

A Comparative Study of Intracranial EEG Files Using Nonlinear Classification Methods

MARIA TITO,¹ MERCEDES CABRERIZO,¹ MELVIN AYALA,¹ PRASANNA JAYAKAR,² and MALEK ADJOUADI¹

¹Center for Advanced Technology and Education, Department of Electrical and Computer Engineering, Florida International University, 10555 W. Flagler Street, Miami, FL 33174, USA; and ²Brain Institute Department, Miami Children's Hospital, 3100 South West 62nd Avenue, Miami, FL 33155, USA

(Received 8 May 2009; accepted 6 October 2009; published online 17 October 2009)

Abstract—This study is a comparative evaluation of nonlinear classification methods with a focus on nonlinear decision functions and the standard method of support vector machines for seizure detection. These nonlinear classification methods are used on key features that were extracted on subdural EEG data after a thorough evaluation of all the frequency bands from 1 to 44 Hz. The sensitivity, specificity, and accuracy of seizure detection reveal that the gamma frequencies (36–44 Hz) are most suitable for detecting seizure files using a unique 2D decisional plane. We evaluated 157 intracranial EEG files from 14 patients by calculating the spectral power using nonoverlapping 1-s windows on different frequency bands. A key finding is in establishing a 2D decision plane, where duration of the seizure is used as the first dimension (x coordinate) and the maximum of the gamma frequency components is used as the second dimension (y coordinate). Within this 2D plane, the best results were observed when the nonlinearity degree is three for the proposed nonlinear decision functions, with a sensitivity of 96.3%, a specificity of 96.8%, and accuracy of 96.7%.

Keywords—Intracranial EEG (iEEG), Epilepsy, Seizure detection, Gamma frequency, Nonlinear decision functions.

INTRODUCTION

Epilepsy is the second most prevalent neurological disorder in humans after stroke. According to the World Health Organization, approximately 1% of the world population is affected by this neurological disorder. Epilepsy is a disorder with many possible causes, many of which remain unknown. Anything that disturbs the normal pattern of neuronal activity, from illness to brain damage to abnormal brain development, can unfortunately lead to seizures. Current

treatments for epilepsy include: anti-epileptic drugs, brain therapies that involve direct electrical brain stimulation. However, not all patients benefit from these treatments, in which case brain surgery may be the only resort, especially for intractable seizures. Most of the treatments for intractable seizures are very limited. The most critical involves focal resections of abnormal brain tissue when the epileptogenic region can be accurately defined. This is a critical task that requires subdural EEG recordings of seizures to define their onset, electrodes of interest, and their region of involvement.

The main objective of this study endeavor is focused on developing an automated algorithm for the detection of seizures offline, based on subdural EEG data that would satisfy a high sensitivity (i.e., minimum number of false negatives), and a high specificity (i.e., minimum number of false positives). The stopping conditions are implemented in order to achieve the highest accuracy, so the main objectives are satisfied.

Since epileptic seizures occur erratically and randomly, enormous amounts of subdural EEG needs to be read and precisely analyzed offline in order to detect the origin of the seizures. This is a critical challenge that can be automated through reliable and computationally efficient seizure detection paradigms.

Several studies related to automated seizure detection paradigms focusing mainly on scalp EEG have been published with diverse level of success and inherent challenges.^{3,7,11,12,19,20,32} These investigations employ the use of multichannel trends, neural networks applications, the use of orthogonal transforms, such as the Walsh Transform, genetic programming, and all dwell in either time or frequency domains. The results are based on EEG recorded from the scalp that has lower signal-to-noise ratio compared to the EEG recorded from the cortex. Since one of the main goals in the area of epilepsy is to detect and ultimately to predict seizures, the use of several features has been

Address correspondence to Mercedes Cabrerizo, Center for Advanced Technology and Education, Department of Electrical and Computer Engineering, Florida International University, 10555 W. Flagler Street, Miami, FL 33174, USA. Electronic mail: cabreriz@fiu.edu

adopted by various research groups for many years with varying degrees of success. In the context of this study, many of the methods currently available in the specialized literature have been tested yielding different results. In this field, the issue of debate is not based in the implementations of all these features, but in determining which ones are more suitable. In this initial assessment of the study, it was determined in a previous paper that the correlation sum is the temporal measure that performed best, while the gamma power, as reported specifically in this study, is found to be the most revealing frequency range for seizure detection purposes.

EEG signals have been analyzed over the years with much effort toward a better understanding of the functional characteristics of the brain, including the complex and yet to be resolved problem of seizure prediction.^{14,18,24,26,30} Researchers have thus considered different approaches using a diversity of linear and nonlinear parameters in order to automate processes of seizure detection, eliciting a better understanding of the chaotic dynamics in biological systems,^{6,15,21,23} and where promising results have been substantiated.^{1,2,16,29,31}

In using subdural EEG data in this study, we have taken into consideration the fact that there is a considerable attenuation effect of the skull on the scalp EEG. This data generally reveal very little fast activity exceeding the beta range [> 30 Hz], limiting as a result the application scope of seizure detection algorithms that rely on scalp EEG recordings since there is a limitation in the use of higher frequency components. A sustained increase in the very high frequency activity exceeding 30 Hz, defined here as the gamma band, is seen only at ictal state onset and during early evolution of the seizure. These empirical facts served as the foundation of this research endeavor.

In this study, we, therefore, explore the role of the gamma frequency band in developing a reliable offline seizure detection algorithm for EEG recorded intracranially, by establishing a unique 2D decision space and implementing nonlinear decision functions to confront the complex nature of iEEG data. The proposed method is based on aggregating the power in the 36–44 Hz frequency range and analyzing its behavior in time using windows of 1 s of duration, looking for patterns indicative of seizure evolution. The performance of the algorithm, which was evaluated by means of the receiver operating characteristics (ROC) terminology, relied on two primary aims: (1) establishing a decision space most suitable for iEEG data classification, and (2) implementing to their full extent generalized decision functions that are operational under any number of dimensions and for any degree of nonlinearity with the gradient descent method used as

means to generate the weights for the highest classification accuracy possible.

METHODS

Data Collection

The data used in this study were obtained sequentially from a significant sample of 14 patients who underwent two-stage epilepsy surgery with subdural recording. The age of the subjects varied from 3 to 17 years. An overview of the patients' clinical information is provided in Table 1. The number and configuration of the subdural electrodes differed between subjects, and were determined by clinical judgment at the time of implantation. Grid, strip, and depth electrodes were used, with a total number of contacts varying between 20 and 88. The amount of data available for analysis was influenced by recording duration, and by the degree to which the interictal EEG was "pruned" prior to storage in the permanent medical record. The intracranial EEG (iEEG) data were recorded at Miami's Children Hospital (MCH) using XLTEK Neuroworks Ver.3.0.5, equipment manufactured by Excel Tech Ltd., Ontario, Canada. The data were collected at 500 Hz sampling frequency, and the DC component was removed.

Data Analysis

Data Preprocessing

For all patients, the length of the files was approximately 10 min. Out of the total 157 files considered, 35 (21 interictal and 14 ictal) iEEG data files or 22% were used initially in a training phase to ascertain the reliability of the gamma power in the seizure detection process. The remaining 122 iEEG data files or 78% were then used in the testing phase to assess the merits in selecting gamma power as means to detect a seizure. In retrospect, the ictal and interictal files used for each patient are summarized in Table 2. Each file was categorized by whether or not it contained a seizure, and were randomly assigned to avoid unwanted biases to either the small training set (22% of the data) or the large testing set (the remaining 78% of the data).

Extracting the Power of Gamma Frequency

The gamma frequency power of each electrode was calculated from the EEG data using consecutive 1-s windows (500 samples). Fourier Transform was applied to each window in order to extract the frequency components of interest.

TABLE 1. Patients' clinical information.

Patient	DOB (m/y)	Sex	Clinical overview	EEG SPECS
1	9/92	M	Tuberous sclerosis with intractable epilepsy since 2 years of age. S/p resection of R frontal tuber/hamartoma resection but not seizure-free.	Right frontotemporoparietal craniotomy, L parietal craniotomy.
2	4/87	M	Intractable seizures that is now s/p craniotomy. 1 seizure per day.	PMH intractable seizures. S/p craniotomy Right craniotomy, frontal lobectomy and removal of electrodes.
3	1/01	F	R partial resection. Abnormal movements (tremors, shivering, shaking), ADHD, and ADD.	Normal EEG study obtained in the awake and drowsy states.
4	10/90	M	Intractable epilepsy and L temporal cortical dysplasia. Three antiepileptics (lamictal, trileptal, and topamax without response). L Craniotomy with L anterior temporal lobe resection.	Intractable seizure unresponsive to medical therapy. Craniotomy with L anterior temporal lobe resection.
5	5/96	M	ADHD and partial epilepsy, s/p right FT resection. Seizures are now associated with more vocalization and are now only nocturnal. They recur two to three times per day.	Seizures in the R hemisphere. Interictal abnormalities are limited to some slowing involving the RF central region and some rare sharply contoured slow waves over the R frontal region and L temporal slowing.
6	6/94	F	Intractable epilepsy. Left frontotemporal craniotomy.	Sharp waves are recorded in the L frontotemporal region. Epileptiform activity over the mesial temporal surface. Ictal EEG discharges involve the left temporal and central regions. EcoG shows frequent epileptiform abnormalities over the pole and superior temporal gyrus.
7	7/91	M	Seizures since age 10. Develops sensory symptoms at onset, paresthesias behind right thigh.	Slower frequencies noted maximal in the L central region, also L parietal region during sleep, Some sharp waves in the L central and frontal regions, in sleep.
8	7/98	M	Seizures (1–3/week) despite a tritherapy. Seizures started at 6 months of age.	EEG and clinical findings are consistent with frontal lobe epilepsy with interictal abnormalities and may be ictal behavior suggestive of L frontal origin.
9	12/97	M	Simple partial seizures since 3 years of age. Syndrome: Symptomatic partial seizure disorder. Etiology: Lesion in R central area.	Slower frequencies in both central regions, R > L. Focal seizure pattern with build up over the R central region, with propagation to the L frontal/central region.
10	9/91	F	Intractable epilepsy. History of glioma and hydrocephalus s/p shunt placement and intractable seizures S/p right frontal craniotomy, right temporal lobectomy and subdural electrode removal. Surgery was successful.	Seizure began in 11/96. It is believed to have seizures arising for the R temporal lobe with secondary propagation to the supra-sylvian region. Seizure foci identified in R temporal lobe.
11	1/93	M	History of L temporal lobe ganglioglioma. Semiology: Complex partial seizures. Syndrome: Symptomatic partial seizure disorder. Etiology: Lesion in the left temporal region.	Epileptiform activity in the L temporal region, rare sharp waves in the L posterior quadrant. Focal ictal patterns in the L temporal region, with late hemispheric spread.
12	1/01	M	Intractable epilepsy since 1.5 year of age. Freq is 3–5 per day. Duration is 30 seconds. Seizure classification: Complex partial seizures of the R central area. Epilepsy syndrome: localization-related.	Rare spikes seen over R central region. Three seizures that are stereotyped with C4 onset and LUE flexion and behavioral arrest.
13	1/97	M	Intractable seizures, having two-stage epilepsy surgery. Severe cerebral dysfunction in the region of the upper parietal interhemispheric region.	Frequent epileptiform activity was recorded inferior to the upper parietal interhemispheric region as well as on the midsection of the parietal convexity behind the grid.
14	2/03	F	Tuberous sclerosis and medically intractable seizures. Seizures are clinically brief periods of behavioral change. Seizure type/semiology: complex partial seizures. Syndrome: Symptomatic partial seizure disorder. Etiology: Tuberous sclerosis.	Multifocal epileptiform activity adding emphasis over the L temporal region. Multiple seizures having nonlocalizing patterns at ictal onset, with predominant involvement of the L frontotemporal region late in the event.

Due to the high volume of information contained in the prefiltered iEEG data files, two key preprocessing steps were performed in order to (1) reduce

the data to be analyzed, reducing as a consequence the computational requirements, and (2) seek a transformation of the raw iEEG data in order to enhance the

TABLE 2. Number of ictal and interictal files used.

Patient	Training files		Testing files		All files	
	Interictal	Ictal	Interictal	Ictal	Interictal	Ictal
1	0	3	0	2	0	5
2	0	1	0	1	0	2
3	0	2	0	1	0	3
4	0	1	0	1	0	2
5	6	3	30	3	36	6
6	5	2	10	1	15	3
7	10	2	5	1	15	3
8	0	0	12	2	12	2
9	0	0	9	5	9	5
10	0	0	15	2	15	2
11	0	0	4	2	4	2
12	0	0	5	3	5	3
13	0	0	1	1	1	1
14	0	0	4	2	4	2
Total	21	14	95	27	116	41

accuracy, specificity, and sensitivity of the seizure detection algorithm. In step (2), the transformation chosen is that of gamma frequency component after a thorough evaluation of several other standard parameters in the time domain, such as mobility, complexity, and activity,²² as well as in the frequency domain by evaluating all other frequency ranges. The iEEG data files were further analyzed with 1 s timed windows, and the power of the gamma frequency component was extracted for all these 1-s windows and for each electrode.

The power P_g of the gamma frequency spectrum was computed as given in Eq. (1):

$$P_g = \int_{b_{\text{start}}}^{b_{\text{end}}} |F(w)|^2 dw \quad (1)$$

where b_{start} and b_{end} are its starting and ending frequencies in the 36–44 Hz band, with $F(w)$ defining the complex fast Fourier coefficient at frequency w .

Figure 1 shows the gamma power near the time of seizure onset using iEEG data from four different patients. It shows the intricate and yet informative nature of the gamma power signal. At the time of seizure onset (vertical line), there is an abrupt change in magnitude almost in synchrony for all electrodes. The vertical line represents the seizure onset previously labeled at the observation room by the EEG expert. It should be noted that the onset as marked by the EEG expert and the results of the synchronized increment in magnitude of the gamma frequency do not coincide exactly in time, which only heightens the relevance and need for an automated seizure detection process. Further clinical evaluations reveal that the

synchronized increment in magnitude of the gamma frequency does actually coincide with the actual clinical onset of the seizure. This is viewed as another interesting finding of this study.

Aggregating EEG Features

As stated earlier, 35 files were randomly assigned for use in the so-called training phase, and the remaining 122 files were used for testing. If the file contained a seizure, it was considered to be an “ictal” file. Otherwise, it was considered to be “interictal.” It is worthy to note that even though a file was classified as “ictal,” the files usually lasted longer than an individual seizure. Therefore, it was possible for “ictal” files to include some interictal data, which nonetheless the classifier needed to handle correctly.

Table 1 shows the distribution of patients along with the number of ictal and interictal files that were selected for both the training and testing phases. The table is set up such that patients (1–7) were those that were randomly selected for the training phase, and patients (8–14) were those used subsequently in the testing phase in order to validate the classifier’s ability to perform well on an inter-patient level.

In order to handle the variable number of electrodes used from patient to patient, we averaged the power in the gamma frequency range across all electrodes. This averaging process, which is referred to as the inter-electrode mean, was used as input to the classifier. Our use of the inter-electrode mean is a result of the experimental studies^{4,5,9,27} that reveal that electrodes tend to interlock in behavior at the onset moment of a seizure. Therefore, this average process of all electrodes did not distort the results, and yet allowed for uniformity in the implementation process across patients independent of the varied number of electrodes used for each.

With this fact, it is emphasized that the concept of averaging for a representative signal does not sidetrack from the main intent of detecting a seizure with the highest accuracy, specificity, and sensitivity possible. At the same time, such a step minimizes to a great extent the computational burden¹⁰ that would have been required in dealing with all of the iEEG data as input to the classifier, and simplifies greatly the seizure detection process as only one representative signal is fed into the classifier. Figure 2 illustrates this assertion by comparing the contributions of individual electrodes to the behavior of gamma frequency for every single electrode used for a particular patient, and the results obtained using averaging or the so-called inter-electrode mean signal S_μ for an arbitrary section of iEEG.

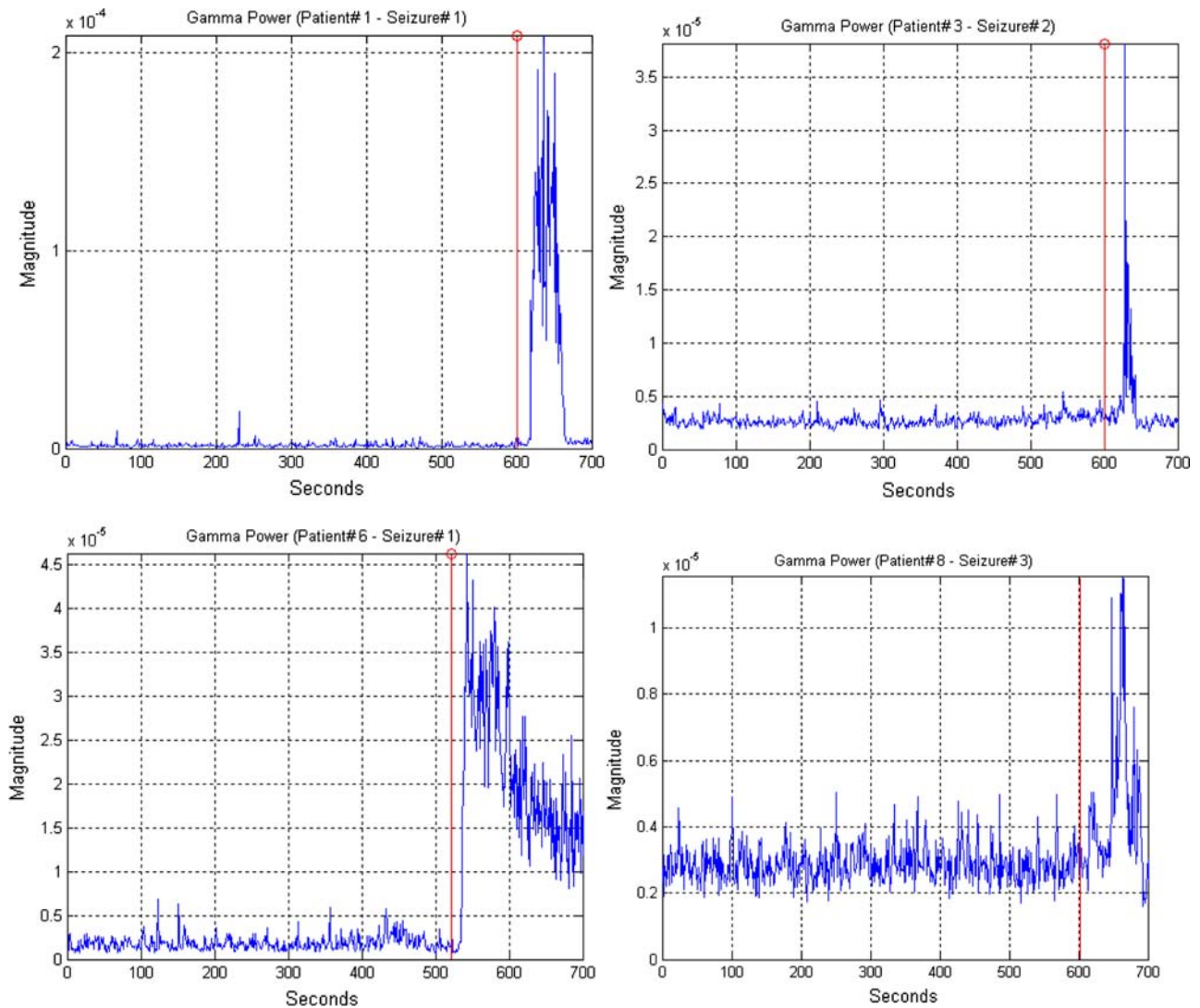


FIGURE 1. An illustrative example of gamma power for all electrodes versus time (shown in terms of samples) for four different seizures from different patients. The vertical line is the seizure onset as identified by medical experts. Note that the scales are different for the different subjects, thus making absolute thresholds impossible to use.

Determining the Seizure Detection Rules

Critical Issue of Thresholding

A threshold must be established before the inter-electrode mean of the power in the gamma frequency band can be used to detect a seizure. If the threshold was crossed at any point during a file, then the entire file is classified as a possible “seizure” file. If the threshold was never crossed, then it was classified as a “nonseizure” file. At stake is therefore an objective choice of this threshold.

Threshold Determination

The seizure detection method tests a threshold based on the inter-electrode mean signal S_{μ} , which has a

natural variability even in a single patient, and even in the absence of a seizure. Using the same threshold in many patients makes the variability even greater, but increases the clinical usefulness of the test. This is essentially a dilemma that is faced due to unreliable and changing thresholds and varying standard deviations that can be experienced even within a single patient.

With these observable facts, the problem becomes difficult to contain not only in terms of these noted variations, but also in ascertaining in a meaningful way the performance evaluation of the classifier. The example given in Fig. 3 illustrates perfectly this dilemma. Note how different is the variation of the magnitude of the inter-electrode mean signal between patient 1 and patient 2. However, if one is to rescale the y -axis for patient 2, it will reveal that an ictal

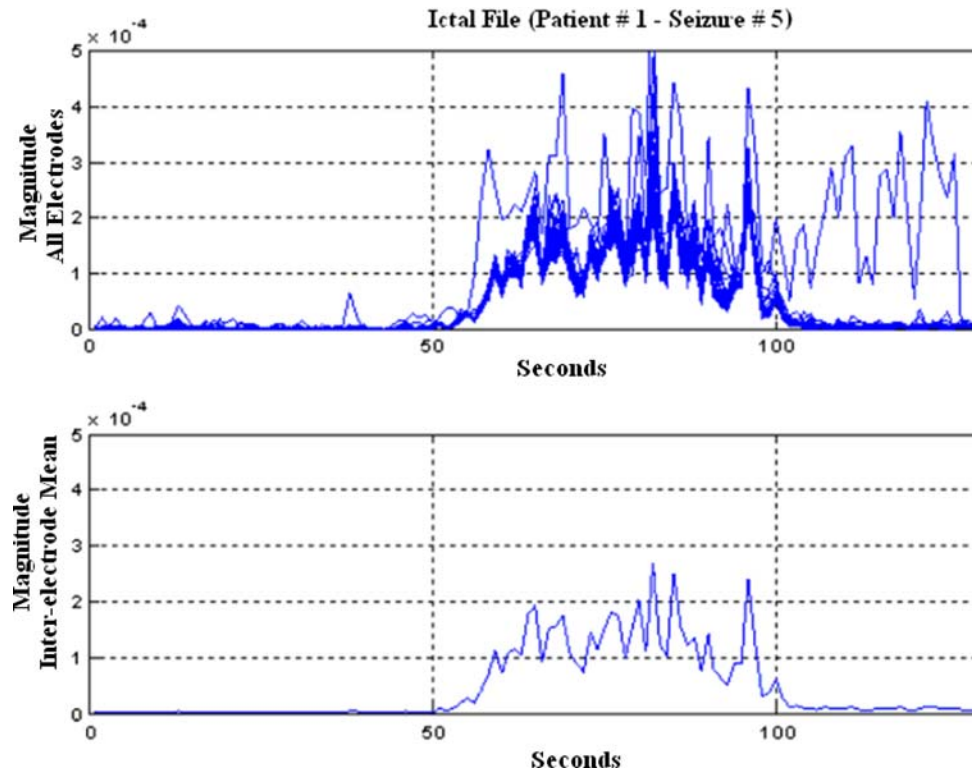


FIGURE 2. Top figure shows the behavior of the gamma power for each of the 48 electrodes used for subject 1, seizure 5, and the bottom figure displays its respective inter-electrode mean signal.

change similar to that of patient 1 is indeed present, which only amplifies the fact that a singular threshold computed on the basis of patient 1 would have missed the ictal change in patient 2 observed after rescaling.

Therefore, a generalized statistical threshold was established for the power of the gamma frequency that will work across all patients independently of its magnitude. This threshold is thus defined by the average of inter-electrode mean plus one standard deviation of the inter-electrode mean signal as defined in Eq. (2).

$$T = \overline{S_\mu} + \sigma \quad (2)$$

In terms of the proposed classifier, any point which exceeded the threshold was considered a point belonging to a potential seizure, with the first such point together with all consecutive points exceeding this threshold defining the duration of such seizure. Such a determination constitutes a first and most critical requirement for the proposed algorithm.

The next stage of the algorithm is to determine additional means to validate that such potential seizure can indeed be declared an actual seizure.

Establishing the 2D Decisional Space

At this stage of the investigation, two measurements were taken into consideration. The first measurement

is the duration in time in which the signal S_μ was consistently above the aforementioned threshold, and the second measurement is found to be the maximum value of S_μ . These two measurements were deemed sufficient for the detection algorithm to work properly.

With these two measurements in place, a table was constructed to train the seizure detector. The table contains as many records as data files were used in the training, whereas each record contains three values: the two aforementioned measurements and also a target (+1 for seizure file and -1 for nonseizure file). Recall that nonseizure data files that did not meet the first requirement (not having a single point that passed the set threshold) were not used in this table to begin with, since they were already identified as true negatives.

Nonlinear decision functions were derived using the training data. The results are displayed in the 2D decisional space, which helps in visualizing the geometrical placement of the two pattern classes with respect to the resulting decision function. All points on the decision function curve itself are considered of an undetermined class, since the decision function is zero for those points. In the result plots, positive and negative signs denote seizure and nonseizure files, respectively. In this, the 2D decisional space, the x -axis represents the duration which was divided by a normalization factor of 1000 in order to accelerate the

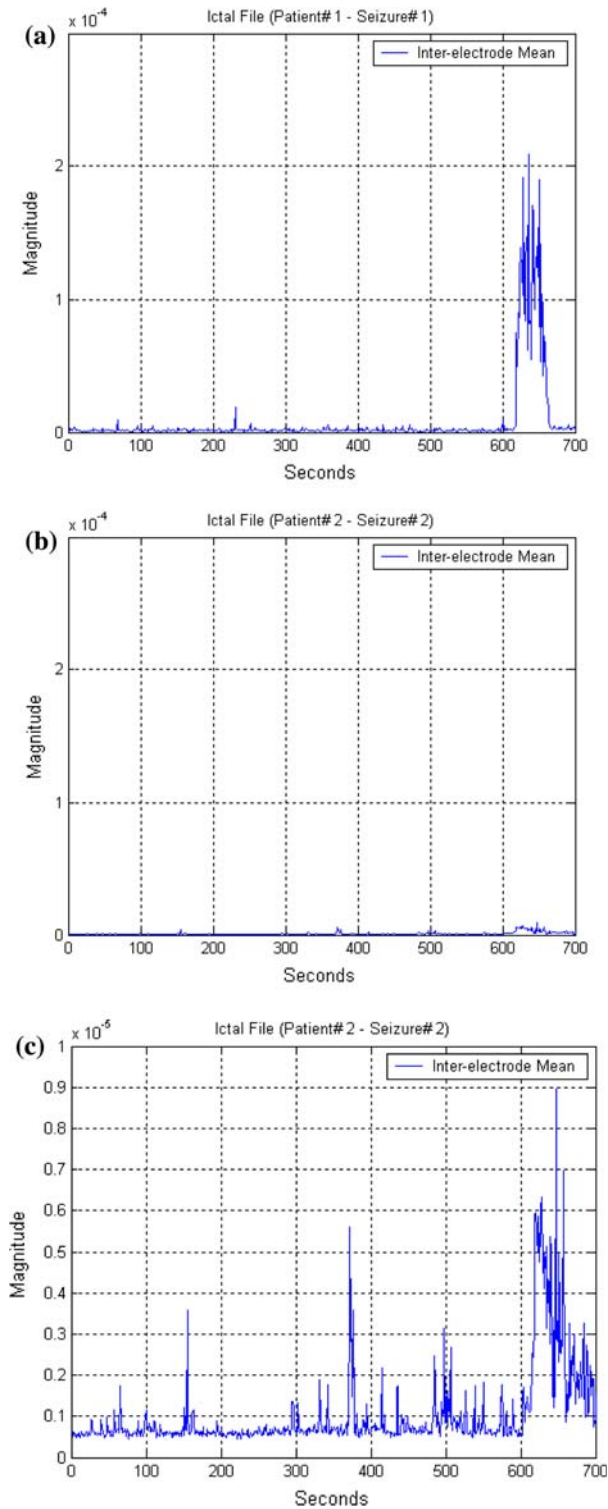


FIGURE 3. Illustration of the variation of the inter-electrode mean signal S_μ within the same patient: (a) seizure 1 of patient 1, (b) seizure 2 of patient 2, and (c) zoomed in view of seizure 2 of patient 2.

convergence of the gradient descent, and thus facilitate the determination of the optimal weights of the decision function; while the y -axis represents the maximum

of S_μ . When this 2D space is chosen appropriately, which was the most challenging part of this research problem, patterns of the same class will tend to cluster together and the nonlinear decision functions will yield optimal results.³³

Derivations of the Nonlinear Decision Functions

To address this problem and begin its implementation steps, each nonlinear decision function is obtained from the training data using the gradient descent method. The gradient descent method is known as an optimization algorithm that seeks a local minimum for a given function. The problem of classification is therefore one that is reduced to identify the boundary $D(\vec{X}) = 0$, which defines the decision function itself, in order to partition the pattern space into different classes. Vector \vec{X} is the pattern with up to n dimensions that needs to be classified. In the case treated here, the classification rules are as follows:

- If $D(\vec{X}) > 0$, $\vec{X} \in$ the class of seizure files
- If $D(\vec{X}) < 0$, $\vec{X} \in$ the class of nonseizure files
- If $D(\vec{X}) = 0$, indeterminate condition

Nonlinear decision functions have long been established in the literature,⁴ but it is their implementation that is regrettably lacking in the resolution of real-world problems such as the one addressed in this study. The general recursive formula as expressed in Eq. (3) allows for deriving any nonlinear decision function with any number of dimension (n) and with any degree of complexity (r), and is structured as follows:

$$D^r(\vec{X}) = \sum_{P_1=1}^n \sum_{P_2=P_1}^n \dots \sum_{P_r=P_{r-1}}^n W_{P_1 P_2 \dots P_r} X_{P_1} X_{P_2} \dots X_{P_r} + D^{r-1} = 0 \quad (3)$$

where $W_{P_1 P_2 \dots P_r}$ represents the weights and $X_{P_1} X_{P_2} \dots X_{P_r}$ are the dimensions of the function. This equation which is recursive in r is defined such that $D^0(\vec{X}) = W_{n+1}$, which is the last element of the augmented weight vector. In this particular study, different results are later provided for different values of r for both the proposed nonlinear decision functions as well as for the well-established SVM method under different complexities of the polynomial kernel.

The way the proposed classification process was set up resulted in decision functions that are derived using two variables X_1 and X_2 which define the two identified dimensions of the decision plane. Thus, for the problem at hand, Eq. (3) with $n = 2$ and $r = 2$ (second degree polynomial) for example, reduces to Eq. (4):

$$D(\vec{X}) = W_{11}X_1^2 + W_{12}X_1X_2 + W_{22}X_2^2 + W_1X_1 + W_2X_2 + W_3 = 0 \quad (4)$$

where W_{11} , W_{12} , W_{22} , W_1 , W_2 , and W_3 are the weights; and X_1 and X_2 are representatives of the first and second dimensions, respectively. To obtain the optimum parameters of this function (3), a cost function C was established as the sum of the square errors between the function output and the targets as follows:

$$C = \sum_{p=1}^N (D_p(\vec{X}) - T_p)^2 \quad (5)$$

where p serves as the index for the data points, where each data point is represented by its dimensions X_1 and X_2 in reference to Eq. (4). N is the number of data points in the files that have met the first requirement for seizure detection; T_p is the target for point p (-1 or $+1$), and $D_p(\vec{X})$ is the resulting value of the decision function evaluated at point p . The optimal weights, which define the ultimate nonlinear decision function, are those that yielded a minimal value for the cost function given in Eq. (5).

Nonlinear decision functions are far superior to linear decision functions and guarantee better separability in complex and often overlapping datasets, which is the case in most real-world datasets. Other good candidates to deal with high data overlap are support vector machines (SVMs) which have known significant success in these types of data classification.^{8,13,34,35}

SVM method can best be described as generalized linear classifiers which map input vectors to a higher dimensional space (feature space) where a maximal separating hyperplane is sought to separate two different classes. The idea is thus to minimize the classification risk by maximizing the inter-class margins or distances. The mapping of the data into the so-called feature space is accomplished by means of kernel functions and the decision function is expressed in terms of the so-called support vectors. However, SVM rely on a set of prescribed kernels that can be used for this problem of complex mapping. As opposed to SVM, which remains a powerful classification tool, the proposed nonlinear decision method allows us to take full ownership of its implementation with the flexibility of changing the complexity order (r) and investigating any number of potential dimensions (n) and assess their individual merit in the classification process of the dataset and can accommodate any number of dimensions as can be seen from Eq. (2). It is this flexibility of the nonlinear decision functions and their ability to handle complex overlapping data that led to the proposed nonlinear approach. Furthermore, the minimization technique represented by Eq. (4) was selected because it allowed for defining any type of decision function, and yet optimization expressions can be established for any specific measures for seizure detection.

The minimization of the cost function for all data points is achieved by setting the partial derivatives of this cost function with respect to all the weights to zero. This generates a system of equations that can be solved iteratively following the gradient descent method approach, in which the weight increments are dictated by a learning rate that is chosen to be small enough to simultaneously seek high accuracy and fast convergence rates. Thus, for a given point p , the notation as used in Eq. (5) simplifies to the following equation:

$$C = (D - T)^2 \quad (6)$$

to which the chain rule is applied in the following way:

$$\frac{\partial C}{\partial W} = \frac{\partial C}{\partial D} \frac{\partial D}{\partial W} \quad (7)$$

where

$$\frac{\partial C}{\partial D} = 2(D - T) \quad (8)$$

and evidently, the value of $\partial C/\partial D$ depends on the specific decision function used. For simplicity in the notation, parameter (\vec{X}) is excluded in the formulation of C and D .

By subsequently applying this procedure with respect to all the weights, the following set of equations are obtained:

$$\left[\begin{array}{l} \frac{\partial D}{\partial W_{11}} = X_1^2 \Rightarrow \Delta W_{11} = 2\lambda(T - D)X_1^2 \\ \frac{\partial D}{\partial W_{12}} = X_1X_2 \Rightarrow \Delta W_{12} = 2\lambda(T - D)X_1X_2 \\ \frac{\partial D}{\partial W_{22}} = X_2^2 \Rightarrow \Delta W_{22} = 2\lambda(T - D)X_2^2 \\ \frac{\partial D}{\partial W_1} = X_1 \Rightarrow \Delta W_1 = 2\lambda(T - D)X_1 \\ \frac{\partial D}{\partial W_2} = X_2 \Rightarrow \Delta W_2 = 2\lambda(T - D)X_2 \\ \frac{\partial D}{\partial W_3} = 1 \Rightarrow \Delta W_3 = 2\lambda(T - D) \end{array} \right] \quad (9)$$

In these derivations, λ represents the learning rate that is assumed during the training phase of the algorithm. A set of weights was evaluated and recalculated throughout the weights optimization process. As a compromise between faster convergence rates and maintaining high accuracy in the classification results, λ was empirically chosen as 0.1.

With all the weights determined, the decision functions expressed through Eq. (3) can now be established. Since the training files constitute a relative small percentage of the total number of data files,

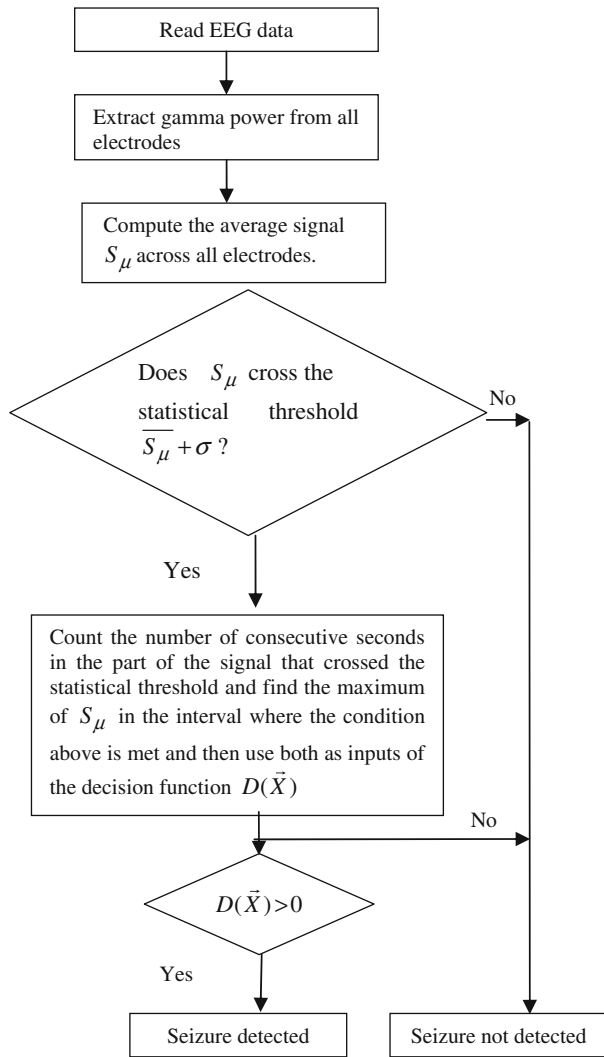


FIGURE 4. Flowchart of the procedural steps followed for each file.

training was set to stop when classification accuracy of all patterns no longer improved with subsequent iterations.

The different steps of the procedure that were followed each time a specific file is processed are as displayed in Fig. 4. It is noted that there are essentially two conditions that need to be satisfied for a file to be declared properly as containing a seizure: (a) crossing the set statistical threshold, which is the main critical condition that needs to be met first, and (b) yielding a positive value for the nonlinear decision function.

Testing Results

Once the detector is trained, the remaining data were consequently used to test the accuracy of the classification process using the same decision function

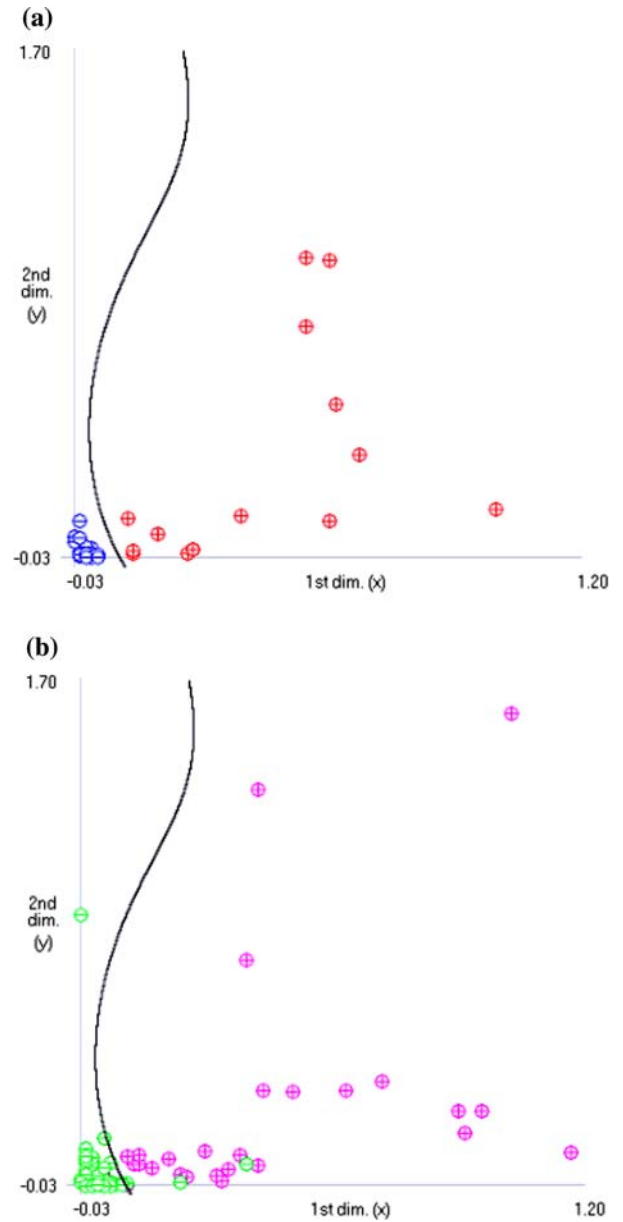


FIGURE 5. Positive (ictal files) and negative (interictal files) points used for training and testing and plot of the decision function. The x -axis represents the duration, whereas the y -axis represents the maximum value of the inter-electrode gamma power. (a) training data ($r = 3$), (b) testing data ($r = 3$).

generated in the training phase. Figure 5 shows plots of the decision functions and the training and testing points used to create and test the detector, respectively. In these figures, each point represents a file that meets the first requirement needed for a file to contain a seizure. The x -axis represents the first measurement (duration or X_1 of the generalized Eq. 3); and the y -axis represents the second measurement (maximum of S_μ or X_2 of the generalized Eq. 3). In the training set, the x -axis (duration) was normalized from zero to

one and in the testing set the duration axis was normalized with respect to the training set. However, since the normalization was done with respect to the training set, normalized values of the testing set with respect to the training set could exceed the values of 1. The 1.2 value in the testing plot Fig. 5b represents the maximum value of the testing set already normalized with respect to the training data (Fig. 5a).

The plots show a remarkable separation of seizure and nonseizure data files with their unique dispersion and clustering characteristics.

The confusion matrices that were obtained are displayed in Table 3 with their corresponding accuracy, specificity, and sensitivity values. As it can be observed from Table 3, the total amount of false detections (FN and FP) is small, and the TP value is rather high. Also accuracy is above 96%. Notice that patients 12 and 14 show the lowest sensitivity. Interestingly, these two patients were not included in the training phase and they are the youngest from that group. They both have complex partial seizures.

Based on the inter-patient nature of this procedure, the results can be considered as highly significant.

Comparative Analysis of Nonlinear Classifiers Under Different Frequency Bands

In order to demonstrate the superiority of the gamma frequency power, other frequency bands were also analyzed and compared in terms of their classification performance. From Table 4, it can be observed that as an overall, the best results were obtained under the gamma frequency. At this juncture, it is noted that the data are assumed without normalization and that the degree of nonlinearity is set to 2 (quadratic) for all frequency bands as we seek uniformity in this

TABLE 3. Results obtained after testing on the 122 files across all patients.

Patient	TN	FP	FN	TP	Accuracy	Sensitivity	Specificity
1	N/A	N/A	0	2	N/A	1.0	N/A
2	N/A	N/A	0	1	N/A	1.0	N/A
3	N/A	N/A	1	0	N/A	0.0	N/A
4	N/A	N/A	0	1	N/A	1.0	N/A
5	12	0	0	2	1.0	1.0	1.0
6	10	0	0	1	1.0	1.0	1.0
7	5	0	0	1	1.0	1.0	1.0
8	29	1	0	3	0.9697	1.0	0.9667
9	9	0	0	5	1.0	1.0	1.0
10	15	0	0	2	1.0	1.0	1.0
11	3	1	0	2	0.8333	1.0	0.7500
12	5	0	1	2	0.8750	0.6667	1.0
13	1	0	0	1	1.0	1.0	1.0
14	4	0	1	1	0.8333	0.5000	1.0
Total	93	2	3	24	0.9590	0.8889	0.9789

comparative study. Through similar experimental evaluations, the same conclusion is drawn for higher values of (r).

Classification Results: Nonlinear Decision Functions vs. Support Vector Machines

As the gamma frequency is firmly established as the band most revealing in terms of seizure detection, the gamma power is thus computed for all the data files, and the classification process can now proceed. Through the gamma frequency band, the input data files were thus classified using the determined nonlinear decision functions and the results obtained were compared to the results produced using the SVM method under different degrees of nonlinearity. Table 5 summarizes these results. As noted earlier, the training dataset which is used as input for generating the weights of the nonlinear decision functions were normalized to accelerate the convergence rate. The testing data were also normalized but with respect to the training set. Figure 6 depicts a 2D decision plane showing the spread of the training dataset through the gamma frequency band using normalized plane and also shows a zooming view of the normalized plane for better visualization of the overlapped “-” points.

TABLE 4. Performance metric of the classifier for different frequency bands using nonnormalized input data.

Nonlinear decision functions ($n = 2, r = 2$)							
BAND	TN	FP	FN	TP	Accuracy %	Sensitivity %	Specificity %
Delta	93	2	14	13	86.89	48.15	97.89
Theta	95	0	11	16	90.98	59.26	100.00
Alpha	94	1	14	13	87.70	48.15	98.95
Beta I	95	0	10	17	91.80	62.96	100.00
Beta II	92	3	6	21	92.62	77.78	96.84
Gamma	93	2	3	24	95.90	88.89	97.89

TABLE 5. Performance evaluation of the nonlinear decision functions (NDF) vs. SVM method using gamma power with 22% of the normalized data used for training and 78% of the data used for testing.

	Accuracy %	Sensitivity %	Specificity %
SVM			
Polynomial ($r = 2$)	95.08	88.89	96.84
Polynomial ($r = 3$)	95.08	96.3	94.74
Polynomial ($r = 4$)	94.26	92.59	94.74
NDF			
Polynomial ($r = 2$)	92.62	77.78	96.84
Polynomial ($r = 3$)	96.72	96.30	96.84
Polynomial ($r = 4$)	96.72	92.59	97.89

The results shown are those obtained for the testing data only with the same decision function established in the training phase.

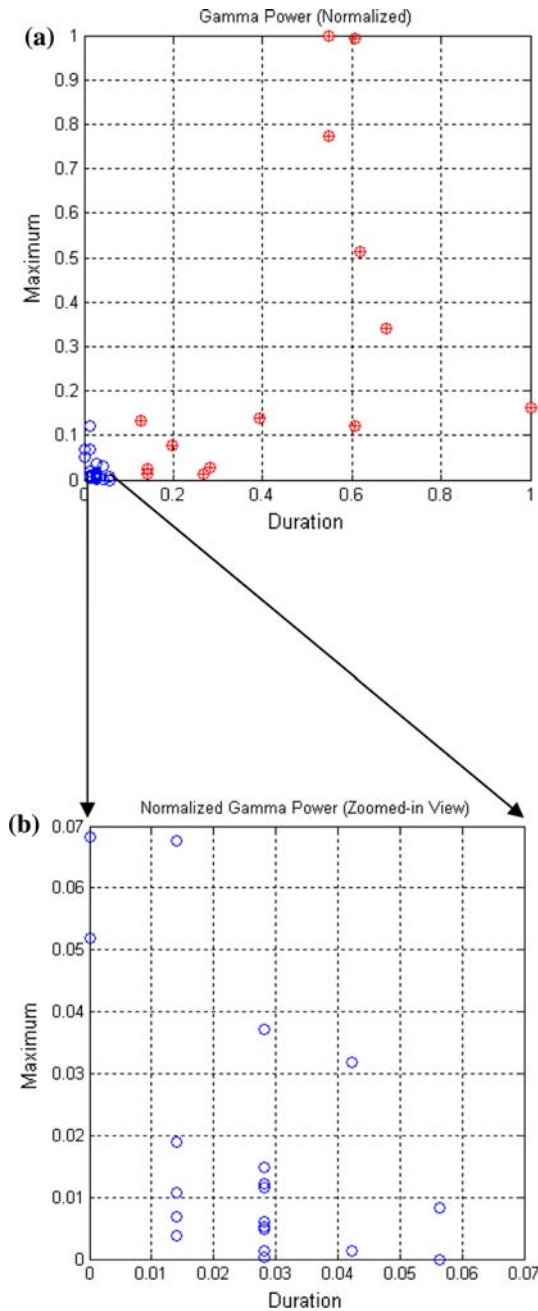


FIGURE 6. 2D decision plane showing the spread of the training dataset through the gamma frequency band using (a) normalized plane displayed for convenience, and (b) a zooming view of the normalized plane in (a) for better visualization of the overlapped “-” points. The x-axis represents the duration ($x \times 1000$ s), whereas the y-axis represents the maximum value of the inter-electrode gamma power.

To ensure the validity of such a normalization process, observe in Fig. 6 that the spread of the data remains the same for both nonnormalized and normalized datasets.

It can be concluded that the nonlinear decision function implemented with $r = 3$ performed better

across the board in terms of accuracy, sensitivity, and specificity over SVM. For the data considered, experimental results reveal that a higher degree in the nonlinearity ($r > 3$) does not necessarily improve the classification results, but that an increase in processing time is experienced as anticipated: for $r = 2$ it took about 6 min and for $r = 3$ it took about 7 min), since it takes more time to generate more weight elements. But it is important to be able to change this degree of nonlinearity if other more data is to be taken into consideration in the future, as more subjects consent to be part of this study. Furthermore, unlike the gradient descent methods, the computational complexity of SVM does not depend on the dimensionality of the input space, which is a powerful characteristic of SVM. However, with the rescaling (normalization) of the data, the NDF method performed at a much faster rate. It is important to emphasize again that once the decision function is generated in the training phase, the same decision function is used in the testing phase. More importantly, with the use of NDF, the process of transforming the input data to a new feature space as required in SVM prior to the classification process is no longer necessary for the case of NDF.

CONCLUSIONS

Our study demonstrated the feasibility of detecting seizures on iEEG using an automated detection algorithm based on the gamma frequency range and embedding nonlinear decision functions for classification purposes in a 2D decisional space. We have shown that the power measurement in the gamma range from 36 to 44 Hz contains the necessary information to discriminate seizures with a sensitivity of 96.3%, a specificity of 96.8%, and an accuracy of 96.7%. These results were obtained with a nonlinearity degree of 3 (these are polynomials of the third order), and where the two most discriminating features that constituted the 2D decisional space were determined to be: (1) the time duration (the number of consecutive points) where the value of each given point in the inter-electrode mean signal S_μ exceeded the set statistical threshold $T = \overline{S_\mu} + \sigma$, and (2) The maximum value of S_μ in that specific interval.

Of particular value is the generalized nature of the algorithm, and its feasibility in the absence of patient-specific training data. This feature is demonstrated most clearly by the test characteristics in patients who were not part of the training data. In general, it is worth mentioning that although only 22% of the files were used randomly for creating the detector, high measures in sensitivity, specificity, and accuracy were achieved. It is also interesting to note that no

cross-validation was performed during the training phase of the detector (decision function); yet, good testing results were obtained.

The computational requirements for creating the nonlinear decision functions during the training phase and the ensuing results during the testing phase reveal additional findings that are quite interesting: the data clusters of seizure files seem more spread out than those data clusters of files not leading to seizures, which clearly prove that seizures which are atypical events obviously vary greatly among subjects, and that such nonlinear decision functions are most capable of delineating such wide-ranging behaviors.

The study has so far included 14 patients who underwent two-stage epilepsy surgery with subdural recording, and whose iEEG data were obtained sequentially. We anticipate gaining more insight into the findings reported in this study as more patients in the future will consent to be included in this study. It is expected that as more data are collected, the more understanding will be gained into the clustering characteristics of epileptogenic data. Our future efforts will also be directed toward extending this approach by using the principal component analysis^{17,25,28} to establish other discriminating dimensions in order to increase the domain of analysis from 2D to 3D or even higher dimensions for more elaborate decision planes.

ACKNOWLEDGMENTS

The authors appreciate the support provided by the National Science Foundation under grants HRD-0833093, CNS-0426125, CNS-0520811, CNS-0540592, and IIS-0308155. The authors are also thankful for the clinical support provided through the Ware Foundation and the joint Neuro-Engineering Program with Miami Children's Hospital.

REFERENCES

- ¹Abend, N. S., D. Dlugos, and S. Herman. Neonatal seizure detection using multichannel display of envelope trend. *Epilepsia* 49(2):349–352, 2008.
- ²Adjouadi, M., M. Cabrerizo, M. Ayala, D. Sanchez, P. Jayakar, I. Yaylali, and A. Barreto. Detection of interictal spikes and artifactual data through orthogonal transformations. *J. Clin. Neurophysiol.* 22(1):53–64, 2005.
- ³Adjouadi, M., M. Cabrerizo, M. Ayala, D. Sanchez, P. Jayakar, I. Yaylali, and A. Barreto. Interictal spike detection using the Walsh transform. *IEEE Trans. Biomed. Eng.* 51(5):868–873, 2004.
- ⁴Albano, A. M., C. J. Cellucci, R. N. Harner, and P. E. Rapp. Optimization of embedding parameters for prediction of seizure onset with mutual information. In: *Nonlinear Dynamics and Statistics: Proceedings of an Isaac Newton Institute*, edited by A.I. MEES. Boston: Birkhauser, 2000, pp. 435–451.
- ⁵Arnhold, J., K. Lehnertz, P. Grassberger, and C. E. Elger. A robust method for detecting interdependences: Application to intracranially recorded EEG. *Physica D* 134:419, 1999.
- ⁶Bezerianos, A., S. Tong, and N. Thakor. Time-dependent entropy estimation of EEG rhythm changes following brain ischemia. *Ann. Biomed. Eng.* 31:221–232, 2003.
- ⁷Bragin, A., C. L. Wilson, T. Fields, I. Fried, and J. Engel, Jr. Analysis of seizure onset on the basis of wideband EEG recordings. *Epilepsia* 46:59–63, 2005.
- ⁸Burges, C. J. C. A tutorial on support vector machines for pattern recognition. *Data Min. Knowl. Disc.* 2(2):121–167, 1998.
- ⁹Cabrerizo, M., M. Adjouadi, M. Ayala, and M. Tito. Pattern extraction in interictal EEG recordings towards detection of electrodes leading to seizures. *Biomed. Sci. Instrum.* 42:243–248, 2006.
- ¹⁰Cabrerizo, M., M. Tito, M. Ayala, M. Adjouadi, A. Barreto, and P. Jayakar. An analysis of subdural EEG parameters for epileptic seizure evaluation. In: *CD Proceedings of the 16th International Conference on Computing CIC-2007*, Mexico City, Mexico, 2007.
- ¹¹Calvagno, G., M. Ermani, R. Rinaldo, and F. Sartoretto. Multiresolution approach to spike detection in EEG. In: *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. 6, pp. 3582–3585, 2006.
- ¹²Chander, R., E. Urrestarrazu, and J. Gotman. Automatic detection of high frequency oscillations in human intracerebral EEGs. *Epilepsia* 47(4):37, 2006.
- ¹³Cristianini, N., and J. Shawe-Taylor. *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods*. Cambridge, UK: Cambridge University Press, 2000, 208 pp.
- ¹⁴D'Alessandro, M., R. Esteller, G. Vachtsevanos, A. Hinson, J. Echaz, and B. Litt. Epileptic seizure prediction using hybrid feature selection over multiple intracranial EEG electrode contacts: a report of four patients. *IEEE Trans. Biomed. Eng.* 50(5):603–615, 2003.
- ¹⁵Frank, G. W., T. Lookman, M. A. H. Nerenberg, C. Essex, J. Lemieux, and W. Blume. Chaotic time series analyses of epileptic seizures. *Physica D* 46:427–438, 1990.
- ¹⁶Gabor, A. J. Seizure detection using a self-organizing neural network: validation and comparison with other detection strategies. *Electroencephalogr. Clin. Neurophysiol.* 107(1):27–32, 1998.
- ¹⁷Ghosh-Dastidar, S., H. Adeli, and N. Dadmehr. Principal component analysis-enhanced cosine radial basis function neural network for robust epilepsy and seizure detection. *IEEE Trans. Biomed. Eng.* 55(2–1):512–518, 2008.
- ¹⁸Good, L. B., S. Sabesan, S. T. Marsh, K. Tsakalis, L. D. Iasemidis, and D. M. Treiman. Automated seizure prediction and deep brain stimulation control in epileptic rats. *Epilepsia* 48(6):278, 2007.
- ¹⁹Gotman, J. Automatic detection of seizures and spikes. *J. Clin. Neurophysiol.* 16:130–140, 1999.
- ²⁰Gotman, J. Automatic recognition of epileptic seizures in the EEG. *Electroencephalogr. Clin. Neurophysiol.* 54:530–540, 1982.
- ²¹Guevara, M. R. Chaos in electrophysiology. In: *Concepts and Techniques in Bioelectric Measurements: Is the Medium*

- Carrying the Message? edited by A.-R. LeBlanc, and J. Billette. Montreal: Editions de l'Ecole Polytechnique de Montreal, 1997, pp. 67–87.
- ²²Hjorth, B. The physical significance of time domain descriptors in EEG analysis. *Electroencephalogr. Clin. Neurophysiol.* 34:321–325, 1970.
- ²³Iasemidis, L. D., A. Barreto, R. L. Gilmore, B. M. Uthman, S. A. Roper, and J. C. Sackellares. Spatiotemporal evolution of dynamical measures precedes onset of mesial temporal lobe seizures. *Epilepsia* 35:133–134, 1994.
- ²⁴Iasemidis, L. D., J. C. Sackellares, R. L. Gilmore, and S. N. Roper. Automated seizure prediction paradigm. *Epilepsia* 39:56, 1998.
- ²⁵Kobayashi, K., I. Merlet, and J. Gotman. Separation of spikes from background by independent component analysis with dipole modeling and comparison to intracranial recording. *Clin. Neurophysiol.* 112(3):405–413, 2001.
- ²⁶Lai, Y. C., M. A. Harrison, M. G. Frei, and I. Osorio. Inability of Lyapunov exponents to predict epileptic seizures. *Phys. Rev. Lett.* 8:91–96, 2003.
- ²⁷Larter, R., and B. Speelman. A coupled ordinary differential equation lattice model for the simulation of epileptic seizures. *Chaos* 9:795–804, 1999.
- ²⁸Levan, P., L. Tyvaert, and J. Gotman. Model-free detection of bold changes related to epileptic discharges using independent component analysis. *Epilepsia* 48(6):397–398, 2007.
- ²⁹Martinerie, J., C. Adam, and M. Le Van Ouyen. Epileptic seizures can be anticipated by non-linear analysis. *Nat. Med.* 4:1173–1176, 1998.
- ³⁰Sackellares, J.C., and L.D. Iasemidis. Seizure warning and prediction. Patent number US 6304775 B1, 2001.
- ³¹Shoeb, A., H. Edwards, J. Connolly, B. Bourgeois, T. Treves, and J. Guttag. Patient-specific seizure onset detection. In: Proceedings of the 26th Annual International Conference of the IEEE EMBS, San Francisco, CA, USA, September 1–5, 2004.
- ³²Smart, O., H. Firpi, and G. Vachtsevanos. Genetic programming of conventional features to detect seizure precursors. *Eng. Appl. Artif. Intel.* 20(8):1070–1085, 2007.
- ³³Tou, J. T., and R. C. Gonzalez. *Pattern Recognition Principles*. Reading, MA: Addison-Wesley, 1974.
- ³⁴Vapnik, V. N. *The Nature of Statistical Learning Theory*. New York: Springer, 1995.
- ³⁵Vapnik, V. N., and A. J. Chervonenkis. *Theory of Pattern Recognition*. Moscow: Nauka, 1974.