Epileptic Seizure Detection in EEGs Using Time–Frequency Analysis

Alexandros T. Tzallas, *Member, IEEE*, Markos G. Tsipouras, and Dimitrios I. Fotiadis, *Senior Member, IEEE*

***Abstract*—The detection of recorded epileptic seizure activity in EEG segments is crucial for the localization and classification of epileptic seizures. However, since seizure evolution is typically a dynamic and nonstationary process and the signals are composed of multiple frequencies, visual and conventional frequency-based methods have limited application. In this paper, we demonstrate the suitability of the time–frequency (t-f) analysis to classify EEG segments for epileptic seizures, and we compare several methods for t-f analysis of EEGs. Short-time Fourier transform and several t-f distributions are used to calculate the power spectrum density (PSD) of each segment. The analysis is performed in three stages: 1) t-f analysis and calculation of the PSD of each EEG segment; 2) feature extraction, measuring the signal segment fractional energy on specific t-f windows; and 3) classification of the EEG segment (existence of epileptic seizure or not), using artificial neural networks. The methods are evaluated using three classification problems obtained from a benchmark EEG dataset, and qualitative and quantitative results are presented.**

***Index Terms*—Artificial neural networks (ANNs), EEG, epilepsy, seizure detection, time–frequency (t-f) analysis.**

I. Introduction

APPROXIMATELY 1% of the world's population suffers from epilepsy, a disorder of the normal brain function, characterized by the existence of abnormal synchronous discharges in large ensembles of neurons in brain structures [1]. These discharges are often referred as "paroxysmal activity" and appear either during seizures (ictal periods) or between seizures (interictal periods) [2]. Epileptic seizures are manifestations of epilepsy, which are due to the sudden development of synchronous neuronal firing in the cerebral cortex and are recorded using the EEG, which is a measure of brain electrical activity. Epileptic seizures may occur in the brain locally (partial seizures), which are seen only in a few channels of the EEG recording, or involving the whole brain (generalized seizures), which are seen in every channel of the EEG recording [3].

Manuscript received December 18, 2007; revised June 19, 2008 and November 27, 2008. First published March 16, 2009; current version published September 2, 2009.

A. T. Tzallas is with the Unit of Medical Technology and Intelligent Information Systems, Department of Material Science and Technology, University of Ioannina, Ioannina 45110, Greece, and also with the Department of Medical Physics, Medical School, University of Ioannina, Ioannina 45110, Greece (e-mail: atzallas@cc.uoi.gr).

M. G. Tsipouras is with the Unit of Medical Technology and Intelligent Information Systems, Department of Material Science and Technology, University of Ioannina, Ioannina 45110, Greece (e-mail: markos@cs.uoi.gr).

D. I. Fotiadis is with the Unit of Medical Technology and Intelligent Information Systems, Department of Material Science and Technology, University of Ioannina, Ioannina 45110, Greece, and also with the Biomedical Research Institute, Foundation for Research and Technology-Hellas (FORTH), Ioannina 45110, Greece (e-mail: fotiadis@cs.uoi.gr).

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

Digital Object Identifier 10.1109/TITB.2009.2017939

Clinical neurologists in daily practice commonly examine short recordings (usually 20-min recordings) of interictal periods. The most common forms of the interictal periods are the individual or isolated spikes, the sharp wave, and the spike-and-wave complex. These are perceived in the majority of patients with epilepsy. For this reason, interictal event detection plays a vital role in the diagnosis of epilepsy. However, during an isolated spike, the brain is not in a clinical seizure. A different EEG pattern is observed during the ictal period consisting of rhythmical waveforms for a wide variety of frequencies, polyspike activity, low-amplitude desynchronization, as well as spike-and-wave complexes [3]. Although interictal findings offer evidence of epilepsy, diagnosis of epilepsy is usually based on observed epileptic seizures [4].

The interictal indications of epilepsy can be identified using a short period EEG recording. However, the long-term video-EEG monitoring is necessary due to the relatively infrequent nature of epileptic seizures. Usually, the epileptic seizures are detected either using a patient alarm button or by direct observation [4]. The development of ambulatory EEG allowed the characterization of epileptic seizures and seizure-like events at home [5].

Visual seizure detection has not been proven very efficient. Efficient automated seizure detection schemes facilitate the diagnosis of epilepsy and enhance the management of long-term EEG recordings. Since the early days of automatic analysis of epileptic EEG signals, representations based on Fourier transform and parametric methods have been applied [6]. These approaches are based on earlier observations that the epileptic seizures give rise to changes in certain frequencies bands, such as δ (0.4–4 Hz), θ (4–8 Hz), α (8–12 Hz), and β (12–30 Hz) bands. Since the nature of epileptic EEG signals is nonstationary and multicomponent, these methods are not suitable for the frequency decomposition of these signals. Time–frequency (t-f) based methods were shown to outperform conventional methods of frequency analysis [7].

In this paper, we address the use of t-f analysis for the determination of EEG segments, which contains epileptic seizures. Our approach is based on the t-f analysis of each EEG segment in order to obtain the power spectrum density (PSD) and extract features from it, which correspond to the fractional energy of windows defined on the t-f plane. Then, the features are fed into an artificial neural network (ANN) that classifies the segment as epileptic or not. Features that depict the distribution of the signal's energy over the t-f plane have not been previously applied in epileptic seizure detection. The selection of the t-f distribution (TFD) in the first step is crucial for the efficiency of the proposed approach. We report on

short-time Fourier transform (STFT) and 12 different TFDs in three common classification problems.

II. Related Work

Automatic analysis of EEG recordings in the diagnosis of epilepsy started in the early 1970s. Today, computer-based analysis addresses two major problems: 1) interictal event detection (or spike detection) and 2) epileptic seizure analysis [4]. Many algorithms for spike detection have been proposed, including mimetic- and rule-based approaches [8], frequency-domain methods [9], wavelet transforms [10], ANNs [11]–[13], independent component analysis [14], support vector machines [15], data mining [2], template matching [16], and topographic classification [16]–[18]. On the other hand, epileptic seizure analysis refers collectively to methods for: 1) epileptic seizure detection; 2) epileptic seizure prediction; and 3) automatic focus channel identification [3], [4].

Conventional temporal and frequency analysis measures have been used for the detection of epileptic seizures in EEG recordings [19]–[26]. Many works focus on the quantitative characterization of the underlying nonlinear systems based on some evidence of a deterministic value of the EEG dynamics [27], [28]. Complexity measures of the underlying EEG dynamics, such as correlation dimension [29], Lyapunov exponents [26], [30], [31], and Kolmogorov entropy [32], have been proposed. These measures can then be combined with a classifier [19], [21], [24]–[26], [30], [33] to identify the occurrence of seizures.

For seizure detection, t-f methods have also been used [34]. Schiff and coworkers considered a multiresolution analysis of the EEG [35] and pulse amplitude models (PAMs) [36]. Both examined electrocorticogram (ECoG) recordings and found qualitative seizure-related changes; however, no specific detection algorithm was posed. Williams *et al.* employed reduced interference (RI) distribution for seizure analysis [37]. They noted the presence of preseizure chirps, but did not present specific detection results. Other works reported chirp identification using scalograms and spectrograms [38], and a matched filter approach [39]. Zaveri *et al.* examined 30-s epochs of data with RI distributions and suggested the instantaneous frequency as a detection feature [40]. Other approaches based on wavelets have also been reported (e.g., [41]).

III. Methods

A. Time–Frequency Analysis

STFT and various TFDs are used for the t-f analysis [42], [43]. For STFT, the signal $x(u)$ is prewindowed around a time instant t, and the Fourier transform is calculated for each time instant t

$$\mathrm{STFT}(t,f)=\int_{-\infty}^{+\infty}x(\tau)h(\tau-t)\,e^{-if\tau}d\tau \tag{1}$$

where $h(t)$ is a short time window. STFT suffers from tradeoff between its window length and its frequency resolution. The TFDs used in our study belong to the Cohen's class of distributions. Unlike the STFT t-f representation, the Cohen's class of t-f representations is quadratic

$$\rho(t,f)=\iiint e^{i2\pi\upsilon(u-t)}g(\upsilon,\tau)x^{*}\left(u-\frac{1}{2}\tau\right)\times x\left(u+\frac{1}{2}\tau\right)e^{-i2\pi f\tau}d\upsilon du d\tau \tag{2}$$

where t is the time, f is the frequency, $x(t)$ is the signal, $x^{*}(t)$ is its complex conjugate, and $g(\upsilon,\tau)$ is an arbitrary function called kernel, which is different for each TFD. Table I presents the TFDs, which are used in our study along with the corresponding kernels. The most common TFDs that belong to the Cohen's class have been employed.

TABLE I
Time–Frequency Distributions

	Distribution	Kernel $(g(\upsilon,t))$
1	Margenau-Hill (MH)	$\cos(\pi\upsilon\tau)$
2	Wigner-Ville (WV)	1
3	Rihaczek (RIH)	$e^{i\pi\upsilon\tau}$
4	Pseudo Margenau-Hill (PMH)	$h(\tau)e^{i\pi\upsilon\tau}$ ($h(\tau)$: window function)
5	Pseudo Wigner-Ville (PWV)	$h(\tau)$ ($h(\tau)$: window function)
6	Born-Jordan (BJ)	$\sin(\pi\upsilon\tau)/(\pi\upsilon\tau)$
7	Butterworth (BUT)	$1/\left(1+\left(\frac{\upsilon}{\upsilon_1}\right)^{2N}\left(\frac{\tau}{\tau_1}\right)^{2\mathrm{M}}\right)$ $\left(N,\mathrm{M},\upsilon_1,\tau_1>0\right)$
8	Choi-Williams (CW)	$e^{-(\pi\upsilon\tau)^2/2\sigma^2}$ (σ: scaling factor)
9	Generalized rectangular (GRECT)	$\sin\left(\frac{2\pi\sigma\upsilon}{\lvert\tau\rvert^{\alpha}}\right)/(\pi\upsilon)$ (σ: scaling factor; α: dissemmetry ratio)
10	Reduced interference (RI)	$\int_{-\infty}^{+\infty}h(t)e^{-j2\pi\upsilon\tau t}dt$ ($h(t)$: hamming window)
11	Smoothed pseudo Wigner-Ville (SPWV)	$G(\upsilon)h(\tau)$ ($h(\tau)$: window function)
12	Zhao-Atlas-Marks (ZAM)	$h(\tau)\sin(\pi\upsilon\tau)/(\pi\upsilon\tau)$ ($h(\tau)$: window function)

A major problem inherit with quadratic TFDs is cross-terms (interference), which is the result of the quadratic nature of transformation. Cross-terms can make the interpretation of the t-f representation difficult. Windowing or smoothing the TFD distribution reduces the influence of the cross-terms.

Some of the TFDs employ a frequency and/or a time smoothing window: Margenau–Hill (MH), Wigner–Ville (WV), and Rihaczek (RIH) distributions do not employ smoothing windows, pseudo-MH (PMH) and pseudo-WV (PWV) employ frequency smoothing windows, while all other employ both frequency and time smoothing windows.

All time and/or frequency smoothing windows were set as Hamming 64-point length windows.

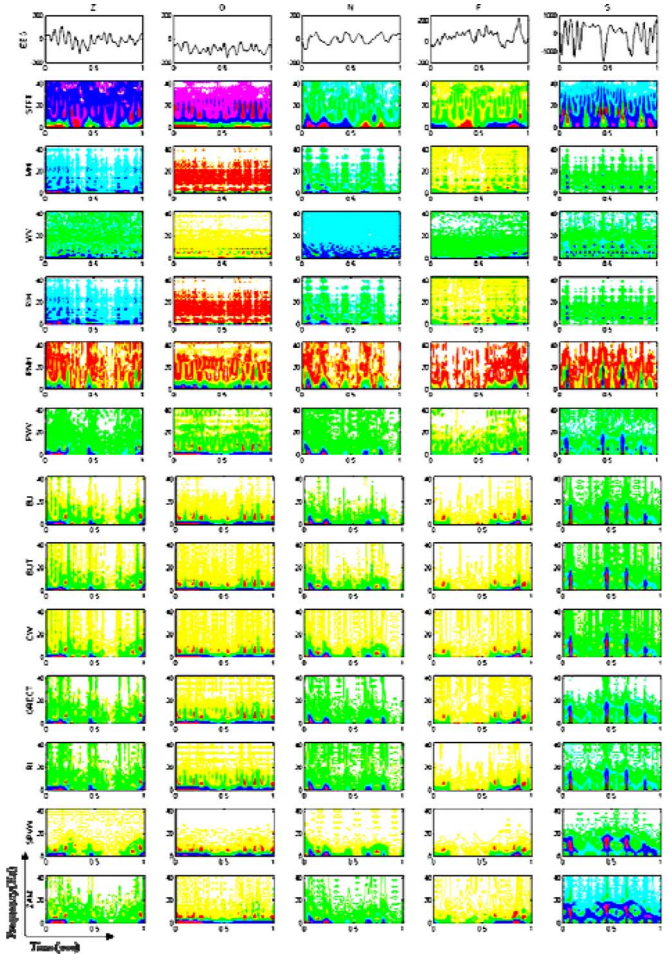

Fig. 1. PSDs calculated for all EEG segments (top figures) using STFT and the TFDs of Table I.

Using t-f analysis, the PSD of the signal is calculated, which represents the distribution of the energy of the signal over the t-f plane. In Fig. 1, the first line of figures presents EEG segments, while all other lines of figures present PSDs generated using STFT (second line of figures) and all TFDs (all other lines of figures).

B. Time–Frequency-Based Feature Extraction

The PSD that is calculated in the previous stage (2) is used to extract several features. A grid is used based on a partition both in the time and the frequency axis. In the time domain, three equal-sized windows were selected while in the frequency domain, the employed partition divided the frequency domain in five subbands; Fig. 2 presents a sample PSD with the analogous grid used for feature extraction.

The frequency subbands, which were defined based on medical knowledge on EEG, are 0–2.5 Hz, 2.5–5.5, 5.5–10.5, 10.5–21.5, and 21.5–43.5 Hz; specific features are expected to be found in certain frequency bands for the EEG segments included in the dataset. The size of the time windows was defined using expert neurologist knowledge, and it is within the range of windows selected in other works [33]. Each feature $f(i,j)$ is calculated as

$$f(i,j) = \int_{t_i}\int_{\omega_j} \mathrm{PSD}_x(t,\omega)\, d\omega dt \tag{3}$$

where PSD_x is the PSD of the signal x calculated using one of the aforementioned methods, t_i is the ith time window, and ω_j is the jth frequency band. The integral in (3) is computed as $f(i,j) = \sum_{t\in t_i}\sum_{\omega\in\omega_j} \mathrm{PSD}_x(t,\omega)$, since $\mathrm{PSD}_x(t,\omega)$ is discrete. Each feature represents the fractional energy of the signal in a specific frequency band and time window; thus, the feature set depicts the distribution of the signal's energy over the t-f plane. It is expected that the feature set carries sufficient information related to the nonstationary properties of the signal. The total energy of the signal is included as an additional feature. Therefore, each feature set is a 16-element vector ($3 \times 5 + 1$). Principal component analysis (PCA), using 1% threshold, was employed to reduce the dimension of the feature set [44] resulting in three to four features in all cases.

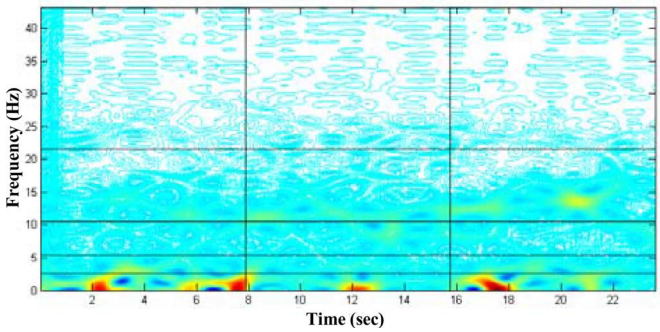

Fig. 2. Time windows and frequency subbands used for feature extraction.

C. Classification Using ANNs

The calculated features are fed into a feedforward ANN. The architecture of the ANN is N inputs (N is the size of the feature vector), one hidden layer with $5 \times N$ neurons, and K outputs (K is the number of the classes), each of them being a real number in the interval [0, 1]. The units in the hidden layer are sigmoid units with hyperbolic tangent as activation function, while the outputs are linear. Each network is trained using a standard backpropagation algorithm [45]. The architecture of the ANNs was defined heuristically, i.e., experiments related to the number of neurons in the hidden layer were conducted in order to determine the architecture with the best performance [46].

Four other classification schemes, naïve Bayes classifier, decision trees, k-nearest-neighbors (k-NNs) and logistic regression were also tested [47].

1) *Naïve Bayes classifier* is based on Bayes theory and assumes that the effect of an attribute value on a given class is independent of the values of the other attributes. This assumption is called class conditional independence.
2) *k-NN classifier* is based on learning by analogy. Given an unknown sample, it searches the pattern space neighbors that are the closest to the unknown sample. Closeness

is defined in terms of distance. The unknown sample is assigned the most common class among its neighbors.

3) *Decision trees* find explicit and understandable rules-like relationships among the input and output variables using search heuristics. Search heuristics use recursive partitioning algorithms to split the original data into finer subsets. The algorithm has to find the optimum number of splits and determine where to partition the data to maximize the information gain. The fewer the splits, the more explainable the output as there are fewer rules to understand. Decision trees are built of nodes, branches, and leaves that indicate the variables, conditions, and outcomes, respectively. The most predictive variable is placed at the top node of the tree. Specifically, the C4.5 decision tree induction algorithm was employed.
4) *Logistic regression* can be considered a special case of linear regression models. However, the binary response variable violates normality assumptions of the general regression models. A logistic regression model specifies that an appropriate function of the fitted probability of the event is a linear function of the observed values of the available explanatory variable.

IV. Results

The dataset described in [48] is used for training and evaluation of our approach. The dataset includes five subsets (denoted as Z, O, N, F, and S) each containing 100 single-channel EEG segments of 23.6 s duration. Sets Z and O have been acquired from surface EEG recordings of five healthy volunteers, with eyes open and closed, respectively. Signals in subsets F and N have been measured in seizure-free intervals, from five patients in the epileptogenic zone (F) and from the hippocampal formation of the opposite hemisphere of the brain (N). Subset S contains seizure activity, selected from all recording sites exhibiting ictal activity. Sets Z and O have been recorded extracranially, using standard electrode positioning (according to the international 10–20 system [49]), whereas sets N, F, and S have been recorded intracranially. More specifically, depth electrodes are implanted symmetrically into the hippocampal formation. EEG segments of subsets N and F were taken from all contacts of the relevant depth electrode [48]. In addition, strip electrodes are implanted onto the lateral and basal regions (middle and bottom) of the neocortex. EEG segments of the subsets S were taken from contacts of all electrodes (depth and strip). All EEG signals were recorded with the same 128-channel amplifier system using an average common reference. The data were digitized at 173.61 samples per second using 12 bit resolution, and they have the spectral bandwidth of the acquisition system, which varies from 0.5 to 85 Hz. Typical EEG segments (one from each category of the dataset) are shown in Fig. 1 (first row of figures).

In our analysis, we use the previously described dataset to create three different classification problems to evaluate our method.

1) In the first problem, two classes are examined: normal and seizure. The normal class includes only the Z-type EEG segments while the seizure class includes the S type. Thus, the dataset used for the first classification problem consists of 200 EEG segments.
2) The second classification problem includes three classes: normal, seizure-free, and seizure. The normal class includes the Z-type EEG segments; the seizure-free class includes the F-type EEG segments and the seizure class includes the S type. In the dataset of the second classification problem, 300 EEG segments are included.
3) In the third problem, all five classes are used, including all EEG segments from the previously described dataset (thus 500 EEG segments).

These different problems are examined since the medical interest is different for each one of them and are also the most widely used in the literature (one or more of them is employed in [19]–[26], [28], [30], and [32]). Therefore, we have selected them for the evaluation to be able to compare our approach with other proposed approaches.

The three classification problems, described before, are used to evaluate the proposed method. STFT and all 12 TFDs were tested for each classification problem. These result to a total of 39 different test cases. In order to avoid overfitting phenomena, cross validation was employed. Thus, for each test case, evaluation was performed using a holdout (with 50% split), ten-times random subsampling technique. Thus, for each case, ten ANNs were trained and tested; using half of the data for training (randomly selected) and the remaining for testing, ten confusion matrices were thus obtained. The size of the confusion matrix depends on the classification problem: 2×2 for the first classification problem, 3×3 for the second, and 5×5 for the third. For each of them, results for each class i are presented in terms of sensitivity (Sens_i) and selectivity(Sel_i)

$$\mathrm{Sens}_i = \frac{\#\ \text{of patterns of class } i \text{ classified in class } i}{\text{total } \#\ \text{of patterns in class } i} \quad (4)$$

$$\mathrm{Sel}_i = \frac{\#\ \text{of patterns of class } i \text{ classified in class } i}{\text{total } \#\ \text{of patterns classified in class } i}. \quad (5)$$

The results obtained in terms of sensitivity and selectivity (shown in Table II), for each classification problem and t-f analysis method, are calculated as the average value of the respective results obtained for each of the ten confusion matrices. STFT presents high Sens and Sel results for the first two problems (99.8% for both average Sens and Sel for all categories in the first classification problem, and 91.8% and 92.0% average Sens and Sel for all categories, respectively, in the second classification problem), while there is a reduction in the classification accuracy in the third classification problem, providing 65.3% average Sens and 66.2% average Sel for all categories. The MH and RIH TFDs, which do not include smoothing windows, present similar results in all classification problems, being much lower than the respective TFDs including time and/or frequency smoothing windows. The WV distribution presents results similar to more complex TFDs, without using smoothing windows. The best results for the first classification problem are obtained using the RI and smooth PWV (SPWV) distributions (100% for both average Sens and Sel), for the second using RI (100% for

TABLE II
OBTAINED SENSITIVITY (SENS) AND SELECTIVITY (SEL) RESULTS (IN PERCENT) FOR STFT AND VARIOUS TFDS FOR THE THREE CLASSIFICATION PROBLEMS

Distribution	(%)	Classification Problems												
		1			2				3					
		Classes			Classes				Classes					
		Z	S	average	Z	N	S	average	Z	O	N	F	S	average
STFT	Sens	99.8	99.8	99.8	94.0	94.2	87.2	91.8	66.2	60.0	52.6	83.8	64.0	65.3
	Sel	99.8	99.8	99.8	87.2	99.0	89.9	92.0	57.9	60.4	57.4	95.7	59.8	66.2
MH	Sens	71.6	67.6	69.6	68.8	90.0	65.6	74.8	59.6	20.6	70.2	85.6	37.0	54.6
	Sel	68.9	70.4	69.6	65.4	97.2	64.2	75.6	39.4	40.4	63.6	90.9	39.8	54.8
WV	Sens	98.4	94.2	96.3	95.6	93.8	93.4	94.3	67.2	81	87.2	89.8	92.2	83.5
	Sel	94.4	98.3	96.4	96	93.1	93.8	94.3	76	72.5	86.7	93	90.4	83.7
RIH	Sens	86.8	60.6	73.7	91.2	92.6	54.8	79.5	73.4	10.8	71.0	89.4	45.2	58.0
	Sel	68.8	82.1	75.4	66.5	97.9	80.4	81.6	40.2	41.9	71.7	93.3	46.6	58.7
PMH	Sens	94.2	96.0	95.1	95.0	100	98.0	97.7	56.8	86.0	86.0	100	92.4	84.2
	Sel	95.9	94.3	95.1	97.9	100	95.2	97.7	77.0	70.8	93.9	99.8	81.8	84.7
PWV	Sens	98.0	100	99.0	98.0	100	99.8	99.3	67.6	77.4	89.4	99.8	98.0	86.4
	Sel	100	98.0	99.0	99.8	100	98.0	99.3	76.8	72.7	97.8	99.0	86.4	86.6
BJ	Sens	96.4	99.8	98.1	97.6	100	99.4	99.0	70.8	83.2	90.4	100	97.4	88.4
	Sel	99.8	96.5	98.2	99.4	100	97.6	99.0	84.3	76.3	95.8	99.8	86.7	88.6
BUT	Sens	98.0	100	99.0	98.0	100	100	99.3	72.0	78.6	87.8	99.8	97.6	87.2
	Sel	100	98.0	99.0	100	100	98.0	99.3	79.8	75.9	97.6	98.6	84.9	87.3
CW	Sens	96.4	100	98.2	96.0	100	98.6	98.2	63.0	72.8	88.8	100	99.2	84.8
	Sel	100	96.5	98.3	98.6	100	96.1	98.2	71.0	71.7	98.7	100	82.9	84.8
GRECT	Sens	96.2	100	98.1	96.4	100	100	98.8	**76.2**	**77.6**	**93.0**	**99.8**	**97.2**	**88.8**
	Sel	100	96.3	98.2	100	100	96.5	98.8	**79.4**	**79.5**	**97.7**	**98.6**	**88.4**	**88.7**
RI	Sens	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**70.4**	**85.2**	**92.0**	**99.8**	**97.6**	**89.0**
	Sel	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**83.2**	**82.6**	**97.3**	**99.4**	**83.3**	**89.1**
SPWV	Sens	**100**	**100**	**100**	95.0	100	100	98.3	66.2	82.0	92.0	99.8	100	88.0
	Sel	**100**	**100**	**100**	100	100	95.2	98.4	80.2	77.2	100	100	83.8	88.2
ZAM	Sens	99.8	97.4	98.6	**100**	**100**	**99.6**	**99.9**	69.8	74.8	93.0	100	98.0	87.1
	Sel	97.5	99.8	98.6	**99.6**	**100**	**100**	**99.9**	74.3	74.1	100	100	87.5	87.2

both average Sens and Sel) and ZAM (99.9% for both average Sens and Sel), while in the case of the third classification problem, using the RI distribution (89% average Sens and 89.1% average Sel) and generalized rectangular (GRECT) (88.8% average Sens and 88.7% average Sel).

In addition, the classification accuracy (Acc) defined as

$$\text{Acc} = \frac{\#\ \text{of correctly classified patterns}}{\text{total}\ \#\ \text{of patterns}} \tag{6}$$

was calculated for the three classification problems and all employed t-f analysis methods. Again, the final Acc, for each classification problem and t-f analysis method, is calculated as the average of the obtained Acc for each one of the ten confusion matrices. The Acc results along with the respective standard deviation are presented in Table III. The number of features that are employed, after the application of PCA, is also presented in Table III. In addition, for each classification problem, overall results have been derived, i.e., for STFT and all TFDs, the minimum and maximum accuracies are calculated as well as the average accuracy and the standard deviation.

V. DISCUSSION

The use of different methods for t-f analysis (STFT and several TFDs) is examined in the detection of epileptic seizures in EEG recordings. Our approach is based on the t-f analysis of the EEG segments, extraction of several features from the PSD of the signal, and classification of the EEG segments using an ANN. The use of various TFDs is evaluated using three different classification problems. The comparison of several different techniques for t-f analysis addresses the problem of identifying the most appropriate technique to deal with the nonstationarity arising in the EEG signals. To our knowledge, there is no study in the literature either related to t-f analysis and feature extraction, reflecting the energy distribution over the t-f plane, or trying to identify the most appropriate t-f analysis technique, for epileptic seizure detection.

The obtained results indicate high classification ability in epileptic seizure detection. For the first and second classification problems, almost all TFDs present excellent classification accuracy results (95%–100%), except MH and RIH distributions; both of them do not employ smoothing windows, and thus, the cross-terms introduced reduce the quality of the obtained features, and subsequently, the classification accuracy. Concerning the third classification problem, the results for classification accuracy vary from 54.6% to 89%, again with the TFDs that employ smoothing windows presenting the best results (84.8%–89%). STFT also presented excellent results for the first classification problem (99.8%) and very good results for the second (91.8%), but it had a reduction in the third (65.3%), while WV distribution presents satisfactory results for all three classification problems. TFDs employing both time and frequency smoothing windows indicate the highest performance: 98.8%, 99%, and 87.6% average accuracy for the three classification problems, respectively. A comparison of the obtained

TABLE III
OBTAINED ACCURACY (IN PERCENT) (STANDARD DEVIATION)/NUMBER OF FEATURES (AFTER PCA APPLICATION) FOR STFT AND VARIOUS TFDS FOR THE THREE CLASSIFICATION PROBLEMS

	Classification problems		
Distribution	**1**	**2**	**3**
Short time Fourier transform	99.8 (0.42) / 4	91.8 (1) / 4	65.32 (0.89) / 4
Margenau-Hill	69.6 (1.78) / 3	74.8 (2.13) / 3	54.6 (0.91) / 3
Wigner-Ville	96.3 (0.82) / 3	94.3 (0.29) / 3	82.6 (1.07) / 3
Rihaczek	73.7 (1.64) / 3	79.53 (1.04) / 3	57.96 (0.51) / 3
Pseudo Margenau-Hill	95.1 (0.32) / 4	97.67 (0.35) / 4	84.24 (0.57) / 4
Pseudo Wigner-Ville	99 (0.37) / 3	99.27 (0.21) / 3	86.44 (0.55) / 3
Born-Jordan	98.1 (0.32) / 3	99 (0.35) / 3	88.36 (0.3) / 3
Butterworth	99 (0.33) / 3	99.33 (0.27) / 3	87.16 (0.67) / 3
Choi-Williams	98.2 (0.63) / 3	98.2 (0.45) / 3	84.76 (0.58) / 3
Generalized rectangular	98.1 (0.32) / 3	98.8 (0.28) / 3	88.76 (0.69) / 3
Reduced interference	**100 (0)** / 3	**100 (0)** / 3	**89 (0.54)** / 3
Smoothed pseudo Wigner-Ville	100 (0) / 3	98.33 (0.65) / 3	88 (0.27) / 3
Zhao-Atlas-Marks	98.6 (0.7) / 3	99.87 (0.28) / 3	87.12 (0.7) / 3
Maximum	100	100	89
Minimum	69.6	74.8	54.6
Average	94.27	94.68	80.33
Standard deviation	10.17	8.17	12.34

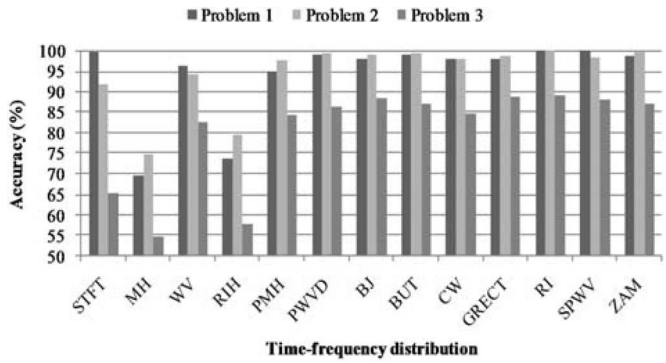

Fig. 3. Accuracy results for STFT and the TFDs shown in Table I for the three classification problems.

TABLE IV
COMPARISON OF CLASSIFICATION ACCURACY (IN PERCENT) OBTAINED BY FIVE CLASSIFIERS USING THE SET OF FEATURES OBTAINED BY THE RI DISTRIBUTION

	Classification Problems		
Classifiers	**1**	**2**	**3**
Naïve Bayes classifier	88.9	79.47	68.96
5-NN	92.8	83.8	75.88
Decision tree (C4.5)	94	85.53	74.08
Logistic regression	95.6	90.73	75.44
ANN	**100**	**100**	**89**

accuracy results for all classification problems and t-f analysis methods is presented in Fig. 3. Statistical analysis (two-tailed t-test) was performed between the accuracy obtained from the STFT and the accuracy obtained using the reduced interference (RID) for all classification problems. The results indicated a statistical significant difference for the second and third classification problems but not for the first one.

We have compared the results obtained using ANNs with the results obtained by four widely used classification techniques: naïve Bayes classifier, decision trees, k-NNs, and logistic regression. Decision trees were implemented using the C4.5 algorithm. Postpruning was employed using the pessimistic error-rate-based method (subtree replacement). The confidence factor for pruning was set to 0.25 and the minimum number of instances in a leaf was 2. The k-NN classifier was implemented with $k = 5$ (5-NN). Table IV presents the average accuracy of the ten runs for the features obtained using the RI distribution for all five classifiers. The use of ANNs produces the best results; however, it should be mentioned that ANNs present the highest computational complexity and training time while the 5-NN classifier presents the highest evaluation time.

In Fig. 1, the obtained PSDs for five EEG segments are presented. PSDs obtained from the MH and RIH distributions do not accurately reflect the rhythm changes in the respective signal segments. Especially in the segment belonging to the S category, main frequency changes are vaguely depicted. The same applies to PSDs obtained from the STFT. In WV distribution, the PSD of the S segment has several components corresponding to frequency changes while PSDs obtained from the TFDs employing time and/or frequency smoothing windows more clearly reflect these frequency changes. Especially for TFD employing both time and frequency smoothing windows, the obtained PSDs are quite similar. However, the presence of cross-terms is obvious in several cases, such as in PMH and PWV distributions (sixth and seventh lines of figures in Fig. 1). Thus, the selection of the appropriate TFD is crucial for the improvement of our approach and this depends on the classification problem under discussion. However, TFDs that employ smoothing windows tend to generate feature sets that are able to discriminate among the EEG categories more accurately.

In this study, high-frequency components (over 43.5 Hz) were not measured. The accuracy achieved for all TFDs for the epileptic seizure detection is more than satisfactory and its automated nature also makes it suitable to be used in real clinical conditions. Besides the feasibility of a real-time implementation of the proposed approach, diagnosis can be made more accurate

TABLE V
COMPARISON OF CLASSIFICATION ACCURACY (IN PERCENT) OBTAINED BY OUR APPROACH FOR THE DETECTION OF EPILEPTIC SEIZURES COMPARED TO THE CLASSIFICATION ACCURACIES (IN PERCENT) OBTAINED BY OTHER RESEARCHERS

Authors	Method	Dataset	Accuracy
Nigam et al. [20]	Nonlinear pre-processing filter-Diagnostic neural network	Z, S	97.2
Srinivasan et al. [19]	Time & frequency domain features-Recurrent neural network	Z, S	99.6
Kannathal et al. [32]	Entropy measures-Adaptive neuro-fuzzy inference system	Z, S	92.22
Kannathal et al. [28]	Chaotic measures-Surrogate data analysis	Z, S	~90
Polat et al. [21]	Fast Fourier transform-Decision tree	Z, S	98.72
Subasi [23]	Discrete wavelet transform-Mixture of expert model	Z, S	95
This work	**Time frequency analysis-Artificial neural network**	**Z, S**	**100**
Güler et al. [30]	Lyapunov exponents-Recurrent neural network	Z, F, S	96,79
Sadati et al. [24]	Discrete wavelet transform–Adaptive neural fuzzy network	Z, F, S	85,9
This work	**Time frequency analysis-Artificial neural network**	**Z, F, S**	**100**
Güler et al. [25]	Wavelet transform-Adaptive neuro-fuzzy inference system	Z, O, N, F, S	98.68
Güler et al. [26]	Wavelet transform, Lyapunov exponents-Support vector machine	Z, O, N, F, S	99.28
Übeyli et al. [22]	Eigenvector methods–Modified of Mixture of expert model	Z, O, N, F, S	98.60
This work	**Time frequency analysis-Artificial neural network**	**Z, O, N, F, S**	**89**

by increasing the number of parameters. A system that may be developed as a result of this study may provide feedback to the experts for the classification of the EEG signals quickly and accurately.

Table V presents a comparison between our approach and other methods proposed in the literature. Only methods that are evaluated using the same dataset are included. The results presented in this paper are the ones obtained using the RI TFD. For the two-class problem, the results obtained from the evaluation of our method are the best presented for this dataset. The difference between our results and all other results proposed in the literature varies from 0.4% to 10%. Concerning the three-class problem, the results obtained from our approach are again the best presented for this dataset, indicating an improvement from 3.21% to 14.1% from other approaches proposed in the literature. However, in the third classification problem, our results are not satisfactory, being almost 89%. This is mainly due to the high misclassification rates between sets Z and O. It should be mentioned that this does not have a major impact on the study, since both these sets are obtained from healthy individuals (with eyes open and closed, respectively). Still, the results demonstrate a limitation of the proposed methodology, since the obtained sets of features do not carry sufficient information to correctly classify these sets. The best reported result for this dataset is 99.28% [31].

VI. CONCLUSION

We presented a comparison of using STFT and other 12 well-known TFDs to access the nonstationary properties of the EEG signal with respect to epileptic seizure detection. We utilized an approach based on t-f analysis and extraction of features reflecting the distribution of the signal's energy over the t-f plane. Both of these aspects have not been employed for epileptic seizure detection, while the results obtained using a publicly available dataset demonstrate the added value of the proposed approach. Future work will focus on the employment of high-frequency components, such as gamma activity (30–60 Hz), and their importance concerning epileptic seizure detection. Also, alternative techniques (such as autoregressive models) and $t-f$ partitions for feature extraction must be examined. Finally, an extension of the proposed method in order to automatically detect epileptic seizures in long-term EEG recordings, and subsequently, classify them into different epileptic categories will be of high clinical value.

ACKNOWLEDGMENT

The authors are grateful to Prof. S. Konitsiotis for his advice in several issues related to the medical knowledge.

REFERENCES

[1] F. Mormann, R. G. Andrzejak, C. E. Elger, and K. Lenhnertz, "Seizure prediction: The long and the winding road," *Brain*, vol. 130, no. 2, pp. 314–333, 2007.

[2] T. P. Exarchos, A. T. Tzallas, D. I. Fotiadis, S. Konitsiotis, and S. Giannopoulos, "EEG transient event detection and classification using association rules," *IEEE Trans. Inf. Technol. Biomed.*, vol. 10, no. 3, pp. 451–457, Jul. 2006.

[3] J. Gotman, "Automatic detection of seizures and spikes," *J. Clin. Neurophysiol.*, vol. 16, no. 2, pp. 130–140, 1999.

[4] N. McGrogan, "Neural network detection of epileptic seizures in the electroencephalogram," Ph.D. dissertation, Oxford Univ., Oxford, U.K., Feb. 1999.

[5] E. Waterhouse, "New horizons in ambulatory electroencephalography," *IEEE Eng. Med. Biol. Mag.*, vol. 22, no. 3, pp. 74–80, May/Jun. 2003.

[6] A. Subasi and E. Erçelebi, "Classification of EEG signals using neural network and logistic regression," *Comput. Methods Programs Biomed.*, vol. 78, no. 2, pp. 87–99, 2005.

[7] B. Boashash, M. Mesbah, and P. Golditz, *Time-Frequency Detection of EEG Abnormalities*. Amsterdam, The Netherlands: Elsevier, pp. 663–669, 2003, ch. 15 (article 15.5).

[8] S.B. Wilson and R. Emerson, "Spike detection: A review and comparison of algorithms," *Clin. Neurophysiol.*, vol. 113, pp. 1873–1881, 2002.

[9] J. Gotman and P. Gloor, "Automatic recognition and quantification of interictal epileptic activity in the human scalp EEG," *Electroencephalogr. Clin. Neurophysiol.*, vol. 41, pp. 513–529, 1976.

[10] M. Dumpelmann and C. E. Elger, "Automatic detection of epileptiform spikes in the electrocorticogram: A comparison of two algorithms," *Seizure*, vol. 7, pp. 145–152, 1998.

[11] S. B. Wilson, C. A. Turner, R. G. Emerson, and M. L. Scheuer, "Spike detection II: Automatic, perception-based detection and clustering," *Clin. Neurophysiol.*, vol. 110, pp. 404–411, 1999.

[12] A. T. Tzallas, P. S. Karvelis, C. D. Katsis, D. I. Fotiadis, S. Giannopoulos, and S. Konitsiotis, "A method for classification of transient events in EEG recordings: Application to epilepsy diagnosis," *Methods Inf. Med.*, vol. 45, no. 6, pp. 610–621, 2006.

[13] W. R. Webber, B. Litt, K. Wilson, and R. P. Lesser, "Practical detection of epileptiform discharges (EDs) in the EEG using an artificial neural

network: A comparison of raw and parameterized EEG data," *Electroencephalogr. Clin. Neurophysiol.*, vol. 91, pp. 194–204, 1994.
[14] K. Kobayashi, I. Merlet, and J. Gotman, "Separation of spikes from background by independent component analysis with dipole modeling and comparison to intracranial recording," *Clin. Neurophysiol.*, vol. 112, pp. 405–413, 2001.
[15] N. Acir and C. Guzelis, "Automatic spike detection in EEG by a two-stage procedure based on support vector machines," *Comput. Biol. Med.*, vol. 34, pp. 561–575, 2004.
[16] R. Sankar and J. Natour, "Automatic computer analysis of transients in EEG," *Comput. Biol. Med.*, vol. 22, pp. 407–422, 1992.
[17] M. Feucht, K. Hoffmann, K. Steinberger, H. Witte, F. Benninger, M. Arnold, and A. Doering, "Simultaneous spike detection and topographic classification in pediatric surface EEGs," *Neuroreport*, vol. 8, pp. 2193–2197, 1997.
[18] M. Adjouadi, D. Sanchez, M. Cabrerizo, M. Ayala, P. Jayakar, I. Yaylali, and A. Barreto, "Interictal spike detection using the Walsh transform," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 5, pp. 868–872, May 2004.
[19] V. Srinivasan, C. Eswaran, and N. Sriraam, "Artificial neural network based epileptic detection using time-domain and frequency domain features," *J. Med. Syst.*, vol. 29, no. 6, pp. 647–660, 2005.
[20] V. P. Nigam and D. Graupe, "A neural-network-based detection of epilepsy," *Neurol. Res.*, vol. 26, no. 6, pp. 55–60, 2004.
[21] K. Polat and S. Güneş, "Classification of epileptiform EEG using a hybrid system based on decision tree classifier and fast Fourier transform," *Appl. Math. Comput.*, vol. 32, no. 2, pp. 625–631, 2007.
[22] E. D. Übeyli and İ. Güler, "Features extracted by eigenvector methods for detecting variability of EEG signals," *Pattern Recognit. Lett.*, vol. 28, no. 5, pp. 592–603, 2007.
[23] A. Subasi, "Signal classification using wavelet feature extraction and a mixture of expert model," *Exp. Syst. Appl.*, vol. 32, no. 4, pp. 1084–1093, 2007.
[24] N. Sadati, H. R. Mohseni, and A. Magshoudi, "Epileptic seizure detection using neural fuzzy networks," in *Proc. IEEE Int. Conf. Fuzzy Syst.*, Jul.16–21, 2006, pp. 596–600.
[25] İ. Güler and E. D. Übeyli, "Adaptive neuro-fuzzy inference system for classification of EEG signals using wavelet coefficients," *J. Neurosci. Methods*, vol. 148, no. 2, pp. 113–121, 2005.
[26] İ. Güler and E. D. Übeyli, "Multiclass support vector machines for EEG signals classification," *IEEE Trans. Inf. Technol. Biomed.*, vol. 11, no. 2, pp. 117–126, Mar. 2007.
[27] L. D. Iasemidis and J. C. Sackellares, "Chaos theory and epilepsy," *Neuroscience*, vol. 2, pp. 118–126, 1996.
[28] N. Kannathal, U. R. Acharya, C. M. Lim, and P. K. Sadasivan, "Characterization of EEG—A comparative study," *Comput. Methods Prog. Biomed.*, vol. 80, no. 1, pp. 17–23, 2005.
[29] D. E. Lerner, "Monitoring changing dynamics with correlation integrals: Case study of an epileptic seizure," *Phys. D*, vol. 97, no. 4, pp. 563–576, 1996.
[30] N. F. Güler, E. D. Übeyli, and İ. Güler, "Recurrent neural networks employing Lyapunov exponents for EEG signals classification," *Exp. Syst. Appl.*, vol. 29, no. 3, pp. 506–514, 2005.
[31] V. Srinivasan, C. Eswaran, and N. Sriraam, "Approximate entropy-based epileptic EEG detection using artificial neural networks," *IEEE Trans. Inf. Technol. Biomed.*, vol. 11, no. 3, pp. 288–295, May 2007.
[32] N. Kannathal, M. L. Choo, U. R. Acharya, and P. K. Sadasivan, "Entropies for detection of epilepsy in EEG," *Comput. Methods Prog. Biomed.*, vol. 80, no. 3, pp. 187–194, 2005.
[33] H. Qu and J. Gotman, "A patient-specific algorithm for the detection of seizure onset in long-term EEG monitoring: Possible use as a warning device," *IEEE Trans. Biomed. Eng.*, vol. 44, no. 2, pp. 115–122, Feb. 1997.
[34] H. P. Zaveri, W. J. Williams, and J. C. Sackellares, "Energy based detection of seizures," in *Proc. 15th Ann. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Oct. 28–31, 1993, pp. 363–364.
[35] S. J. Schiff, J. Heller, S. R. Weinstein, and J. G. Milton, "Controlled wavelet transforms for EEG spike and seizure localization," *Proc. SPIE*, vol. 2242, pp. 762–775, Mar. 1994.
[36] E. S. Hughes, W. D. Gaillard, G. M. Jacyna, and S. J. Schiff, "Feature-based detection of epileptic seizures using ECoG recordings and a PAM," in *Proc. 17th Ann. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Sep. 20–23, 1995, vol. 2, pp. 933–934.
[37] W. J. Williams, H. P. Zaveri, and J. C. Sackellares, "Time-frequency analysis of electrophysiology signals in epilepsy," *IEEE Eng. Med. Biol.*, vol. 14, no. 2, pp. 133–143, Mar./Apr. 1995.
[38] J. J. Benedetto and D. Colella, "Wavelet analysis of spectrogram seizure chirps," *Proc. SPIE*, vol. 2569, pp. 512–521, Sep. 1995.
[39] J. W. Creekmore, J. Conry, D. Colella, and S. J. Schiff, "Matched-filter based detection and localization of epileptic seizures from spectrographic chirps in ECoG recordings," in *Proc. 17th Ann. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Sep. 20–23, 1995, vol. 2, pp. 891–892.
[40] H. P. Zaveri, R. B. Duckrow, N. C. de Lanerolle, and S. S. Spencer, "Distinguishing subtypes of temporal lobe epilepsy with background hippocampal activity," *Epilepsia*, vol. 42, pp. 725–730, 2001.
[41] M. Sun, L. S. Pon, M. L. Scheurer, and R. J. Sclabassi, "Application of time-frequency and wavelet analysis to the diagnosis of epilepsy," *Proc. SPIE*, vol. 4056, pp. 210–219, 2000.
[42] L. Cohen, "Time-frequency distributions-a review," *Proc. IEEE*, vol. 77, no. 7, pp. 941–981, Jul. 1989.
[43] F. Auger, P. Flandrin, P. Goncalves, and O. Lemoine, *Time-Frequency Toolbox Tutorial*. Paris, France/Houston, TX: CNRS/RICE Univ., 1995–1996.
[44] I. H. Witten and E. Frank, *Data Mining: Practical Machine Learning Tools and Techniques*. San Fransisco, CA: Morgan Kaufmann, 2005.
[45] C. M. Bishop, *Neural Networks for Pattern Recognition*. New York: Oxford Univ. Pres, 1995.
[46] H. Demuth and M. Beale, *Neural Network Toolbox User's Guide for Use With MATLAB*. Natick, MA: The Math Works, Inc., 2002.
[47] M. Kantardzic, *Data Mining Concepts, Models, Methods, and Algorithms*. Hoboken, NJ: Wiley–IEEE Press, 2003.
[48] R. G. Andrzejak, K. Lehnertz, F. Mormann, C. Rieke, P. David, and C. E. Elger, "Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state," *Phys. Rev. E*, vol. 64, no. 6, pp. 061907-1–061907-8, 2001.
[49] H. H. Jasper, "Ten-twenty electrode system of the international federation," *Electroencephalogr. Clin. Neurophysiol.*, vol. 10, pp. 371–375, 1958.

Alexandros T. Tzallas (M'09) was born in Athens, Greece, in 1977. He received the B.S. degree in physics in 2001 from the University of Ioannina, Ioannina, Greece, where he is currently working toward the Ph.D. degree in medical physics.

His current research interests include biomedical signal processing, analysis of EEGs, epilepsy diagnosis, and biomedical applications.

Markos G. Tsipouras was born in Athens, Greece, in 1977. He received the Diploma and the M.Sc. degree in computer science in 1999 and 2002, respectively, from the University of Ioannina, Ioannina, Greece, where he is currently working toward the Ph.D. degree in the Department of Computer Science.

His current research interests include biomedical engineering, decision support and medical expert systems, and biomedical applications.

Dimitrios I. Fotiadis (M'01–SM'07) was born in Ioannina, Greece, in 1961. He received the Diploma in chemical engineering from the National Technical University of Athens, Athens, Greece, in 1985, and the Ph.D. degree in chemical engineering from the University of Minnesota, Minneapolis, in 1990.

From 1995 to early 2008, he was with the Department of Computer Science, University of Ioannina, Ioannina, Greece, where he is currently a Professor in the Department of Materials Science and Technology, the Director of the Unit of Medical Technology and Intelligent Information Systems, and the President of the Science and Technology Park of Epirus. He is also with the Biomedical Research Institute, Foundation for Research and Technology-Hellas (FORTH), Ioannina. His current research interests include biomedical technology, biomechanics, scientific computing, and intelligent information systems.