding remarks are given in Section VI.

II. Background

A. Epilepsy Diagnosis

There are many diseases that cause changes in brain behavior and can be confused with epilepsy [3]; specifically, several medical conditions can cause seizures or seizurelike episodes. Therefore, the evaluation of patients with these symptoms is aimed at determining the type (epileptic or nonepileptic) and cause of the seizures. Out of the commonly used medical tests such as blood tests, magnetic resonance imaging, positron emission tomography, as well as studying the patient's medical history, EEG reading is *the most important* part of epilepsy diagnosis because it directly detects electrical activity in the brain. During an EEG test, electrodes are attached to specific locations on the patient's scalp (see Fig. 1 for simplified scalp EEG recordings). A routine EEG test for epilepsy diagnosis usually records about 20–30 min of brain waves; however, in most cases, the results of routine EEG studies are often inconclusive, even in people known to have epilepsy, so prolonged (24-h) EEG monitoring can be necessary. Epilepsy diagnosis can be very complicated, particularly in immediate life-and-death situations [e.g., in emergency rooms (ERs)] when the decision needs to be made promptly. Also, in many cases of coma, trauma, and surgical critical care, the medical diagnostic tools used to differentiate epileptic seizures from other symptoms that have similar EEG morphological patterns are very critical to the patients' welfare. Inaccurate diagnosis and treatment could have severe consequences, particularly in life-threatening situations in ERs and intensive care units. An accurate quick EEG analysis that can identify whether a patient has epilepsy could drastically improve the accuracy of diagnosis of epilepsy and thereby save patients' lives. There is a desperate need for a new technology providing quick and accurate epilepsy screening, which would serve as an initial medical diagnostic tool.

This study is of significant importance to improving current brain diagnosis as it offers great potential for the development of computerized techniques to differentiate between EEG patterns from epileptic patients and those with different diseases but morphologically similar EEG patterns. This study presents a framework for a quick EEG analysis and screening process that has an overarching potential to help save lives, improve the accuracy of medical diagnosis, and reduce associated healthcare costs.

B. Pattern Recognition for Neurophysiological Signals

Over the past decade, there have been a number of research studies in quantitative signal-processing techniques (both univariate and multivariate) applied to neurophysiological data such as EEG. For EEG analysis, linear univariate techniques (e.g., power spectrum analysis and time–frequency analysis) have often been used in conjunction with nonlinear methods which incorporate high-order statistics, nonlinear dynamics (e.g., chaos theory), and information-theoretical quantification. New similarity measures called the Kullback–Leibler discrimination information and the Chernoff information for discrimination between multivariate series in the multivariate non-Gaussian case were proposed in [5]. Subsequently, neural network methods were developed for a nonlinear aspect of analysis for multivariate time series, where a unified view of nonlinear principal component analysis, nonlinear canonical correlation analysis, and nonlinear singular spectrum analysis techniques were presented [6]. A constructive induction method for classifying time series was proposed in [7], where the scope of attribute-value learning was expanded by using metafeatures to the domains that contain instances with recurring substructure. A new technique that vectorizes components of a correlation coefficient matrix of multidimensional time-series data was proposed in [8]. Recently, there have been a few studies regarding detecting changes in spatiotemporal data [9], [10]. One such study was conducted to detect and identify changes in brain waves through event-related potential data [10]. However, the problem of detecting spatiotemporal changes in known EEG patterns is much easier than the problem of finding unseen patterns in EEGs. Despite converging evidence and consistency

in reported findings on the potential usefulness of these univariate and multivariate techniques, their added value to the diagnosis of brain disorders in clinical settings remains questionable. Complex brainwave patterns and their relationships to brain disorders may be very specific for an individual patient, yet vary from one patient to another. This suggests the need for a new development of data-analysis and signal-pattern-recognition techniques that allow the identification of more complex relationships.

C. EEG Classification

Over the past few years, there has been a substantial body of work in EEG classification. In the previous studies by our group [11]–[14], it is suggested that EEG signals of epileptic patients during normal and preseizure states may be classifiable. In those studies, we employed a feature-extraction technique based on the chaos theory to characterize nonlinear dynamical patterns in EEG signals. That technique was motivated by mathematical models used to characterize multidimensional complex systems and by the prospect of reduction of the dimensionality of the EEG data [15]–[19]. In our first attempt to classify EEG signals, we implemented the SVM approach to classify normal and preseizure EEGs with some degree of success [11]. However, the classification results were inferior to a standard nearest neighbor classification. In a more recent study, we proposed a time-series k-nearest neighbor approach with advanced time-series similarity approaches (e.g., t-index and dynamic time warping) and tested it on a larger EEG data set [12]. The main drawback of traditional pattern-based (or instance-based) classification techniques like k-nearest neighbor is their sensitivity to noisy features because most techniques do not incorporate the feature-selection process. Although the feature-selection process can be carried out as a separate step before classification, the process is usually not trained (supervised) to select the best combination of features (which requires combinatorial optimization) as each feature is evaluated on a one-to-one basis. Our group has recently proposed a new optimization framework for supervised feature selection with k-nearest neighbor classification [13], [14]. The framework has been very effective in classifying normal and preseizure EEGs. Based on our initial discovery of detectable EEG pattern changes before a seizure, we believe that the concept of EEG classification may be applicable to the epilepsy diagnosis problem. It is important to note that, in all of our previous studies, the classification was done on individual patients, where we used the training and testing data from the same patient. In addition, we tested the hypothesis that there were significant changes in EEG patterns prior to seizure onsets. In this paper, however, we focus on the classification of EEG recordings from epileptic and nonepileptic patients, which are much more challenging than what we have studied in the past. We make an attempt to test the hypothesis that epileptic and nonepileptic EEG recordings during a nonseizure episode can be classified. It is even more challenging when the training has to be done across multiple patients. In addition, this study presents the new NSVM as a multidimensional time-series classification algorithm, which has not been introduced elsewhere.

D. Advances in Classification

Feature selection and classification are supervised learning, which constructs a predictive function or model from training data. Generally, classification deals with a set of desired input–output pairs by trying to find a global mapping from the collected inputs to outputs to the highest possible extent and then making predictions of future outputs. A decision tree is an algorithm that creates a mapping model to data instances based on some feature values. In a decision tree, nodes represent classification features and branches represent conjunctions of features that lead to those classifications. Starting at the root node, the data instances are sorted based on their feature values. The most well-know algorithm to generate decision trees is an algorithm called C4.5 [20]. C4.5 builds decision trees from a set of training data by using the concept of Shannon entropy [21], which is a measure of the uncertainty associated with a random variable. Ruggieri [22] provided an improved version of C4.5, called EC4.5, which was claimed to be able to compute the same decision trees as C4.5 with a performance gain of up to five times. Yildiz and Dikmen [23] presented three parallel C4.5 algorithms which are applicable to large data sets. Baik and Bala [24] presented a distributed version of decision trees. SVM is a widely used technique for data classification and regression [25]. The key concept of SVM is its projection of input data instances into a higher dimensional space and division of the space with a continuous separation hyperplane while iteratively minimizing the distance of misclassified data instances from the hyperplane. In other words, SVM generally tries to construct a hyperplane that minimizes the upper bound on the out-of-sample error. There have been many variations of SVM models. In practice, most data sets are not perfectly separable. For this reason, one should try to approximate the goal of maximizing margin by minimizing an average sum of violations. This leads to the development of robust linear programming formulation [26]. A number of linear programming formulations for SVMs have also been used to explore the properties of the structure of the optimization problem and solve large-scale problems [27], [28]. The SVM technique proposed in [28] was also demonstrated to be applicable to the generation of complex space partitions similar to those obtained by C4.5 [20] and CART [29]. SVMs have been applied to many real-life problems including handwritten digit recognition [30], object recognition [31], speaker identification [32], face detection in images [33], and text categorization [34]. SVM can also be extended to multiclass problems [30], [35], [36]. Kernel transformation, also known as kernel trick, is one of the most successful approaches applied to SVM. It uses the idea that once a data set is transformed into a high-dimensional space, each data instance can be classified by a separating plane if the new dimension is sufficiently high enough [37]. A good separation is achieved by a hyperplane with the largest distance from the neighboring data points of both classes. This concept is very intuitive because, in general, the larger the margin, the better the generalization error of the classifier. The most simple kernel function is the linear kernel $k(x, y) = x \cdot y$. The decision function takes the formula $f(x) = wx + b$. In time-series prediction, the linear kernel can be interpreted as a statistical autoregressive model of

the order $k(AR[k])$. 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 the RBF kernel can be interpreted as their Euclidian distance. Recently, there have been a number of research studies proposing the use of kernel functions for single time-series transformation such as speech recognition [38]–[42]. To the best of our knowledge, a formal time-series classification study based on data mining was introduced in the late 1990s [43], [44]. A new type of self-organizing neural network was developed to classify control-chart time-series data [43]. A new representation of time series for faster classification was proposed [44]. Subsequently, the feature-extraction technique was introduced to classify time-series data [45]. More sophisticated classification techniques such as classification trees and SVM were employed with some degree of success [46], [47]. In later studies [48], [49], the research focus was shifted to time-series motif discovery where new algorithms were developed to find an efficient and effective discrete approximation of the time series. The dynamic-time-warping measure has been applied and improved to efficiently classify time series [50]. In a recent study [51], a disk-aware algorithm was employed to find exact time-series motifs in large-scale databases. Most recently, an efficient online classification tool for time-series data based on support vector regression (SVR) was developed [52]. Whenever a sample is added to or removed from the training set, this technique is capable of updating a trained SVR model efficiently without retraining the entire training data.

Although there has been an expanded body of work in time-series classification, most studies only deal with single time-series data (univariate) and very few methods are applicable to multidimensional time series (multivariate) like EEG recordings. Similarly, most studies on EEG analysis in the literature focus on univariate analysis of the recordings. The NSVM and SFM approaches proposed in this study are designed to work with multidimensional time-series data. The NSVM approach captures the interactions between different pairs of time series while the SFM approach uses baseline patterns and fuse the similarities of individual series into a single framework. In addition, the SFM approach can be viewed as an ensemble classification version of the nearest neighbor approach for time series. The feature-extraction step of the NSVM approach can be viewed as a new way to generate features that are more interpretable to the end users (e.g., physicians).

III. Classification Techniques for Multichannel EEG Signals

We propose two new classification techniques for multidimensional time-series data that will be used to identify epileptic and nonepileptic patients from EEG data samples. Each EEG sample is represented by a multidimensional time series, shown in Fig. 1, where each trace represents a time series of an electrode. The key idea of both techniques is to integrate both spatial and temporal features of EEG data into the classification models.

A. SFM

The key idea of SFM is the integration of accuracy optimization into feature selection and the nearest neighbor classification in the training phase. The procedure of SFM is described by the following steps.

1) Step 1—Apply the Nearest Neighbor Rule: The nearest neighbor rule is a very intuitive classification method, which assigns an unlabeled sample to the class whose baseline samples are the closest. During the training phase, we have two groups of baseline (labeled) EEG data samples, epileptic and nonepileptic. Since each EEG data sample is in the form of multidimensional time-series data, we employ the nearest neighbor rule based on time-series similarity measures to quantify the closeness between data samples. Generally, time-series similarity measures deal with a single time-series profile. Here, we apply an ensemble classification concept to modify the nearest neighbor approach for multidimensional time series [12]. Given multiple decisions from multiple features (electrodes), we employ two commonly used schemes, *distance averaging* and *majority voting*, to combine these decisions in classifying an unlabeled sample. In the distance-averaging scheme, for *every feature*, each class gets a score equal to the statistical distance from each of its training samples to all other baseline samples of the same class. The overall score of each class is equal to the summation of the scores for all features (electrodes). The sample is then classified to the class with the lowest overall score. In the majority-voting scheme, for *every feature*, a class (category) gets one vote if the nearest neighbor rule classifies the training sample to that class. The sample is, in turn, classified to the class with the maximum number of votes, i.e., majority of features/electrodes.

Here, we employ two commonly used time-series similarity measures, *Euclidean distance* and *T-Statistical distance (T-Statistics)*. Euclidean distance is the most commonly used similarity measure. It measures the degree of similarity in terms of the amplitude of the data. The Euclidean distance between EEG samples u and v of length t at electrode j is defined as $ED^j_{uv} = (\sum_{i=1}^{t}(u_i^j - v_i^j)^2)/t$. T-statistical distance is a measure of statistical distance between two time series derived from the t-test, which is frequently used to determine if two time series differ from each other in a significant way under the assumptions that the paired differences are independent and identically normally distributed. The t-index can be seen as a ratio of the difference between the two means or averages and a measure of the variability or dispersion of the scores. The t-index between EEG samples u and v of length t at electrode j is defined as $T^j_{uv} = (\sum_{i=1}^{t}|u_i^j - v_i^j|/\sqrt{t}\sigma_{|u^j - v^j|})$, where $\sigma_{|u^j - v^j|}$ is the sample standard deviation of the absolute difference between EEG time series u and v estimated over a window with length t.

In the training phase, since we already know the true class (label) for each of the training samples, we use the nearest neighbor rule to evaluate the classification accuracy of every electrode on every training sample. The SFM optimization models are then formulated to incorporate all the classification decisions made by all the electrodes and select the best subset of electrodes that maximizes the classification accuracy. For

the distance-averaging scheme, the input information of SFM includes $n \times m$ *intraclass-distance* matrix (D) and $n \times m$ *interclass-distance* matrix $(\bar{D})$. The entry of intraclass matrix d_{ij} is the average statistical distance between training sample i and all other training samples from the same class at electrode j. The entry of interclass matrix $\bar{d}_{ij}$ is the average statistical distance between training sample i and all other training samples from different classes at electrode j. For the majority-voting scheme, the SFM input is an accuracy $n \times m$ matrix (A), where n is the number of training samples and m is the number of features. The entry $a_{ij} = 1$ indicates that the nearest neighbor rule correctly classifies training sample i at feature j and is zero otherwise. The time complexity of the matrix-generation procedure for both schemes is $O(n^2 m\tau)$, where n^2 is required for a pairwise comparison of all data, m is required for all electrodes, and τ is required for the calculation of time-series similarity measures. Here, $\tau = O(t)$ for Euclidean distance and $\tau = O(t^2)$ for T-Statistical distance.

2) Step 2—Optimization Models of SFM: The SFM optimization model is formulated to select a group of features (electrodes) that maximizes classification accuracy based on the nearest neighbor rule. To formally formulate the SFM optimization problem into a mathematical programming model, we define the following sets and decision variables. Denote $i \in I$ as a set of training samples where $|I| = n$ and $j \in J$ as a set of features where $|J| = m$. We let $x_j \in \{0, 1\}$ be a decision variable indicating if feature j is selected by SFM and $y_i \in \{0, 1\}$ be a decision variable indicating if training sample i is correctly classified.

Averaging SFM (A-SFM): The input of A-SFM are the intra- and interclass distance matrices D and $\bar{D}$ generated in Step 1. The objective function in (1) is to maximize the total number of correctly classified samples. There are two sets of constraints in (2) and (3) to ensure that the training samples are classified based on the distance-averaging nearest neighbor rule. There is a set of logical constraints in (4) to ensure that at least one feature (electrode) is selected in the distance-averaging nearest neighbor rule. The mixed-integer program (MIP) of A-SFM is given by

$$\max \sum_{i=1}^{n} y_i \tag{1}$$

$$\text{s.t.} \quad \sum_{j=1}^{m} \bar{d}_{ij} x_j - \sum_{j=1}^{m} d_{ij} x_j \le M_{1i} y_i, \qquad \forall i \in I \tag{2}$$

$$\sum_{j=1}^{m} d_{ij} x_j - \sum_{j=1}^{m} \bar{d}_{ij} x_j \le M_{2i}(1 - y_i), \qquad \forall i \in I \tag{3}$$

$$\sum_{j=1}^{m} x_j \ge 1 \tag{4}$$

$$x \in \{0, 1\}^m \quad y \in \{0, 1\}^n \tag{5}$$

where d_{ij} is the average statistical distance between training sample i and all other training samples from the same class at electrode j (*intraclass distance*), $\bar{d}_{ij}$ is the average statistical distance between training sample i and all other training samples from different classes at electrode j (*interclass distance*), $M_{1i} = \sum_{j=1}^{m} \bar{d}_{ij}$, and $M_{2i} = \sum_{j=1}^{m} d_{ij}$.

Voting SFM (V-SFM): The input of V-SFM is the accuracy matrix A generated in Step 1. The objective function of V-SFM in (6) is the same as that of A-SFM. There are two sets of constraints in (7) and (8) to ensure that the training samples are classified based on the voting scheme. There is a set of logical constraints in (9) to ensure that at least one feature is used in the voting nearest neighbor rule. The MIP of V-SFM is given by

$$\max \sum_{i=1}^{n} y_i \tag{6}$$

$$\text{s.t.} \quad \sum_{j=1}^{m} a_{ij} x_j - \sum_{j=1}^{m} \frac{x_j}{2} \le M y_i, \qquad \forall i \in I \tag{7}$$

$$\sum_{j=1}^{m} \frac{x_j}{2} - \sum_{j=1}^{m} a_{ij} x_j + \epsilon \le M(1 - y_i) \qquad \forall i \in I \tag{8}$$

$$\sum_{j=1}^{m} x_j \ge 1 \tag{9}$$

$$x \in \{0, 1\}^m \quad y \in \{0, 1\}^n \tag{10}$$

where $a_{ij} = 1$ if the nearest neighbor rule correctly classified training sample i at feature j and is zero otherwise, n is the total number of training samples, m is the total number of features, $M = m/2$, and $0 < \epsilon < 1/2$ is used to break ties during voting.

In the training phase (Steps 1 and 2), the SFM optimization models of A-SFM and V-SFM are individually solved using an off-the-shelf CPLEX optimization solver. It is very important to note that these optimization models are very compact. The space complexity grows linearly with the number of training samples and the number of features (electrodes), specifically $O(n + m)$. The majority of computational efforts will be in Step 1. This depends solely on the number of features and the length (number of data points) of EEG epochs. After the SFM models are solved, a group of optimally selected features (electrodes) that maximize the classification accuracy will be obtained and used in Step 3.

3) Step 3—Using SFM to Classify Unlabeled EEG Sample: In the testing phase, the classification will be done based on the features (electrodes) selected in the training phase. A-SFM classifies an unlabeled EEG sample to the class whose baseline training EEG data are the nearest (closest) based on the average distance of selected electrodes. Similarly, V-SFM classifies an unlabeled EEG sample to the class with the highest votes counted only from selected electrodes.

B. NSVM

NSVM employs a new network modeling technique to incorporate SVM with spatiotemporal EEG analysis by representing an EEG sample as a "*Brain Graph*" or "*Brain Network*." Essentially, NSVM maps each multichannel EEG sample into a network representation and applies the SVM classification to the mapped data. Fig. 2 shows a hypothetical schematic

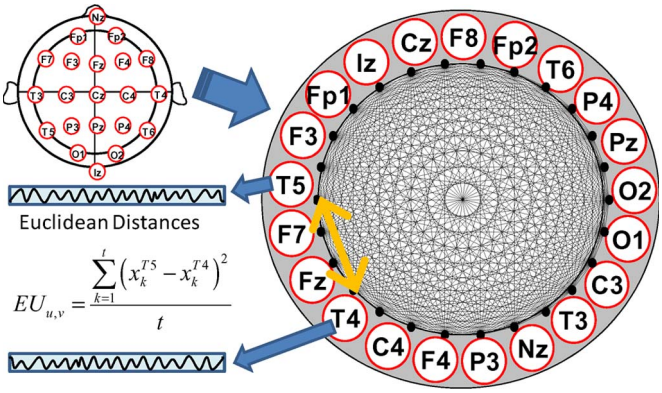

Fig. 2. Example of brain network models of all pairwise edges of scalp EEG data, where the similarity of EEG time series between two electrodes $T4$ and $T5$ is measured by Euclidean distance.

view of the network-based modeling procedure of NSVM. The procedure of NSVM is described in the following steps.

1) Step 1—Network Modeling (Feature Extraction): In the first step, we extract important features of EEG recordings by using the new brain networking model, which can be formulated in graph-theoretic terms as follows. Let G be an undirected graph with vertices $V_1, \ldots, V_m$, where V_i represents electrode i. There is an edge (link) with the weight w_{ij} for every pair of nodes V_i and V_j corresponding to the similarity of the EEG signals between the two electrodes. In this paper, we employ Euclidean distance as a measure of similarity/distance between two electrodes. The similarity or distance between each electrode pair is hypothesized to be the degree of changes in the neuronal synapses in the local circuitry under the recording electrodes.

With this networking model, the attributes of SVM inputs are the pairwise distances between time-series profiles of electrode pairs rather than the time stamps of time-series profiles. Not only can NSVM capture the global interaction among different electrode sites, but it can also reduce the dimensionality of the EEG data. For example, given m electrodes, each with t time stamps, NSVM decreases the number of attributes (SVM features) by $2(m-1)/t$ times. To be more precise, let n be the total number of EEG data samples. The dimensionality can be reduced from $A \in \Re^{n\times m\times t}$ to $\bar{A} \in \Re^{n\times(m(m-1)/2)}$. In Fig. 2, all pairwise edges of a brain graph are illustrated, where each edge weight is calculated by measuring the Euclidean distance between the time-series profiles of two electrodes. In fact, the network modeling concept can be viewed as a kernel function that maps the EEG data from the input space X into another feature space $\mathcal{X}$ by a function $\Phi : X \to \mathcal{X}$. Such a kernel function k calculates the inner product in the feature space $k(u,v) = \Phi(u) \cdot \Phi(v)$ of EEG time series u and v. The kernel function is a similarity (distance) measure in the input space [53]. The similarity between the samples x and y can be shown as the kernel function $k(x,y)$ as $d^2(u,v) = (\Phi(u) - \Phi(v))^2 = k(u,u) - 2k(u,v) + k(v,v)$.

2) Step 2—Modeling SVM: After a network representation of EEG samples is obtained, we use the edge weights as input features of SVM. The procedure of SVM is formally defined as follows. Let all the data samples be represented as n-dimensional vectors (or points in the n-dimensional space). SVM finds a hyperplane that separates all the vectors (points) in set A (epileptic) from the vectors in set B (nonepileptic) in an n-dimensional space. If a hyperplane is defined by the standard expression $x^{\mathrm{T}}\omega = \gamma$, where $\omega = (\omega_1, \ldots, \omega_n)^{\mathrm{T}}$ is an n-dimensional vector of real numbers and γ is a scalar, then this plane will separate all the elements A from B. Thus, discrimination rules can be formulated as an optimization problem to determine vectors ω and γ such that the separating hyperplane defines two open half spaces, which contain most of the data points in A and B, respectively. However, in practice, it is usually not possible to perfectly separate two sets of elements by a plane. For this reason, it is intuitive to minimize the average measure of misclassifications. The violations of these constraints are modeled by introducing nonnegative variables x and y. The mathematical model for SVM that minimizes the total average measure of misclassification errors is given by

$$\min_{\omega,\gamma,x,y} \quad \frac{1}{m}\sum_{i=1}^{m} x_i + \frac{1}{k}\sum_{j=1}^{k} y_j \tag{11}$$

$$\text{s.t. } A\omega + x \geq e\gamma + e \tag{12}$$

$$B\omega - y \leq e\gamma - e \tag{13}$$

$$x \geq 0 \quad y \geq 0 \quad \omega, \gamma \in \mathbb{R}^n. \tag{14}$$

As one can see, this is a linear programming problem. The decision variables here are the geometrical parameters of the separating plane ω and γ, as well as the variables representing misclassification error x and y.

IV. Materials and Methods

A. Data Acquisition and Sampling

In this paper, two EEG data sets were acquired. The first EEG data set was acquired from short-term EEG recordings of ten patients, five epileptic and five nonepileptic. The second EEG data set was acquired from ten patients; however, only five of those patients were known to be epileptic while the other five patients were not known to be epileptic or nonepileptic. In practice, it was hard to find a clean data set in which patients were definitely labeled as nonepileptic. Thus, in this study, we used the second data set as a validation set and considered only the recordings from the five epileptic patients. The recordings from ambiguous patients were discarded before the ground truth of the patient labels were not known.

Tables I and II present the recording characteristics of the first and second EEG data sets, respectively. All EEG recordings were obtained from two separate hospitals in New Jersey during the patients' scheduled EEG checks as part of a routine epilepsy and brain-disorder diagnosis. Each recording used a total of 19–22 scalp electrodes, placed according to the extended International 10–20 System. In the first data set, the 18-channel bipolar montage was used to acquire signals for our analysis. In the second data set, different bipolar montages were used to acquire signals from different patients; ultimately, there were only 14 consistent bipolar electrodes, which were used in this study. The EEG recordings were 13–45 min in duration, digitized at 250-Hz sampling rate (for the first data

TABLE I
CHARACTERISTICS OF EEG DATA SET I

Patient	EEG Recordings	
	Duration	Data Dimension
Epileptic 1	28.71 mins	18 (channels) × 430,650 (data points)
Epileptic 2	29.87 mins	18 (channels) × 448,025 (data points)
Epileptic 3	20.89 mins	18 (channels) × 313,375 (data points)
Epileptic 4	30.19 mins	18 (channels) × 452,875 (data points)
Epileptic 5	29.94 mins	18 (channels) × 449,150 (data points)
Non-Epileptic 1	31.50 mins	18 (channels) × 472,464 (data points)
Non-Epileptic 2	32.90 mins	18 (channels) × 493,464 (data points)
Non-Epileptic 3	28.06 mins	18 (channels) × 420,964 (data points)
Non-Epileptic 4	21.90 mins	18 (channels) × 328,464 (data points)
Non-Epileptic 5	33.33 mins	18 (channels) × 499,464 (data points)

TABLE II
CHARACTERISTICS OF EEG DATA SET II

Patient	EEG Recordings	
	Duration	Data Dimension
Epileptic 1	18.38 mins	14 (channels) × 220,600 (data points)
Epileptic 2	28.57 mins	14 (channels) × 342,800 (data points)
Epileptic 3	13.43 mins	14 (channels) × 161,200 (data points)
Epileptic 4	29.65 mins	14 (channels) × 355,800 (data points)
Epileptic 5	21.63 mins	14 (channels) × 259,600 (data points)

set) and 200-Hz sampling rate (for the second data set). All recordings in the first data set were viewed by two independent epileptologist electroencephalographers to determine if the patient had epilepsy. The recordings in the second data set were viewed by an epileptologist electroencephalographer to identify obvious epileptic patients and ambiguous patients. The first data set was used to assess the performance of the proposed classification techniques that predict whether a patient is epileptic or nonepileptic. It is very important to note that only the recordings from epileptic patients in the second data set were used as hold-out samples only to validate the classification models trained by using the first data set.

Based on the noncentral F-distribution, with Type I error $\alpha = 0.05$, a minimum of eight subjects in each group were required to detect a difference of two standard deviations with a test power of at least 90% (i.e., Type II error < 0.1). Since we had only five subjects in each group, in order to justify the prediction power, we used a Monte Carlo sampling method to randomly and uniformly sample five l-minute epochs of EEG recordings from each subject. In this paper, we use $l = 1, 2, 3,$ and 4 min to test the prediction power of different sizes of EEG epochs. Each sampled epoch can be represented by an $n \times m$ matrix, where $n = 18$ electrodes and $m = l \times 60 \times 250$ data points for the first data set and $n = 14$ electrodes and $m = l \times 60 \times 200$ data points for the second data set. There were two main reasons why we varied the window size between 1 and 4 min. First, in clinical settings, obtaining a second opinion in a timely fashion can be very useful for clinicians. Second, the epoch has to be large enough to capture significant patterns (e.g., sharp waves, transients, and spindles) yet small enough to run our classification algorithms in a timely fashion.

B. Assessment and Validation

This phase involved the design of experiments to assess and validate the efficacy and robustness of the proposed classification techniques.

Training and Testing: In order to reduce the bias of training and testing data, in previous work, cross-validation techniques have been extensively used as a method to estimate the generalization error based on "resampling," by assessing how well classification techniques have been learned in the training phase and how well they are going to perform on future as-yet-unseen data in the testing phase. The n-fold cross-validation is commonly used to divide the data into n subsets of (approximately) equal size. Then, the classification techniques will be trained and tested n times, in which one of the subsets from training is left out each time and tested on the omitted subset [54]. However, n-fold cross-validation is not appropriate in this study because the various folds of the training and the fold of the testing may include data samples that were drawn from the same patients. This situation would increase the likelihood of bias and, thereby, might artificially increase the classification accuracy. Thus, to avoid such a situation, we employed a leave-one-patient-out cross-validation methodology in order to avoid the potential bias of having EEG samples from the same patients in both the training and testing data. Essentially, the proposed cross-validation methodology ensures that all five EEG samples drawn from the same patient are included in the same fold. In turn, five EEG samples corresponding to an individual patient are left out of the training data, consisting of the data from the remaining nine patients, and used as testing data. This is repeated for all ten patients, so that, ultimately, there are ten *leave-one-patient-out cross-validation runs* for each data set. This process ensures that each patient is left out exactly once, removing the potential for bias.

Performance Evaluation: We evaluated the performance of classification techniques in terms of sensitivity (1 - Type II error), specificity (1 - Type I error), and accuracy. 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. Accuracy measures the overall fraction of samples that are correctly classified. In this paper, in the first data set, we labeled the EEG samples from epileptic patients as positive and those from nonepileptic patients as negative. In the second data set, we labeled the samples from epileptic patients as positive and those from ambiguous patients as negative. For the testing samples, we considered four subsets of classification results: 1) true positives (*TP*) denoting correct classifications of positive cases; 2) true negatives (*TN*) denoting correct classifications of negative cases; 3) false positives (*FP*) denoting incorrect classifications of negative cases as positive; and 4) false negatives (*FN*) denoting incorrect classifications of positive cases as negative. We measured the performance of classification techniques as follows: $Sensitivity = (TP/(TP + FN)) = 1 - Type\ \ II\ \ error$; $Specificity = (TN/(FP + TN)) = 1 - Type\ \ I\ \ error$; and $Accuracy = (TP + TN/(TP + TN + FP + FN))$.

Hold-out Samples: Only recordings from epileptic patients in the second data set were used as a hold-out data set. We trained all classification techniques using recordings from all ten patients in the first data set. Once they were trained, we tested their accuracies (sensitivities) in detecting recordings

from epileptic patients. Note that the sensitivity is extremely important because, in medical diagnosis, a false negative is much more costly than a false positive. In other words, it is better to overdiagnose the patients than to not pick up if the patient indeed has the disease (i.e., epilepsy in this case). It is important to note that the sampling rates of EEG recordings in the two data sets were different. We employed a spline function to downsample the EEG recordings in both data sets to 100 and 200 Hz.

C. Computational Implementation

All optimization problems associated with SFM were modeled in MATLAB through a callable GAMS library and solved using ILOG CPLEX version 10.0 with the default setting. All of the SFM experiments were implemented and performed on an Intel Xeon 3.0-GHz workstation with 3 GB of memory running Windows XP. All calculations and algorithms were implemented and run on MATLAB version R2007a. The computational time required to solve the SFM model was, on average, less than 5 min, and the time required to process an unknown EEG epoch to classify it to a patient group was, on average, less than 3 min. All the calculations associated with SVM were done using the WEKA Workbench [55]. In order to employ the WEKA algorithms, we converted the EEG data into the Attribute Relation File Format, which is the WEKA default format, using a Java program. Due to the very large dimensionality of the EEG data, the WEKA GUI could not be employed on our local workstation. The WEKA algorithms were executed on a supercomputer named Cobalt at the National Center for Supercomputing Applications in Champaign–Urbana, IL. Cobalt contains 96 GB of globally accessible memory and therefore provided substantial resources to conduct the experiments in this study.

V. Computational Results

We employed the leave-one-patient-out cross-validation to assess the classification performance of A-SFM, V-SFM, and NSVM on both data sets I and II independently. Specifically, for each data set, ten classification iterations were performed, in each of which, all five EEG samples drawn from one patient were used as the testing set while the rest of the EEG samples were used as the training set. In each iteration, sensitivities and specificities were measured for different EEG epoch sizes. The sensitivities and specificities were then averaged across ten iterations. In order to show the superiority of the proposed techniques, we implemented various different classification algorithms in order to discover which model most accurately predicted an individual patient's diagnosis. In the initial steps of the experiment, a few different algorithms were selected in order to explore the space of algorithm performance and eventually narrow the selection down to the most effective algorithms. The algorithms selected for initial experimentation were J48 (a Decision-Tree Learner), SVM, OneR (Holte's OneR), Naive Bayes classifier, Instance-Based Learning, the Decision Stump, and the Hoeffding tree algorithm. We note that the Hoeffding tree algorithm is a time-series classification, which appears to be an appropriate choice of EEG classification. Because the Hoeffding tree algorithm works by assuming a stream of data consisting of many records, the leave-one-patient-out cross-validation is not applicable in this case. In our experiments, we performed bootstrapping to oversample the EEG epochs to obtain a stream of 1 million instances. This resampled stream was used as both the training and testing data sets while a different seed was used for random selection from the stream during testing to ensure that the Hoeffding tree algorithm used a different set of data during training and testing. Given the various results, the most effective algorithms selected for further experimentation in this study were J48 and linear and quadratic SVMs. J48 is a decision-tree algorithm in which nonterminal nodes indicate tests on one or more attributes using an attribute selected as the best differentiator and terminal nodes indicate classifications. Note that the implementation of standard linear and quadratic SVMs was very similar to the one of NSVMs. The main difference is the input data, as for the nonnetworking implementation, we simply concatenated the time-series data from all electrodes for each sample as a large vector.

A. Classification Results of Data Set I

The classification results of A-SFM, V-SFM, SVM, NSVM, and Decision Tree (J48) on the testing data of data set I are summarized in Table III. We note that, in every experiment, all of the classification techniques yielded 100% training accuracy. From the table, the various consistencies of classification results across different epoch sizes are observed. For J48 (decision-tree algorithm), the classification performance appears to decrease as the epoch size increases. This might be due to the number of input features that grows with the epoch size, which might lead to data overfitting. Both linear and quadratic SVMs consistently achieved very high sensitivity and specificity. NSVM also consistently provided very high competitive sensitivity and specificity, except in the 4-min epoch case. It is also important to note that the size of the input vectors for NSVM is much smaller than the size of the vectors used as input for standard SVM. Specifically, as the features of NSVM are the pairwise distances between electrode pairs, each input data sample of NSVM contains only $18 \times 17 = 306$ features, whereas each input data sample of SVM contains $18 \times 45\,000 = 810\,000$ features. Most importantly, it only required a fraction of time to solve (compared with standard SVM) and did not require the use of large computing clusters. As for SFM approaches, A-SFM consistently outperformed V-SFM. We speculated that V-SFM might not perform well when there were not many features in the model. Thus, in order to get consensus voting, a larger number of features may provide more accurate reliable classification. These results indicate that the optimization component in SFMs is very effective at capturing the dynamic interactions of the EEG's spatial components, i.e., electrode interactions of epileptic and nonepileptic patients. These results also validate the need for feature selection through spatial component optimization in multidimensional time-series classification. All in all, the proposed SFM and NSVM approaches were capable of separating and identifying EEG samples collected

TABLE III
PERFORMANCE CHARACTERISTICS (TESTING RESULTS) OF DECISION TREE (J48), SVM, AND SFM WITH DIFFERENT PARAMETER SETTINGS TESTED ON EEG DATA SET I (EPILEPTIC VERSUS NONEPILEPTIC)

Classification Techniques	1-min Epoch		2-min Epoch		3-min Epoch		4-min Epoch	
	Sensitivity	Specificity	Sensitivity	Specificity	Sensitivity	Specificity	Sensitivity	Specificity
Decision Tree (J48)	80%	92%	72%	80%	72%	80%	68%	84%
SVM (Linear)	100%	92%	100%	96%	100%	96%	100%	100%
(Quadratic)	100%	92%	100%	92%	100%	96%	100%	92%
NSVM (Linear)	92%	100%	92%	100%	96%	100%	92%	88%
(Quadratic)	92%	100%	92%	100%	96%	100%	92%	84%
A-SFM (Euclidean)	92%	88%	100%	96%	96%	100%	100%	88%
(T-Statistics)	96%	84%	100%	84%	100%	88%	100%	92%
V-SFM (Euclidean)	92%	88%	96%	88%	96%	84%	96%	80%
(T-Statistics)	92%	88%	92%	84%	92%	88%	96%	88%

TABLE IV
ACCURACIES OF J48, SVM, AND SFM IN CLASSIFYING INDIVIDUAL PATIENTS (TESTING RESULTS) USING 3-min EEG EPOCHS OF THE FIRST DATA SET

Patients	Decision Tree	SVM		NSVM		A-SFM		V-SFM	
	J48	Linear	Quadratic	Linear	Quadratic	Euclidean	T-Statistics	Euclidean	T-Statistics
Epileptic 1	100%	100%	100%	80%	80%	100%	100%	100%	100%
Epileptic 2	100%	100%	100%	100%	100%	100%	40%	40%	40%
Epileptic 3	60%	100%	100%	100%	100%	100%	100%	80%	100%
Epileptic 4	60%	100%	100%	100%	100%	100%	100%	100%	100%
Epileptic 5	40%	100%	100%	100%	100%	100%	100%	100%	100%
Non-Epileptic 1	60%	100%	100%	100%	100%	100%	100%	100%	100%
Non-Epileptic 2	60%	80%	80%	100%	100%	100%	100%	100%	100%
Non-Epileptic 3	80%	100%	100%	100%	100%	80%	100%	80%	60%
Non-Epileptic 4	100%	100%	100%	100%	100%	100%	100%	100%	100%
Non-Epileptic 5	100%	100%	100%	100%	100%	100%	100%	100%	100%

from epileptic and nonepileptic patients. These approaches required much less computational resources than the decision-tree and standard SVM approaches while providing very competitive results.

B. Classification Results of Individual Patients

Table IV summarizes the accuracy of EEG classification using J48, SVM, NSVM, and SFM based on individual patients from the experiments in Table III. Note that, for individual patients, we classify five data samples; therefore, the accuracies shown in the table are in a multiple of 20%. These results may be useful when estimating the performance of a possible collective classification when making a diagnosis decision. For example, one can apply a majority rule (three out of five samples) to make a diagnosis decision: epileptic or nonepileptic. The results in Table IV may be, in turn, considered as classifier's reliability. In regard to classifier's reliability, we observed that, in most cases, all approaches provide very good classification accuracies for all patients in the first data set. The classification results of J48 appear to be the worst among the different classification techniques in terms of accuracy and reliability. The classification results obtained by standard SVM, NSVM, and A-SFM approaches were consistently accurate with only a few exceptions where 100% accuracy was not obtained.

C. Classification Results of Hold-Out Data Set II

In order to validate that our EEG classification framework can be generalized to data from different acquisition machines, we trained the classifiers using the entire data set I and tested the trained classifiers using data set II as a hold-out data set. Note that due to the difference in sampling rates, we employed a spline function to downsample the EEG recordings from both data sets to 100 and 200 Hz. We also note that although the standard 10–20 montage system is commonly used, the output EEG signals may be extracted from different bipolar settings. In our case, we were able to match the EEG recordings from 12 electrodes in common between data sets I and II. After we obtained both training and testing data sets of the same size from the common electrodes, we implemented all classification techniques as stated in an earlier experiment. The training accuracies of all classification techniques were near perfect, either 100% or 96%, in most test instances. The testing results of using the hold-out data set II are given in Table V. The table reports the classification accuracies of correctly predicting the EEG epochs from epileptic patients. The results when using 100- and 200-Hz spline functions are very comparable while the size of EEG epochs slightly affected the classification accuracies. Nonetheless, the results from 3- and 4-min epochs are quite stable, which, in turn, implies that they might be an appropriate epoch size for this study. From the table, it is observed that NSVM (quadratic) yielded 100% classification accuracy for all cases, regardless of the spline sampling rate and epoch size. In addition, the classification results of NSVM outperformed those of all other classification techniques tested here. In particular, SVM techniques, both quadratic and linear kernels, failed to classify all EEG samples from the epileptic patients in data set II. The results suggest that NSVM may be an appropriate approach for this problem as it is fast, accurate, scalable, and can be parallelized. More importantly, these results may lead to a greater understanding of epileptogenesis. The features that are useful in classifying epileptic and nonepileptic patients may be the interactions among different brain areas.

TABLE V
ACCURACIES OF DECISION TREE (J48), SVM, AND SFM WITH DIFFERENT PARAMETER SETTINGS TESTED ON HOLD-OUT DATA SET II (ONLY EPILEPTIC PATIENTS)

Classification Techniques	100 Hz Spline				200 Hz Spline			
	1-min	2-min	3-min	4-min	1-min	2-min	3-min	4-min
Decision Tree (J48)	0%	0%	84%	96%	0%	0%	0%	92%
SVM (Linear)	0%	0%	0%	0%	0%	0%	0%	0%
(Quadratic)	0%	0%	0%	0%	0%	0%	0%	0%
NSVM (Linear)	88%	88%	88%	88%	0%	100%	100%	100%
(Quadratic)	100%	100%	100%	100%	100%	100%	100%	100%
A-SFM (Euclidean)	0%	0%	0%	0%	0%	0%	40%	0%
(T-Statistics)	0%	0%	0%	0%	0%	0%	0%	0%
V-SFM (Euclidean)	100%	100%	100%	100%	80%	100%	100%	100%
(T-Statistics)	0%	0%	0%	40%	0%	0%	0%	40%

VI. CONCLUSION AND DISCUSSIONS

A quick and accurate epilepsy-screening tool will enormously reduce associated healthcare costs and improve the medical diagnosis, treatment, and prognosis of patients. In this paper, we studied the diagnosis challenges of epilepsy, a brain disease characterized by recurrent seizures. In epilepsy diagnosis, discovering why seizures occur can be difficult, because there are so many possible causes. Seizures often happen infrequently and unpredictably, making it difficult for neurologists to assess them with neurological testing. In addition, some people have more than one condition that can cause seizures or seizurelike episodes. In the case of a coma, the differentiation of epileptic seizures from seizures with other causes with similar EEG morphological patterns is very challenging, yet would be very helpful to physicians. Unfortunately, in up to 70% of recorded cases, the cause of a person's seizures remains unknown [3]. Thus, there is a definite need for a new decision-aided tool to help physicians recognize abnormal hidden patterns from brain activity that signify epilepsy. Such a tool could improve both the diagnosis and treatment of patients, since various types of seizures respond best to specific treatments.

To be able to reliably recognize if a patient has epilepsy, we developed two new time-series classification techniques, namely, SFM and NSVM, to classify EEG data from epileptic and nonepileptic patients. The pattern-based SFM approach is an optimization model that maximizes classification accuracy by selecting electrodes (features) that correctly classify EEG samples in the training phase. Conceptually, the selected electrodes in the training phase have strong class separability through time-series similarity measures and have proven useful in the testing phase. There are two variations of SFM modeling: averaging and voting. The objective of A-SFM is to maximize the number of training data samples whose intraclass distances are smaller than their interclass distances. The classification rule is based on the concept of the nearest neighbor rule. V-SFM can be viewed as an ensemble classification method, in which each classification rule is derived from the nearest neighbor classification for each electrode. The main concern of SFM is its scalability. Although SFM's space complexity grows linearly with the number of training samples and the number of features, it is still ideal to collect and use as many baseline samples as possible as the accuracy of SFM relies mainly on the baseline samples. If EEG epochs were carefully sampled such that they included epileptic patterns such as sharp waves and spikes, SFM should perform much better than other approaches as it uses the morphology of EEG recordings as features. In addition, the size of sampled EEG epochs also plays a very important role in computational time requirement. Because the SFM modeling requires a preprocessing step that uses the nearest neighbor rule to measure the intra- and interclass distances for each feature, the size of EEG epochs were limited to less than 5 min. Nevertheless, the proposed SFM and NSVM approaches still scale much better than the decision-tree and standard SVM approaches. In our empirical study, we were not able to run those approaches in our computing server but rather had to use a supercomputer based at the University of Illinois at Urbana–Champaign. Among all the approaches investigated here, NSVM, which employs a new network-based model and integrates it with SVM modeling, is the most scalable approach. This is because it transforms a multidimensional EEG time series into a single matrix as input features of SVM.

The classification results in this study are quite promising. In a carefully selected data set (data set I), whose ground truth of the data label (epileptic versus nonepileptic) is known, all the classification techniques except decision tree successfully classified all ten patients. However, in the hold-out data set (data set II), only NSVM and V-SFM were able to accurately predict the group of EEG samples while other methods failed to make accurate predictions. In other words, decision tree, A-SFM, and SVM predicted all epileptic patients as nonepileptic patients—obtaining 0% classification accuracy. This is extremely vital in real-life medical diagnosis because the cost of false negative is far higher than the cost of false negative. The NSVM approach appears to be very appropriate as a generalized method to be applied in real life because it is independent of the sampling rate and the signal acquisition settings of the EEG machine. NSVM's features are simply the connectivity relationship among different brain regions. Thus, the EEG time series from individual electrodes are compared among each other from the same acquired recordings. On the other hand, other approaches like decision tree and traditional SVM are very sensitive to the sampling rate and the amplification of the EEG signals. Although simple scaling and/or normalizing can be applied, such preprocessing step might, in turn, remove meaningful patterns in the recordings. In addition, this observation is also confirmed by the contrasting results between V-SFM and A-SFM. V-SFM is less sensitive to noises and biases from individual electrodes due to the nature of voting,

i.e., every selected electrode equally contributes a vote of one. On the other hand, A-SFM may be biased by some selected electrodes with a higher scale of similarity distance. Although the results in this experiment appear to be conclusive, it would be ideal if we were able to obtain a hold-out data set that contains nonepileptic patients. In practice, carefully selected data sets with known labels (epileptic versus nonepileptic) are extremely hard to obtain due to the fact that it takes several months or years for physicians to ascertain whether a patient is epileptic or nonepileptic.

ACKNOWLEDGMENT

The authors would like to thank R. C. Sachdeo, M.D., D. Tikku, M.D., and B. Y. Wu, M.D., Ph.D. for providing insightful information and discussion that has led us to investigate the problem addressed in this study.

REFERENCES

[1] O. Cockerell, I. Eckle, D. Goodridge, J. Sander, and S. Shorvon, "Epilepsy in a population of 6000 re-examined: Secular trends in first attendance rates, prevalence, and prognosis," *J. Neurol., Neurosurg., Psychiatry*, vol. 58, no. 5, pp. 570–576, May 1995.

[2] W.H.O. (WHO), *Epilepsy: Historical Overview*, 2004. [Online]. Available: http://www.who.int/mediacentre/factsheets/fs168/en/

[3] Epilepsy Foundation, *Epilepsy Foundation—Not Another Moment Lost to Seizures*, 2006. [Online]. Available: http://www.epilepsyfoundation.org

[4] C. A. van Donselaar, H. Stroink, and W.-F. Arts, "How confident are we of the diagnosis of epilepsy?" *Epilepsia*, vol 47, no. S1, pp. 9–13, Oct. 2006.

[5] Y. Kakizawa, R. Shumway, and M. Taniguchi, "Discrimination and clustering for multivariate time series," *J. Amer. Stat. Assoc.*, vol. 93, no. 441, pp. 328–340, Mar. 1998.

[6] W. Hsieh, "Nonlinear multivariate and time series analysis by neural network methods," *Rev. Geophys.*, vol. 42, p. RG1 003, 2004.

[7] M. Kadous and C. Sammut, "Constructive induction for classifying time series," in *Proc. ECML*, vol. 3201. Berlin, Germany: Springer-Verlag, 2004, pp. 192–204.

[8] K. Yang, H. Yoon, and C. Shahabi, "A supervised feature subset selection technique for multivariate time series," in *Proc. Int. Workshop FSDM: Interfacing Mach. Learn. With Statist. in conjunction with SIAM Int. Conf. Data Mining (SDM)*, Newport Beach, CA, Apr. 2005.

[9] X. Song, M. Wu, C. Jermaine, and S. Ranka, "Statistical change detection for multi-dimensional data," in *Proc. 13th ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining*, 2007, pp. 667–676.

[10] D. Dou, G. Frishkoff, J. Rong, R. Frank, A. Malony, and D. Tucker, "Development of neuroelectromagnetic ontologies (NEMO): A framework for mining brain wave ontologies," in *Proc. 13th ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining*, 2007, pp. 270–279.

[11] W. Chaovalitwongse, P. Pardalos, and O. Prokoyev, "Electroencephalogram (EEG) time series classification: Applications in epilepsy," *Ann. Oper. Res.*, vol. 148, no. 1, pp. 227–250, Nov. 2006.

[12] W. Chaovalitwongse, Y. Fan, and R. Sachdeo, "On the time series k-nearest neighbor for abnormal brain activity classification," *IEEE Trans. Syst., Man, Cybern. A, Syst., Humans*, vol. 37, no. 6, pp. 1005–1016, Nov. 2007.

[13] W. Chaovalitwongse, Y. Fan, and R. Sachdeo, "Support feature machine for classification of abnormal brain activity," in *Proc. 13th ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining*, 2007, pp. 113–122.

[14] W. Chaovalitwongse, Y. Fan, and R. Sachdeo, "Novel optimization models for abnormal brain activity classification," *Oper. Res.*, vol. 56, no. 6, pp. 1450–1460, Nov./Dec. 2008.

[15] A. Babloyantz and A. Destexhe, "Low dimensional chaos in an instance of epilepsy," *Proc. Nat. Acad. Sci. U.S.A.*, vol. 83, no. 10, pp. 3513–3517, May 1986.

[16] L. Iasemidis, H. Zaveri, J. Sackellares, and W. Williams, "Phase space analysis of EEG in temporal lobe epilepsy," in *Proc. 10th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, 1988, pp. 1201–1203.

[17] N. Packard, J. Crutchfield, and J. Farmer, "Geometry from time series," *Phys. Rev. Lett.*, vol. 45, no. 9, pp. 712–716, Sep. 1980.

[18] P. Rapp, I. Zimmerman, and A. M. Albano, "Experimental studies of chaotic neural behavior: Cellular activity and electroencephalographic signals," in *Nonlinear Oscillations in Biology and Chemistry*, H. Othmer, Ed. Berlin, Germany: Springer-Verlag, 1986, pp. 175–205.

[19] F. Takens, "Detecting strange attractors in turbulence," in *Dynamical Systems and Turbulence*, vol. 898, *Lecture Notes in Mathematics*, D. Rand and L. Young, Eds. Berlin, Germany: Springer-Verlag, 1981, pp. 366–381.

[20] J. Quinlan, *C4.5: Programs for Machine Learning*. San Mateo, CA: Morgan Kaufmann, 1993.

[21] C. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, pp. 379–423, Jul. 1948.

[22] S. Ruggieri, "Efficient C4.5," *IEEE Trans. Knowl. Data Eng.*, vol. 14, no. 2, pp. 438–444, Mar./Apr. 2002.

[23] O. Yildiz and O. Dikmen, "Parallel univariate decision trees," *Pattern Recognit. Lett.*, vol. 28, no. 7, pp. 825–832, May 2007.

[24] S. Baik and J. Bala, "A decision tree algorithm for distributed data mining: Towards network intrusion detection," in *Proc. Int. Conf. Comput. Sci. Appl.*, 2004, vol. 3046, pp. 206–212.

[25] N. Cristianini and J. Taylor, *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods*. Cambridge, U.K.: Cambridge Univ. Press, 2000.

[26] K. P. Bennett and O. L. Mangasarian, "Robust linear programming discrimination of two linearly inseparable sets," *Optim. Methods Softw.*, vol. 1, no. 1, pp. 23–34, 1992.

[27] P. Bradley, U. Fayyad, and O. Mangasarian, "Mathematical programming for data mining: Formulations and challenges," *INFORMS J. Comput.*, vol. 11, no. 3, pp. 217–238, 1999.

[28] O. Mangasarian, W. Street, and W. Wolberg, "Breast cancer diagnosis and prognosis via linear programming," *Oper. Res.*, vol. 43, no. 4, pp. 570–577, Jul./Aug. 1995.

[29] L. Breiman, J. Friedman, R. Olsen, and C. Stone, *Classification and Regression Trees*. Belmont, CA: Wadsworth, 1993.

[30] B. Schölkopf, C. Burges, and V. Vapnik, "Extracting support data for a given task," in *Proc. 1st Int. Conf. Knowl. Discov. Data Mining*, U. M. Fayyad and R. Uthurusamy, Eds., 1995, pp. 252–257.

[31] V. Blanz, B. Scholkopf, H. Bulthoff, C. Burges, V. Vapnik, and T. Vetter, "Comparison of view-based object recognition algorithms using realistic 3D models," in *Proc. ICANN*, vol. 1112. Berlin, Germany: Springer-Verlag, 1996, pp. 251–256.

[32] M. Schmidt, "Identifying speaker with support vector networks," in *Proc. Interface*, Sydney, Australia, 1996.

[33] E. Osuna, R. Freund, and F. Girosi, "Training support vector machines: An application to face detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 1997, pp. 130–136.

[34] T. Joachims, "Text categorization with support vector machines: Learning with many relevant features," in *Proc. 10th Eur. Conf. Mach. Learn.*, Chemnitz, Germany, Apr. 1998, vol. 1398, pp. 137–142.

[35] C.-W. Hsu and C.-J. Lin, "A comparison of methods multi-class support vector machines," *IEEE Trans. Neural Netw.*, vol. 13, no. 2, pp. 415–425, Mar. 2002.

[36] U. Krebel, "Pairwise classification and support vector machines," in *Advances in Kernel Methods—Support Vector Learning*. Cambridge, MA: MIT Press, 1999, pp. 255–268.

[37] C. Burges, "Tutorial on support vector machines for pattern recognition," *Data Mining Knowl. Discov.*, vol. 2, pp. 121–167, 1998.

[38] H. Shimodaira, K. ichi Noma, M. Naka, and S. Sagayama, "Support vector machine with dynamic time-alignment kernel for speech recognition," in *Proc. Eurospeech*, 2001, pp. 1841–1844.

[39] S. Rüping, "SVM kernels for time series analysis," in *Proc. LLWA—Tagungsband der GI-Workshop-Woche Lernen-Lehren-Wissen-Adaptivität*, R. Klinkenberg, S. Rüping, A. Fick, N. Henze, C. Herzog, R. Molitor, and O. Schröder, Eds., 2001, pp. 43–50.

[40] V. Wan and J. Carmichael, "Polynomial dynamic time warping kernel support vector machines for dysarthric speech recognition with sparse training data," in *Proc. Interspeech*, 2005, pp. 3321–3324.

[41] K. M. Borgwardt, S. Vishwanathan, and H.-P. Kriegel, "Class prediction from time series gene expression profiles using dynamical systems kernels," in *Proc. Pacific Symp. Biocomputing*, 2006, pp. 547–558.

[42] K. Yang and C. Shahabi, "A PCA-based kernel for kernel PCA on multivariate time series," in *Proc. ICDM Workshop Temporal Data Mining: Algorithms, Theory and Applications held in conjunction with 5th IEEE ICDM*, 2005.

[43] D. Pham and A. Chan, "Control chart pattern recognition using a new type of self organizing neural network," in *Proc. Inst. Mech. Eng. I, J. Syst. Control Eng.*, 1998, vol. 212, no. 2, pp. 115–127.

[44] E. Keogh and M. Pazzani, "An enhanced representation of time series which allows fast and accurate classification, clustering and relevance feedback," in *Proc. 4th ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining*, 1998, pp. 239–241.
[45] R. Olszewski, "Generalized feature extraction for structural pattern recognition in time-series data," Ph.D. dissertation, Carnegie Mellon Univ., Pittsburgh, PA, 2001.
[46] D. Eads, D. Hill, S. Davis, S. Perkins, J. Ma, R. Porter, and J. Theiler, "Genetic algorithms and support vector machines for time series classification," *Proc. SPIE*, vol. 4787, pp. 74–85, 2002.
[47] P. Geurts, "Contributions to decision tree induction: Bias/variance tradeoff and time series classification," Ph.D. dissertation, Univ. Liege, Liege, Belgium, 2002.
[48] E. Keogh and S. Kasetty, "On the need for time series data mining benchmarks: A survey and empirical demonstration," in *Proc. 8th ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining*, 2002, pp. 102–111.
[49] B. Chiu, E. Keogh, and S. Lonardi, "Probabilistic discovery of time series motifs," in *Proc. 9th ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining*, 2003, pp. 493–498.
[50] X. Xi, E. Keogh, C. Shelton, L. Wei, and C. A. Ratanamahatana, "Fast time series classification using numerosity reduction," in *Proc. 23rd Int. Conf. Mach. Learn.*, 2006, vol. 148, pp. 1033–1040.
[51] A. Mueen, E. Keogh, Q. Zhu, S. Cash, and B. Westover, "Exact discovery of time series motifs," in *Proc. SIAM Int. Conf. Data Mining (SDM)*, 2009, pp. 473–484.
[52] O. Omitaomu, M. Jeong, and A. Badiru, "Online support vector regression with varying parameters for time-dependent data," *IEEE Trans. Syst., Man, Cybern. A, Syst., Humans*, vol. 41, no. 1, pp. 191–197, Jan. 2011.
[53] B. Scholkopf, "The kernel trick for distances," in *Proc. Adv. Neural Inf. Process. Syst.*, 2001, pp. 301–307.
[54] B. Efron, "Estimating the error rate of a prediction rule: Improvement on cross-validation," *J. Amer. Stat. Assoc.*, vol. 78, no. 382, pp. 316–331, Jun. 1983.
[55] I. H. Witten and E. Frank, *Data Mining: Practical Machine Learning Tools and Techniques*, 2nd ed. San Mateo, CA: Morgan Kaufmann, 2005.

Wanpracha Art Chaovalitwongse (M'05–SM'11) received the B.S. degree in telecommunication engineering from King Mongkut's Institute of Technology Ladkrabang, Bangkok, Thailand, in 1999 and the M.S. and Ph.D. degrees in industrial and systems engineering from the University of Florida, Gainesville, in 2000 and 2003, respectively.

He is an Associate Professor of Industrial and Systems Engineering with Rutgers University, Piscataway, NJ. Before joining Rutgers, he was with Corporate Strategic Research, ExxonMobil Research and Engineering Company, in 2004, where he managed research in developing efficient mathematical models and novel statistical data analyses for upstream and downstream business operations.

Rebecca S. Pottenger is a rising junior at Princeton University, Princeton, NJ, in the Department of Computer Science.

She has been interested and involved in data mining research since her junior year of high school. In 2007, she conducted a research project entitled "Classifying String Quartets by Composer via Data Mining" which she presented at the Consortium for Computing Science in Colleges (2007), Fredericksburg, MD, and published in the Proceedings of the IEEE 2006 Lehigh Valley Science Technology Engineering and Math Conference, Bethlehem, PA, in November. She previously worked with Prof. W. Art Chaovaltiwongse at Rutgers University, Piscataway, NJ, through the Rutgers/Center for Discrete Mathematics and Theoretical Computer Science Research Experience for Undergraduates Program, studying the classification of epileptic patients.

Shouyi Wang received the B.S. degree in control science and engineering from Harbin Institute of Technology, Harbin, China, in 2003 and the M.S. degree in systems and control engineering from Delft University of Technology, Delft, The Netherlands, in 2005. Currently, he is working toward the Ph.D. degree in the Department of Industrial and Systems Engineering, Rutgers University, Piscataway, NJ, under the supervision of Prof. W. Art Chaovaltiwongse.

His current research interests include data mining and pattern discovery, machine learning, intelligent decision-making systems, multivariate time-series modeling, and forecasting.

Ya-Ju Fan received the B.B.A. degree in production and operations management from Fu Jen Catholic University, Taipei, Taiwan, and the M.S. degree in industrial engineering from the University of Wisconsin, Madison. In the fall of 2006, she joined the graduate program in the Department of Industrial and Systems Engineering, Rutgers University, Piscataway, NJ, to work with Prof. W. Art Chaovalitwongse. She received the Ph.D. degree from the Rutgers University in 2010.

Leon D. Iasemidis (M'96) received the Diploma degree in electrical and electronics engineering from the National Technical University of Athens, Athens, Greece, in 1982 and the M.S. degree in physics and the M.S. and Ph.D. degrees in biomedical engineering from the University of Michigan, Ann Arbor, in 1985, 1986, and 1991, respectively.

From 1991 to 1993, he was a Postdoctoral Fellow with the Departments of Electrical Engineering and Neurology, University of Michigan, Ann Arbor. From 1993 to 2000, he was the Director of the Clinical Neurophysiology Laboratory and the Founder of the Brain Dynamics Laboratory at the Malcolm Randall VA Medical Center, Gainesville, FL, and a Research Assistant Professor with the Departments of Electrical Engineering, Neurology, and Neuroscience, University of Florida, Gainesville. Since 2000, he has been an Associate Professor of Bioengineering with the Arizona State University (ASU), Tempe, and the Director and Founder of the ASU Brain Dynamics Laboratory. He is recognized as an expert in the dynamics of epileptic seizures, and his research and publications have stimulated an international interest in the prediction and control of epileptic seizures and the understanding of the mechanisms of epileptogenesis.