frontiers in Computational Neuroscience

ORIGINAL RESEARCH
published: 24 March 2015
doi: 10.3389/fncom.2015.00038

Detection of epileptiform activity in EEG signals based on time-frequency and non-linear analysis

Dragoljub Gajic[1,2], Zeljko Djurovic[1], Jovan Gligorijevic[3], Stefano Di Gennaro[2] and Ivana Savic-Gajic[4]*

[1] Department of Signals and Systems, School of Electrical Engineering, University of Belgrade, Belgrade, Serbia, [2] Center of Excellence DEWS, University of L'Aquila, L'Aquila, Italy, [3] Faculty of Engineering, University of Kragujevac, Kragujevac, Serbia, [4] Faculty of Technology, University of Nis, Leskovac, Serbia

We present a new technique for detection of epileptiform activity in EEG signals. After preprocessing of EEG signals we extract representative features in time, frequency and time-frequency domain as well as using non-linear analysis. The features are extracted in a few frequency sub-bands of clinical interest since these sub-bands showed much better discriminatory characteristics compared with the whole frequency band. Then we optimally reduce the dimension of feature space to two using scatter matrices. A decision about the presence of epileptiform activity in EEG signals is made by quadratic classifiers designed in the reduced two-dimensional feature space. The accuracy of the technique was tested on three sets of electroencephalographic (EEG) signals recorded at the University Hospital Bonn: surface EEG signals from healthy volunteers, intracranial EEG signals from the epilepsy patients during the seizure free interval from within the seizure focus and intracranial EEG signals of epileptic seizures also from within the seizure focus. An overall detection accuracy of 98.7% was achieved.

Keywords: seizure detection, epileptiform activity, non-linear analysis, scatter matrices, quadratic classifiers

OPEN ACCESS

Edited by:
Tobias Alecio Mattei, Brain and Spine Center - InvisionHealth - Kenmore Mercy Hospital, USA

Reviewed by:
Peter König, University of Osnabrück, Germany
Germán Mato, Centro Atomico Bariloche, Argentina

****Correspondence:***
Dragoljub Gajic, Department of Signals and Systems, School of Electrical Engineering, University of Belgrade, Bulevar Kralja Aleksandra 73, Belgrade 11000, Serbia
dragoljubgajic@gmail.com

Received: *01 January 2015*
Accepted: *08 March 2015*
Published: *24 March 2015*

Citation:
Gajic D, Djurovic Z, Gligorijevic J, Di Gennaro S and Savic-Gajic I (2015) Detection of epileptiform activity in EEG signals based on time-frequency and non-linear analysis. Front. Comput. Neurosci. 9:38. doi: 10.3389/fncom.2015.00038

Introduction

According to the estimations of the World Health Organization around 50 million people worldwide suffer from epilepsy as the most common disorder of the brain activity (World Health Organization, 2012). It is characterized by sudden and recurrent seizures which are the result of an excessive and synchronous electrical discharge of a large number of neurons. Epileptic seizures can be divided by their clinical manifestation into two main classes, partial and generalized (Tzallas et al., 2007). Partial or focal epileptic seizures involve only a circumscribed region of the brain (epileptic focus) and remain restricted to this region while generalized epileptic seizures involve almost the entire brain. Both classes of epileptic seizures can occur at all ages. An epileptiform activity in EEG signals including spikes, sharp waves, or spike-and-wave complexes can be evident not only during a seizure (the ictal period) but also a short time before (the preictal period) as well as between seizures (the interictal period). Consequently, EEG signals have been the most utilized in clinical assessments of the brain state including both prediction and detection of epileptic seizures (Waterhouse, 2003; Casson et al., 2010). However, the detection of epileptiform activity in EEG signals by visual scanning of EEG recordings usually collected over a few days is a tedious and time-consuming process. In addition, it requires a team of experts to analyze the entire length of the EEG recordings in order to detect epileptiform activity. A reliable technique for detection

of epileptiform activity in EEG signals would ensure an objective and facilitating treatment of patients and thus improve the diagnosis of epilepsy. Furthermore, it would also enable an automated prediction and/or detection of epileptic seizures in real time by a system to be implanted in head of epileptic patients (Jerger et al., 2001). Such a system would significantly improve quality of life of people suffering from epilepsy. Most of the techniques for automated detection of epileptiform activity that have emerged in recent years consist of two key successive steps: extraction of features from EEG signals and then classification of the extracted features for detection of epileptiform activity.

The feature extraction, as the first step, has a direct influence on both precision and complexity of the entire technique. Most common statistical features in time domain, such as the mean, the variance, the coefficient of variation and the total variation, by themselves are not sufficient for a reliable detection of epileptiform activity, and thus are mostly used as statistical measures for features in other domains. The variance and the total variation are considered to have better discriminatory capabilities than the mean, since they are able to detect magnitude of change in a signal over time. Even though we can note a certain periodicity and synchronization between EEG signals from different electrodes, neither the autocorrelation nor the cross-correlation have proved to be reliable features for detection of epileptiform activity. This is especially true in the case of the cortical EEG where the recording electrodes are so close to each other that the synchronization could be noted even when there was no seizure. However, in the literature we can still find several applications of these two features (Niederhauser et al., 2003; Jerger et al., 2005).

Unlike the previous features, the spectral features of EEG signals obtained through the Fourier transform have found wide applications in the field (Polat and Gunes, 2007; Mousavi et al., 2008). Namely, all the research carried out to date clearly indicates that it is much better to identify and extract the features of interest in frequency domain than in time domain, even though the both domains contain identical information. The analysis in time-frequency domain gives even better results considering that it contains, in addition to frequency, also the temporal component of signal which is lost during the Fourier transform. The literature mainly contains techniques based on wavelet transform (Subasi, 2007a,b; Wang et al., 2011; Gajić et al., 2014) which has also been used in the research related to other brain disorders, such as schizophrenia (Hazarika et al., 1997) and Alzheimer's disease (Adeli and Ghosh-Dastidar, 2010). The detection of epileptiform activity based on non-linear analysis, i.e., extraction of the correlational dimension and the Lyapunov exponents as non-linear features can also be noted in some research studies (Iasemidis et al., 2003; Srinivasan et al., 2007; Adeli and Ghosh-Dastidar, 2010).

A precise classification as the second key step directly depends on the previously extracted features. That is, there is no classifier which could in any way make up for the shortcomings which are consequence of the information lost during the feature extraction. Like in the case of the feature extraction, we can come across a very wide range of classifiers starting from the most simple ones with thresholds (Altunay et al., 2010) or rule-based (Gotman, 1999), to linear classifiers (Liang et al., 2010; Iscan et al., 2011) and all the way to those more complex ones based on fuzzy logic and artificial neural networks (Gajić, 2007; Subasi, 2007a; Tzallas et al., 2007). We can also note the use of other techniques for classification based on k nearest neighbors (Guo et al., 2011; Orhan et al., 2011), decision trees (Tzallas et al., 2009), expert models (Ubeyli, 2007; Ubeyli and Guler, 2007) as well as Bayes classifiers (Tzallas et al., 2009; Iscan et al., 2011). Considering that the feature extraction as a process of higher priority can be computationally very demanding it is always more desirable to use simpler classifiers so that the entire decision-making system could ideally work in real time.

In this paper we present an automated technique for detection of epileptiform activity in EEG signals. In contrast with the existing techniques which are mainly based on features from one domain of interest, our new technique optimally integrates features from a few domains and frequency sub-bands of clinical interest in order to increase its robustness and accuracy. We extract features in both time and frequency domain as well as time-frequency domain using discrete wavelet transform which has already been recognized as a very good linear technique for analysis of non-stationary signals such as EEG signals. In addition, by non-linear analysis we extract the correlation dimension and the largest Lyapunov exponent as much better measures of EEG signal non-linearity which is only approximated by other linear techniques such as fast Fourier transform (FFT) and discrete wavelet transform (DWT). After the feature extraction we optimally reduce the feature space dimension to two using scatter matrices and then perform classification in the reduced feature space by quadratic classifiers which have already been known as very robust solutions for classification of random feature vectors.

Materials and Methods

Materials

The EEG signals used to design and test the new technique were recorded at the University Hospital Bonn, Germany with the same 128-channel amplifier system (Andrzejak et al., 2001). After 12 bit analog-to-digital conversion the EEG signals were saved in a data acquisition system at a sampling rate of 173.61 Hz. The amplifier range was adjusted well so that the recordings could be made with 12 bits. The recorded EEG signals were further passed through a low pass filter with the finite impulse response and bandwidth of 0–60 Hz. The frequencies higher than 60 Hz mostly present noise and are a very small part of the signal total energy in the frequency band up to 86.8 Hz saved by the acquisition system. We used 100 segments of epileptic and 200 segments of non-epileptic EEG signals to design and test our new technique. The epileptic EEG signals were recorded using cortical electrodes from 5 epileptic patients during seizure from within the seizure focus, i.e., the region of unhealthy brain tissue that was later removed by surgery. The first 100 segments of non-epileptic EEG signals were also recorded using cortical electrodes from the same epileptic patients and the same unhealthy brain tissue but during seizure-free interval. The remaining 100 segments of non-epileptic EEG signals were recorded using scalp electrodes from 5 healthy volunteers and of course their healthy brain tissue.

So, there was a total of three groups with 100 segments of the EEG signals. All the segments have duration of 4096 samples, i.e., 23.6 s, and were additionally tested on the weak stationarity (Andrzejak et al., 2001) in order to perform non-linear analysis. Since the EEG signals were recorded from different patients and with different electrodes, all extracted EEG signal segments were also additionally normalized in order to have the same zero mean and unit variance as shown in **Figure 1**. In this way, we wanted to design a detection technique that is not dependent on patient and the EEG recording system either.

Methods

There are five broad sub-bands of the EEG signal which are generally of clinical interest: delta (0–4 Hz), theta (4–8 Hz), alpha (8–16 Hz), beta (16–32 Hz), and gamma waves (32–64 Hz). Higher frequencies are often more common in abnormal brain states such as epilepsy, i.e., there is a shift of EEG signal energy from lower to higher frequency bands before and during a seizure (Gajić et al., 2014). These five frequency sub-bands provide more accurate information about neuronal activities underlying the problem. Consequently, some changes in the EEG signal, which are not so obvious in the original full-spectrum signal, can be amplified when each sub-band is considered independently. Thus, we extract features from each sub-band separately and also in time, frequency and time-frequency domain as well as by non-linear analysis. After the feature extraction we reduce dimension of the feature space to two. Finally, two quadratic classifiers able to separate all three groups of the EEG signals from each other are designed. The entire structure of the technique is shown in **Figure 2**.

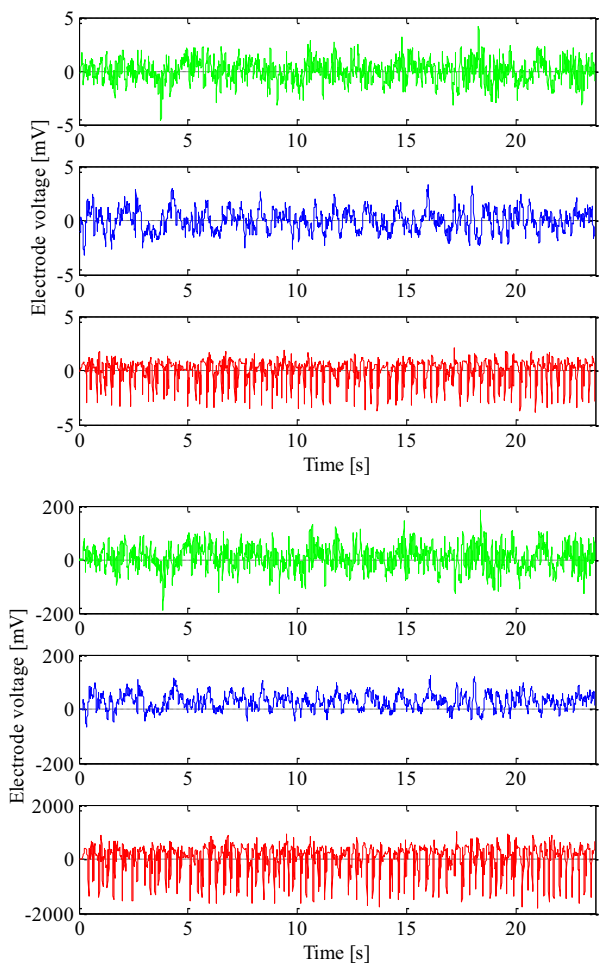

FIGURE 1 | Non-normalized (lower) and normalized (upper) epileptic (in red) and non-epileptic (unhealthy in blue and healthy tissue in green) EEG signals.

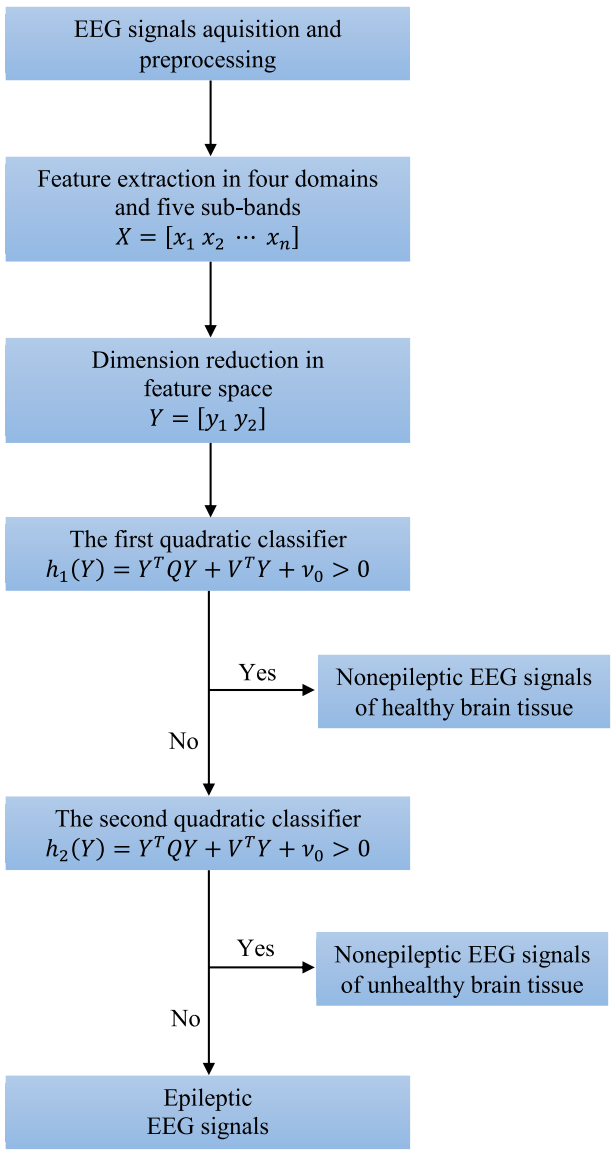

FIGURE 2 | Structure of the new technique consisting of four key steps: preprocessing, feature extraction, dimension reduction, and classification.

Time-Frequency Domain Analysis

Since the segments of the EEG signals have already been normalized and all have zero mean and unit variance, additional extraction of these two features as well as coefficient of variation as function of mean value and variance, does not make any sense. However, we extracted the total variation as another measure of signal variability in the time domain even after normalization since it counts number of signal sign changes or signal polarity. In the case of a signal segment $x[n]$ of N samples, i.e., $n = 1,\ 2 \cdots N$, the total variation is given by:

$$v_x = \frac{1}{N-1} \frac{\sum_{n=2}^{N} |x[n] - x[n-1]|}{(max_x - min_x)} \tag{1}$$

where the signal is essentially normalized by the difference between its maximum and minimum values in the segment of interest. Obviously, the value of the total variation is located in the range between $1/(N-1)$ for slower signals and 1 for signals with very high and frequent changes.

EEG signals, as the outcome of events with different repetition periods, contain signals whose different frequencies cannot be identified in the time domain, since all these signals are shown together. Thus, signal transformation from the time domain to the frequency domain is necessary, which in the case of a signal segment $x[n]$ of N samples is achieved using the fast Fourier transform (FFT) defined by:

$$\mathit{fft}[\omega] = \sum_{n=1}^{N} x[n] e^{-i\omega n},\ \omega = \frac{2\pi m}{N},\ 0 \le m \le N-1 \tag{2}$$

where $\omega = 2\pi f/f_s$ represents the angular frequency discretized in N samples (Proakis and Manolakis, 1996). In order to avoid discontinuities between the end and beginning of the segments and thus spurious spectral frequency components the beginning of each segment was chosen in such a way that the amplitude difference of the last and first data points was within the range of amplitude differences of consecutive data points, and the slopes at the end and beginning of each segment had the same sign. This procedure reduces edge effects that result in spectral leakage in the FFT spectrum. In order to further minimize spectral leakage windowing of signal segments by the Hamming window (the sum of a rectangle and a Hanning window) is used before application of the FFT. Considering the fact that by transforming the signal into the frequency domain we do not lose any original information from the time domain, the signal can completely be reconstructed using the inverse Fourier transform by:

$$x[n] = \frac{1}{N} \sum_{\omega=0}^{2\pi(N-1)/N} \mathit{fft}[\omega] e^{i\omega n},\ 1 \le n \le N \tag{3}$$

Clearly, the longer the segment $x[n]$, i.e., the larger N, the greater the frequency resolution.

Power spectral density is also one of the most important features of the signal in the frequency domain and represents the contribution of each individual frequency component to the power of the whole signal segment $x[n]$. In practice, power spectral density is usually estimated using the coefficients of the fast Fourier transform, i.e., the periodogram (Welch, 1967) given by:

$$per[\omega] = \frac{1}{N} \left|\mathit{fft}[\omega]\right|^2 \tag{4}$$

which is an unbiased and inconsistent estimator. Thus, with the increase in the length of the signal segment, the mean of the estimation tends toward the actual value of power spectral density, which is actually an advantage, unlike variance estimation, which is not reduced, i.e., which does not have a tendency toward zero with the increase in segment length. A periodogram can be further normalized by the total signal power, i.e.,:

$$per_{norm}[\omega] = \frac{1}{N} \left|\mathit{fft}[\omega]\right|^2 \Big/ \sum_{\omega=0}^{2\pi(N-1)/N} per[\omega] \tag{5}$$

where we obtain the relative contribution of each frequency component to the total power of the signal. If the original signal segment $x[n]$ is further divided into P sub-segments of the N/P samples, the periodogram can be calculated as follows:

$$per[\omega] = \frac{1}{P} \sum_{p=0}^{P-1} \frac{P}{N} \left|\mathit{fft}_p[\omega]\right|^2 \tag{6}$$

where $\mathit{fft}_p[\omega]$ is the fast Fourier transform of each of the sub-segments of the N/P sample. In this way, the periodogram is actually an averaged one with a smaller variance, but clearly with a lower resolution in the frequency domain. Based on the periodogram we extracted relative power of all five previously mentioned sub-bands, i.e., delta (0–4 Hz), theta (4–8 Hz), alpha (8–12 Hz), beta (12–30 Hz), and gamma (30–60 Hz), as features of interest in frequency domain.

By analyzing the EEG signals solely in the time domain, extracted features do not contain any information on frequencies, which are, as we will later show, also very important for the proper detection of epileptic EEG signals. On the other hand, by transforming the signals from the time into the frequency domain, any information on time is completely lost, except of course in the case of sequential application on sufficiently short and stationary sub-segments, which also has its disadvantage in terms of the correct choice of the length of these sub-segments which would enable the simultaneous achievement of the desired resolution in both domains. In addition, once selected, the sub-segment length, i.e., the resolution in the time domain, remains fixed throughout the entire frequency bands and cannot be adjusted to the dominant signal frequencies at a specific time. Signal processing using wavelets very accurately resolves this deficiency and results in sufficient information on non-stationary signals, both in the time and frequency domain. We are already familiar with the fact that a signal can be presented as a linear combination of its basic functions. A unit impulse function whose power is limited and whose mean differs from zero is the basic function of the signal in the time domain, whereas in the frequency domain, this role is assigned to the sinusoidal function

that has infinite power, and a zero mean. In the time-frequency domain, the basic function is the wavelet, which is actually a function of limited power, i.e., duration, and a zero mean (Rao and Bopardikar, 1998), and for which the following is valid:

$$\sum_{n=-\infty}^{\infty} |\psi[n]|^2 < \infty, \quad \sum_{n=-\infty}^{\infty} \psi[n] = 0. \tag{7}$$

The wavelet that is moved, or translated, in time for b samples and scaled by the so-called dilation parameter a is given by:

$$\psi_{ab}[n] = \frac{1}{\sqrt{a}}\psi\left[\frac{n-b}{a}\right]. \tag{8}$$

By changing the dilation parameter, the basic wavelet ($a = 1$) changes its width, that is, it spreads ($a > 1$) and contracts ($0 \leq a < 1$) in the time domain. In the analysis of non-stationary signals, the possibility of changing the width of the wavelet represents a significant advantage of this analysis technique, considering the fact that wider wavelets can be used to extract slower changes, i.e., lower signal frequencies, and narrower wavelets can be used to extract faster changes, i.e., higher frequencies. Following the selection of the values of parameters a and b it is possible to transform segments of the signal $x\left[k\right]$ of N samples, that is, to calculate the wavelet transform coefficients in the following way:

$$w_{ab}[n] = \sum_{\tau=1}^{N} x[\tau]\psi_{ab}[n-\tau], \ 1 \leq n \leq N \tag{9}$$

Thus, what is actually being extracted from the signal are only those frequencies that are within the wavelet frequency band $\psi_{ab}[n]$, i.e., the signals are filtrated by the wavelet $\psi_{ab}[n]$. As previously indicated, based on the coefficients obtained in this way, the original signal can be reconstructed using an inverse wavelet transform. Of course, if necessary, it is possible to also independently reconstruct the part of the signal which is filtered, as well as the part that was rejected by the wavelet $\psi_{ab}[n]$ on the basis of the so-called detail coefficients and approximation coefficients respectively, which are of course a function of the transformation coefficients $\psi_{ab}[n]$.

Parameters a and b can continuously change, which is not so practical especially bearing in mind that the signal can be completely and accurately transformed and reconstructed by using a smaller and finite number of wavelets, that is, by using a limited number of discrete values of parameters a and b, which is also known as the discrete wavelet transform (DWT). In this case, parameters a and b are the powers of 2, which gives us the dyadic orthogonal wavelet network with frequency bands which do not overlap each other. The dilation parameter a, as the power of 2, at each subsequent higher level of transformation, doubles in value in comparison to the value from the previous level, which means that the wavelet becomes twice as wide in the time domain, and has a frequency band that is half as narrow and twice as low. This actually decreases the resolution of the transformed signal in the time domain two-fold, increasing it twice as much in the frequency domain. Thus, the signal frequency band from the previous level is split into two halves at every next level, into a higher band which contains higher frequencies and describes the finer changes, or details, and a lower band that contains lower frequencies and actually represents an approximation of the signal from the previous level. This technique is also known as wavelet decomposition of the signal.

Before the application of DWT, it is necessary to choose the type of the basic wavelet as well as the number of levels into which the signal will be decompose. After analysis of several types of the basic wavelets, the fourth-order Daubechies wavelet (Rao and Bopardikar, 1998) was selected for further analysis within this work since it has good localizing properties both in the time and frequency domains (Kalayci and Özdamar, 1995; Petrosian et al., 2000) Due to its shape and smoothing feature this type of the basic wavelet has already shown good capabilities in the field of EEG signal processing. The discrete wavelet decomposition was performed at four levels that resulted into five sub-bands of clinical interest. The standard deviation and the average relative power of the DWT coefficients in each of the sub-bands were extracted as representative features in time-frequency domain.

Non-linear Analysis

EEG signals, as the result of the activities of an extremely complex and non-linear system, in addition to the fairly well-known and previously described linear techniques, can also be analyzed using some of the non-linear techniques. By using linear techniques, any non-linearity that can be found in the signal is only approximated, which can result in the loss of certain pieces of potentially relevant information. If that is the case, the use of non-linear techniques is preferred since they are more reliable for non-linear analyses, despite the fact that they imply weak signal stationarity (Varsavsky et al., 2011), and the fact that they need somewhat longer segments, which leads to their being computationally more demanding than linear techniques.

Let $x[n]$ again represent the signal segment which is to be analyzed, where $n = 1 \cdots N$. Also, let m denote the lag for which we can define two new sub-segments $x[n]$, the first x_k containing samples starting from k up to $N - m$ and the second x_{k+m} with samples starting from $k + m$ to N. Both of these sub-segments contain $N - k - m + 1$ samples and can be represented opposite one another in the phase space with a lag m and the so-called embedding dimension 2. In case of three sub-segments: x_{k+2m}, x_{k+m} and x_k, the embedding dimension of the phase space would be 3. The lagged phase space provides a completely different view of signal evolution in time, where we can note that the signal gravitates to a certain part of the phase space, known as the attractor. With the aim of constructing lagged phase space, i.e., the signal attractor, it is necessary to previously define the values of the lag and the embedding dimension, which although significantly smaller than the real dimension of the non-linear system space, provides an approximation of the signal complexity and non-linearity (Andrzejak et al., 2001). The lag m should be large enough so that these sub-segments would overlap as little as possible, that is, share as little mutual information as possible, but at the same time sufficiently small so that the sub-segments could be long enough for any further useful analysis. An optimal lag is obtained by determining the mutual information coefficient the sub-segments for different values of the lag m. The mutual

information coefficient is defined by Williams (1997):

$$info_m = \sum_{i=1}^{N_s}\sum_{j=1}^{N_s} p\left(x_k[i], x_{k+m}[j]\right) log_2 \frac{p\left(x_k[i], x_{k+m}[j]\right)}{p\left(x_k[i]\right) p\left(x_{k+m}[j]\right)} \quad (10)$$

where N_s represents the number of areas in which the signal is discretized based on the amplitude and p is the corresponding probability that the sub-segment belongs to a certain area. The first local minimum shown in the graph representing the dependence of the mutual information coefficient on lag determines the optimal lag m_o.

After determining the optimal lag, the minimum embedding dimension of the lagged phase space is estimated using Cao's technique (Cao, 1997). In the phase space with a lag m_o and embedding dimension d, the original segment is represented by its phase portraits, which all together make up the attractor defined by the following points in the lagged phase space:

$$y_d[i] = \left[x[i]\, x\left[i + m_o\right] \cdots x\left[i + m_o(d-1)\right]\right] \quad (11)$$

where $i = 1, 2, \cdots, N - m_o(d-1)$. According to the technique developed by Cao, if d is the right dimension, then the two points are also close to each other in phase space dimension d, as well as in the phase space of dimension $d+1$ and are referred to as real neighbors (Cao, 1997). Dimension increases gradually until the number of false neighbors reaches zero, that is, until the Cao's embedding function defined by:

$$e_d = \frac{1}{N - m_o d} \sum_{i=1}^{N-m_o d} \frac{\left\| y_{d+1}[i] - y_{d+1}[n_{i,\,d}] \right\|}{\left\| y_d[i] - y_d[n_{i,\,d}] \right\|} \quad (12)$$

becomes constant, where $i = 1, 2, \cdots, N - m_o d$ and $y_d[n_{i,\,d}]$ represents the nearest neighbor of $y_d[i]$ in the d-dimensional phase space with a lag m_o. In fact, the minimum embedding dimension d_{min} is determined when the ratio between the e_{d+1}/e_d approaches the value of 1. Since this ratio may approach 1 in some other cases, e.g., for completely random signals, an additional check is also carried out where the Cao's embedding function is redefined and given by:

$$e_d^* = \frac{1}{N - m_o d} \sum_{i=1}^{N-m_o d} \left| x\left[i + m_o d\right] - x\left[n_{i,\,d} + m_o d\right] \right| \quad (13)$$

where $x\left[n_{i,\,d} + m_o d\right]$ is the nearest neighbor of $x\left[i + m_o d\right]$. The constant value of the ratio e_{d+1}^*/e_d^* for different values of the embedding dimension indicates that we are dealing with a random signal. The signal is not random, i.e., it is deterministic if this ratio differs from 1 for at least one value of the embedding dimension, which in that case is also the minimum value.

The correlation dimension is a measure of the complexity of the signal attractor in the lagged phase space. This dimension, unlike most others better known dimensions, may have a fractional value and could thus characterize the dimension, that is, the complexity of the attractors with more precision than the embedding dimension; however, it is always less than or equal to the embedding dimension.

Let C_ε be the correlational sum of the signal segment with N samples within the radius ε in its phase space with a lag m_o and minimum embedding dimension d_{min}, i.e., $M = N - m_o d_{min}$ points $y_{d_{min}}$ given by Williams (1997):

$$C_\varepsilon = \lim_{M\to\infty} \frac{1}{M^2} \sum_{i=1}^{M}\sum_{j=1}^{M} H(\varepsilon - \left\| y_{d_{min}}[i] - y_{d_{min}}[j] \right\|) \quad (14)$$

where H is the Heaviside step function that results in 1 if $y_{d_{min}}[j]$ is within the radius ε of $y_{d_{min}}[i]$, i.e.,:

$$\varepsilon - \left\| y_{d_{min}}[i] - y_{d_{min}}[j] \right\| > 0 \quad (15)$$

otherwise it is 0. The correlation dimension d_{corr} is the approximated slope of the natural logarithm of the correlation sum as a function of ε. Given that the total number of possible distances between two points in a lagged phase space equals $M(M-1)/2$, the correlation dimension could directly be obtained by the Takens estimator (Takens, 1981; Cao, 1997) using:

$$d_{corr} = -\left[\frac{2}{M(M-1)} \sum_{i=1}^{M}\sum_{j=1}^{M} log\left(\frac{\left\| y_{d_{min}}[i] - y_{d_{min}}[j] \right\|}{\varepsilon}\right)\right] \quad (16)$$

The largest Lyapunov exponent λ_{max} represents a measure of both chaotic behavior of the attractor and the divergence of the trajectories in phase space, i.e., the predictability of the signal. Attractor divergence is the distance between two closely positioned points in a phase space after a certain period of time of k samples, which is also known as the prediction length. Based on chaos theory, i.e., the so-called butterfly effect, two points close in the phase space of a chaotic system may have completely different trajectories. Thus, the divergence of the trajectories implies a chaotic system, and vice versa. The Lyapunov exponent actually characterizes the exponential growth of that divergence. The number of Lyapunov exponents is equal to the embedding dimension, and each of these Lyapunov exponents represents the rate of a contracting ($\lambda < 0$) or expanding attractor ($\lambda > 0$) in a certain direction of the phase space. In the case of a chaotic system, the trajectories must diverge in at least one dimension, which means that at least one Lyapunov exponent must be greater than zero, when it is, at the same time, the largest Lyapunov exponent. If several Lyapunov exponents are positive, then the largest among them indicates the direction of the maximum expansion of the attractor and its chaotic behavior. The mean of the trajectory divergence after k samples and a sampling period T_s can be calculated by the Wolf's technique (Wolf et al., 1985; Rosenstein et al., 1993) using:

$$d_T = \frac{1}{(M-k)} \sum_{i=1}^{M-k} \frac{\left\| y_{d_{min}}[i+k] - y_{d_{min}}[n_i + k] \right\|}{\left\| y_{d_{min}}[i] - y_{d_{min}}[n_i] \right\|} \quad (17)$$

where $y_{d_{min}}[i]$ and $y_{d_{min}}[n_i]$ represent two close points on different trajectories in the phase space. The largest Lyapunov exponent λ_{max} is in this case an approximation of the slope of the natural logarithmic trajectory divergence as a function of the

number of samples k, i.e., $d_T = d_0 e^{kT_s \lambda_{max}}$ where d_0 stands for the initial divergence. In addition, there is another very similar more practical technique for the evaluation of the largest Lyapunov exponent proposed by Sato et al. where we first calculate the prediction error for several different values of the number of samples k using:

$$p_k = \frac{1}{(M-k)} \sum_{i=1}^{M-k} log_2 \frac{\left\| y_{d_{min}}[i+k] - y_{d_{min}}[n_i+k] \right\|}{\left\| y_{d_{min}}[i] - y_{d_{min}}[n_i] \right\|} \quad (18)$$

after which the λ_{max} is determined as the slope of the middle and approximately linear part of the prediction error p_k as a function of kT_s.

We extract both the correlation dimension and the largest Lyapunov exponent as features that describe complexity and chaotic behavior of the attractor in the lagged phase space. By choosing the radius ε, the phase space is divided into parts of the dimension ε. While the correlation dimension shows how many points can be found in the surrounding areas of the phase space, the Lyapunov exponent describes the distance between each of the trajectories that terminate in different parts of the phase space but start from the same one. In other words, both of these features give us an idea of how complex and predictable EEG signal is, which, of course, they both interpret and quantify in their own characteristic way.

Dimension Reduction in Feature Space

Let an n-dimensional random vector X be transformed through the application of a certain linear transformation into an n-dimensional random vector $Y = A^T X$ where A is the transformational square matrix of the dimension n. Then the mean vector and the covariance matrix of the random vector Y are $M_Y = A^T M_X$ and $\Sigma_Y = A^T \Sigma_X A$. Based on that, the distance function is:

$$\begin{aligned} d_Y^2(Y) &= (Y - M_Y)^T \Sigma_Y^{-1} (Y - M_Y) = (X - M_X)^T \Sigma_X^{-1} \\ &\quad (X - M_X) = d_X^2(X) \end{aligned} \quad (19)$$

that is, the distance function does not change with the linear transformation. If we were to perform the translation of the coordinate system for the mean vector M_X we would obtain the random vector $Z = X - M_X$ whose mean vector is zero and its covariance matrix is the same as Σ_X. If we wanted to determine the random vector Z which maximizes the distance function $d_Z^2(Z) = Z^T \Sigma^{-1} Z$ under the condition that $Z^T Z = 1$, it is necessary to minimize the following criterion:

$$J = Z^T \Sigma^{-1} Z - \mu \left(Z^T Z - 1 \right) \quad (20)$$

where μ is the Lagrange multiplier. By using a partial derivate $\partial J / \partial Z$ and by equating it with zero, we obtain the following:

$$\partial J / \partial Z = 2\Sigma^{-1} Z - 2\mu Z \implies \Sigma Z = \lambda Z \quad (21)$$

where $\lambda = 1/\mu$. With the aim of obtaining a non-zero solution which satisfies the equation:

$$\Sigma Z = \lambda Z \iff (\Sigma - \lambda I) Z = 0 \quad (22)$$

it is further necessary to find such a parameter λ which satisfies the following so-called characteristic equation of a matrix Σ:

$$|\Sigma - \lambda I| = 0 \quad (23)$$

Every λ which satisfies this characteristic equation is known as eigenvalue of the matrix Σ while the vector Z related to specific eigenvalue is known as an eigenvector. When Σ is a symmetric $n \times n$ matrix, then there are n real eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$ and n real eigenvectors $\Phi_1, \Phi_2, \ldots, \Phi_n$ which are mutually orthogonal and for which $\Sigma\Phi = \Phi\Lambda$ and $\Phi^T\Phi = I$ where $\Phi = [\Phi_1 \, \Phi_2 \cdots \Phi_n]$ is the square matrix of the eigenvectors, Λ the diagonal matrix of the eigenvalues:

$$\Lambda = \begin{bmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_n \end{bmatrix} \quad (24)$$

while I is the identity matrix.

If the matrix Φ is used as a transformation matrix during the linear transformation $Y = \Phi^T X$, then the covariance matrix of the random vector Y will be $\Sigma_Y = \Phi^T \Sigma_X \Phi = \Lambda$. This kind of transformation is orthonormal since for the transformation matrix Φ holds $\Phi^T \Phi = I$. In addition, during all these orthonormal transformations, the Euclidean distance does not change, that is $\|Y\|^2 = Y^T Y = X^T \Phi^T \Phi X = X^T X = \|X\|^2$.

Let X be an n-dimensional random vector of the extracted features which could be represented using n linear independent vectors in the following way:

$$X = \sum_{i=1}^{n} y_i \Phi_i = \Phi Y \quad (25)$$

where $\Phi = [\Phi_1 \, \Phi_2 \, \cdots \, \Phi_n]$ and $Y = \left[y_1 \, y_2 \, \cdots \, y_n \right]$ that is Φ_i are the basis vectors of the new n-dimensional space, and the new coordinates y_i are the scalar products of the basis vectors Φ_i and the random vector X. Assuming that the columns of the matrix Φ or in other words the basis vectors Φ_i are orthogonal, the coordinates of the random vector X in the new space can be obtained in the following way:

$$y_i = \Phi_i^T X. \quad (26)$$

Thus, Y represents a mapped random vector and the orthonormal transformation of the original random vector X. The random vector X approximated using only the m ($m < n$) basis vectors, i.e., the mapped features, could be represented in the following way:

$$\widehat{X}(m) = \sum_{i=1}^{m} y_i \Phi_i + \sum_{i=m+1}^{n} b_i \Phi_i \quad (27)$$

where the approximation error becomes:

$$\Delta X(m) = X - \hat{X}(m) = \sum_{i=m+1}^{n} (y_i - b_i) \Phi_i \quad (28)$$

and the mean squared error:

$$\overline{\varepsilon}^2(m) = E\left\{\left\|\Delta X(m)\right\|^2\right\} = \sum_{i=m+1}^{n} E\left\{(y_i - b_i)^2\right\} \tag{29}$$

has its own minimal value for $b_i = E\left\{y_i\right\} = \Phi_i^T E\{X\}$. The optimal mean squared error can then be presented in the following form:

$$\begin{aligned}\overline{\varepsilon}^2_{opt}(m) &= \sum_{i=m+1}^{n} E\left\{(y_i - E\left\{y_i\right\})^2\right\} \\ &= \sum_{i=m+1}^{n} \Phi_i^T E\left\{(X - E\{X\})(X - E\{X\})^T\right\}\Phi_i \\ &= \sum_{i=m+1}^{n} \Phi_i^T \Sigma_X \Phi_i = \sum_{i=m+1}^{n} \lambda_i \end{aligned} \tag{30}$$

where Σ_X is the covariance matrix of the random vector X and λ_i are its eigenvalues. Thus, the minimal mean squared error of approximation is also equal to the sum of the eigenvalues of the leftout coordinates, which actually means that we should leave out coordinates with the smallest eigenvalues. The mapping of the random vector X into the space made up by the eigenvectors of its covariance matrix Σ_X is known as the Karhunen-Loeve (KL) expansion. When reducing the dimension of the feature space using the KL expansion technique we should bear in mind that the performance of each feature is characterized by its eigenvalue. Thus, by rejecting features we should first reject those with the smallest eigenvalue, i.e., with the smallest variance in the new feature space. For example, in the case of dimension reduction from two to one shown in **Figure 3** the feature y_2 would be rejected as less informative even though it has better discriminatory potential than y_1. Also the coordinates y_i are mutually uncorrelated considering that the covariance matrix of the random vector Y is diagonal, i.e.,:

$$\Sigma_Y = \Phi^T \Sigma_X \Phi = \Lambda = diag\left\{\lambda_1 \lambda_2 \cdots \lambda_n\right\}. \tag{31}$$

Unlike the previously outlined method, the reduction of dimension based on scatter matrices (Fukunaga, 1990; Djurovic, 2006) is of special significance for the new detection technique since it takes into consideration the very purpose of the reduction, that is, the classification of the random vectors. Let L be the number of classes which should be classified and M_i and Σ_i, $i = 1 \cdots L$ the mean vectors and the covariance matrices of these classes, respectively. Then the within-class scatter matrix can be defined by:

$$S_W = \sum_{i=1}^{L} P_i E\left\{(X - M_i)(X - M_i)^T/\omega_i\right\} = \sum_{i=1}^{L} P_i \Sigma_i \tag{32}$$

and the between-class scatter matrix as:

$$S_B = \sum_{i=1}^{L} P_i (M_i - M_0)(M_i - M_0)^T \tag{33}$$

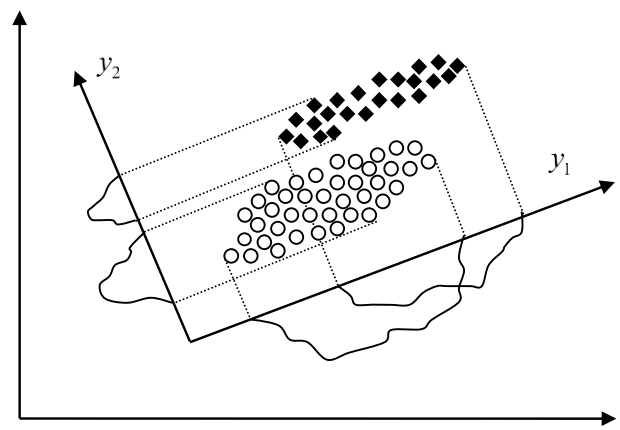

FIGURE 3 | Different approaches to dimension reduction in feature space, the KL expansion technique which rejects the feature y_2 and the technique based on scatter matrices which rejects the feature y_1.

where M_0 is the joint vector of mathematical expectation for all the classes together, that is:

$$M_0 = E\{X\} = \sum_{i=1}^{L} P_i M_i. \tag{34}$$

In addition the mixed scatter matrix can be defined by:

$$S_M = E\left\{(X - M_0)(X - M_0)^T\right\} = S_W + S_B. \tag{35}$$

Then the problem of dimension reduction is reduced to the identification of the $n \times m$ transformation matrix A which maps the random vector X of dimension n onto the random vector $Y = A^T X$ of dimension m and at the same time maximizes the criteria $J = tr(S_W^{-1} S_B)$. This criteria is invariant to non-singular linear transformations and results into transformation matrix that takes the following form:

$$A = [\Psi_1\ \Psi_2\ \cdots\ \Psi_m] \tag{36}$$

where Ψ_i, $i = 1, \ldots, m$ are the eigenvectors of the matrix $S_2^{-1} S_1$ which correspond to the greatest eigenvalues, i.e., $(S_W^{-1} S_B)\Psi_i = \lambda_i \Psi_i$, $i = 1, \ldots, n$, $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$. Dimension reduction based on scatter matrices applied to the case shown in **Figure 3** would result into selection of the feature y_2 that is much better choice than the feature y_1 selected by the KL expansion technique, of course in terms of more accurate classification.

Design of Quadratic Classifiers

Quadratic classifiers are already known to be very good robust solutions to the problems of classification of random vectors whose statistical features are either unknown or change over time. Additionally, quadratic classifiers allow visual insight into the classification results. We design a piecewise quadratic classifier for detection of epileptiform activity, i.e., two quadratic

classifiers, able to separate all three classes of the EEG signals of interest as shown in **Figure 2**. The quadratic classifiers have the same structure defined by the following equation:

$$\begin{aligned} h(Y) &= Y^T Q Y + V^T Y + \nu_0 \\ &= [y_1\, y_2] \begin{bmatrix} q_{11} & q_{12} \\ q_{21} & q_{22} \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} + [\nu_1\, \nu_2] \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} + \nu_0 \end{aligned} \quad (37)$$

where y_1 and y_2 are two features in the reduced feature space. The matrix Q, the vector V and scalar ν_0 are the unknowns which are also need to be determined optimally. The quadratic equation (37) can be represented in a linear form as:

$$h(Y) = [q_{11}\, q_{12}\, q_{22}\, \nu_1\, \nu_2] \begin{bmatrix} y_1^2 \\ 2y_1y_2 \\ y_2^2 \\ y_1 \\ y_2 \end{bmatrix} + \nu_0 = V_z^T Z + \nu_0. \quad (38)$$

In order to also achieve the largest possible between-class and shortest within-class scattering during the dimension reduction in the feature space, for the optimization criterion we have selected the following function (Fukunaga, 1990):

$$f = \frac{P_1 {\eta_1}^2 + P_2 {\eta_2}^2}{P_1 {\sigma_1}^2 + P_2 {\sigma_2}^2} \quad (39)$$

where P_1 and P_2 are probabilities and

$$\eta_l = E\left\{h(Z)/\omega_l\right\} = E\left\{V_z^T Z + \nu_0/\omega_l\right\} = V_z^T M_l + \nu_0 \quad (40)$$

$${\sigma_l}^2 = var\left\{h(Z)/\omega_l\right\} = var\left\{V_z^T Z + \nu_0/\omega_l\right\} = V_z^T \Sigma_l V_z. \quad (41)$$

M_l and Σ_l are the mean vectors and covariance matrices, respectively, of the random vector Z for each of the two classes l that need to be classified. By optimizing the function f, for the optimal vector V_z, i.e., matrix Q and vector V from Equation (37), we have:

$$V_z = \begin{bmatrix} q_{11} \\ q_{12} \\ q_{22} \\ \nu_1 \\ \nu_2 \end{bmatrix} = [P_1 \Sigma_1 + P_2 \Sigma_2]^{-1} (M_2 - M_1) \quad (42)$$

and for the optimal scalar:

$$\nu_0 = - V_z^T (P_1 M_1 + P_2 M_2) \quad (43)$$

which finishes the design of the quadratic classifiers as well as the new technique for detection of epileptiform activity.

Statistical performances such as sensitivity, specificity and accuracy of the designed piecewise quadratic classifier, i.e., the new technique for detection of epileptiform activity, is estimated based on the classification results. The sensitivity is defined as a ratio between the number of correctly classified segments and the total number of the segments for each of the classes separately. The specificity is also calculated for each of these three classes separately and represents the ratio between the number of correctly classified features of the other two classes and the total number of the segments of these two classes. The accuracy is calculated as the ratio between the total number of correctly classified segments and the total number of the segments in all three classes together.

Results

Feature Extraction

In total 30 features for each of 300 analyzed segments of the EEG signals were extracted. All the features together with their mean values and standard deviations for all three different classes of EEG signals of interest are presented in **Table 1**. The extracted features refer to the adequate clinical sub-bands since these sub-bands had better discrimination characteristics compared with the whole frequency band between 0 and 60 Hz. The separability index as a measure of the discriminatory potential was also calculated for all the extracted features. In this case, the separability index is the criteria $J = tr(S_W^{-1} S_B)$ where S_W and S_B are earlier defined within- and between-class scatter matrices, respectively. Based on these matrices, a higher separability index corresponds to better separability between different classes of the EEG signals. Based on these 30 features, each original segment of the EEG signals from time domain can be presented now by its feature vector $X = [x_1 x_2 \cdots x_{30}]^T$, i.e., by the point in the feature space with dimension of 30.

The total variation is the only one feature that we extracted in the time domain. In **Table 1**, it can be noticed that the total variation has a certain potential for the detection of epileptiform activity in EEG signals. However, the total variation is not that much reliable despite the fact that is a pretty well estimated having in mind the duration of each of the analyzed segments.

The periodogram represents a very important feature of the signal in the frequency domain given that based on it we can get a relative contribution of either any individual frequency or a specific frequency band to the total power of the analyzed signal. The periodograms of one epileptic and two non-epileptic (from both unhealthy and healthy tissue) segments of the EEG signals are shown in **Figure 4** where it can be noticed that the EEG signal power of is shifting from lower to higher frequencies in the presence of epileptiform activity.

Using the discrete wavelet transform (DWT) we can completely and independently extract higher and lower frequencies from the signal. All that can be done with different resolution in the time domain, i.e., higher resolution in the time domain for higher frequencies and lower resolution in the time domain for lower frequencies. The EEG signal segments were analyzed at four levels, i.e., the discrete wavelet decomposition was performed at four levels as presented in **Figure 5**. At the first level of decomposition, the original frequency band of the EEG signals (0–60 Hz) was divided into its higher (30–60 Hz) and lower part (0–30 Hz), i.e., the details and the approximation of the signals at the first decomposition level, respectively. Then at the second decomposition level, the frequency band of the approximation from the first level was additionally divided into its higher (15–30 Hz) and lower (0–15 Hz) part, i.e., the

TABLE 1 | Normalized features extracted from different frequency sub-bands.

Index	Feature	Non-epileptic of healthy tissue		Non-epileptic of unhealthy tissue		Epileptic		Separa-bility
		μ	σ	μ	σ	μ	σ	J
x_1	Total variation—delta	0.011	0.002	0.011	0.003	0.019	0.005	1.253
x_2	Total variation—theta	0.027	0.004	0.022	0.006	0.028	0.006	0.300
x_3	Total variation—àlpha	0.044	0.005	0.034	0.011	0.042	0.011	0.215
x_4	Total variation—beta	0.075	0.008	0.057	0.024	0.062	0.023	0.150
x_5	Total variation—gamma	0.149	0.019	0.102	0.047	0.103	0.041	0.335
x_6	Relative power FFT—delta	0.446	0.090	0.628	0.147	0.267	0.220	0.720
x_7	Relative power FFT—theta	0.159	0.049	0.236	0.119	0.390	0.224	0.417
x_8	Relative power FFT—alpha	0.162	0.043	0.086	0.066	0.134	0.057	0.316
x_9	Relative power FFT—beta	0.221	0.075	0.046	0.024	0.205	0.151	0.641
x_{10}	Relative power FFT—gamma	0.012	0.010	0.004	0.003	0.004	0.005	0.264
x_{11}	St. dev. coeff. DWT—delta	2.825	0.275	3.362	0.290	2.507	0.549	0.810
x_{12}	St. dev. coeff. DWT—theta	1.795	0.180	1.709	0.366	2.181	0.505	0.300
x_{13}	St. dev. coeff. DWT—alpha	1.266	0.140	0.766	0.175	1.275	0.288	1.276
x_{14}	St. dev. coeff. DWT—beta	0.556	0.122	0.267	0.072	0.466	0.146	1.057
x_{15}	St. dev. coeff. DWT—gamma	0.154	0.039	0.085	0.028	0.115	0.040	0.596
x_{16}	Relative power DWÒ—delta	0.501	0.097	0.708	0.118	0.408	0.175	0.873
x_{17}	Relative power DWÒ—theta	0.203	0.039	0.190	0.081	0.311	0.132	0.347
x_{18}	Relative power DWÒ—alpha	0.202	0.043	0.077	0.035	0.213	0.097	0.913
x_{19}	Relative power DWÒ—beta	0.081	0.038	0.020	0.011	0.060	0.039	0.613
x_{20}	Relative power DWÒ—gamma	0.013	0.007	0.005	0.003	0.008	0.006	0.291
x_{21}	Correlation dimension—delta	6.979	3.443	6.494	1.605	5.763	1.489	0.045
x_{22}	Correlation dimension—theta	4.621	0.594	4.288	0.925	4.206	0.884	0.048
x_{23}	Correlation dimension—alpha	4.184	0.442	3.701	0.886	3.230	0.833	0.272
x_{24}	Correlation dimension—beta	3.635	0.359	3.097	0.940	2.348	0.832	0.490
x_{25}	Correlation dimension—gamma	6.729	1.248	6.374	1.838	4.003	1.994	0.493
x_{26}	Largest Lyapunov exp.—delta	3.282	0.873	2.910	0.856	4.203	1.102	0.327
x_{27}	Largest Lyapunov exp.—theta	8.213	1.935	8.188	1.914	8.286	1.933	0.000
x_{28}	Largest Lyapunov exp.—alpha	17.58	2.165	17.57	2.160	17.58	2.377	0.000
x_{29}	Largest Lyapunov exp.—beta	32.91	5.991	32.65	5.977	33.04	5.091	0.001
x_{30}	Largest Lyapunov exp.—gamma	11.71	2.985	11.62	2.965	11.89	5.210	0.001

details and the approximation of the signals at the second decomposition level, respectively. After all four decomposition levels, the original band was divided into its five sub-bands, i.e., four sub-bands with the details and one sub-band with the approximation. All these five sub-bands approximately correspond to the earlier defined clinical sub-bands. Power distribution of the EEG signals in the time-frequency domain is quite well described by the DWT coefficients. However, in order to reduce the dimension of the problem and make easier further classification we calculated certain statistics of these coefficients in each sub-band such as the standard deviation and the average relative power, i.e., the square of the absolute values of the DWT coefficients.

Given that the EEG signal also roughly represents a dynamics of a very complex non-linear system such as the brain, the non-linear analysis based on the chaos theory was used in order to extract the information that could not been extracted by any of previously described linear techniques. It is interesting to see that unlike the other feature extraction techniques in the field, a complete agreement about if at all and how to perform a non-linear analysis of the EEG signals has not been achieved yet. Thus, quite often it is possible to find contradictory results of such experiments in the literature. For example, the correlation dimension and the largest Lyapunov exponent have completely different values in Hively et al. (1999), Adeli and Ghosh-Dastidar (2010) and Iasemidis and Sackellares (1991). The feature extraction techniques and non-linear analysis implemented and used in this research are exclusively based on the chaos theory described in the methods part. In addition, there are no any further subjective adjustments applied to the EEG signals, which provides a high level of reproductivity of the obtained results at any time.

At first, the optimal lag and the embedding dimension were determined in order to reconstruct a segment of the EEG signals in its own lagged phase space. The optimal lag m_o was obtained as the first local minimum of the function of the mutual information

coefficients. The value of the optimal lag of the most of analyzed segments varied between 5 and 7. The minimum embedding dimension d_{min} was determined using Cao's technique, i.e., based on the saturation of the embedding function e_d, for example as presented in **Figure 6** in the case of one segment. In other words when a further increase in the embedding dimension does not result in more than 5% of increase in the embedding function. The value of the embedding function of all 300 segments processed approached 1. In fact, this confirms that there is a certain level of chaos present in the segments of the EEG signals. That chaos is not random but deterministic given that the value of the redefined embedding function e_d^* is not constant for all values of the embedding dimension as it can be seen in **Figure 6**. The value of the minimum embedding dimension varied between 4 and 10.

After reconstruction of the EEG signals in the lagged phase space, the correlation dimension of attractor was estimated using

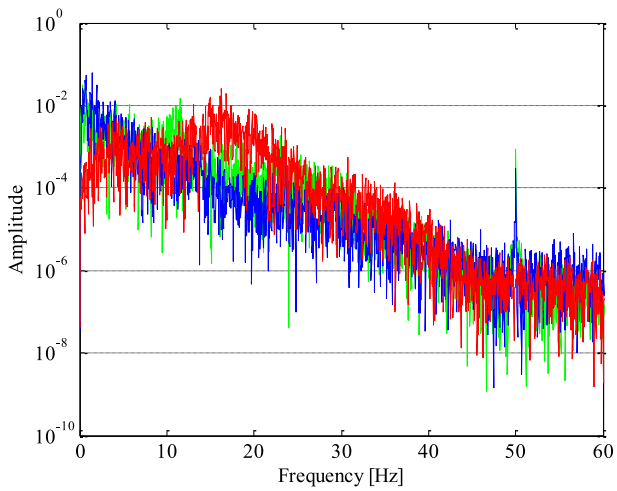

FIGURE 4 | Periodogram of epileptic (in red) and non-epileptic (unhealthy in blue and healthy tissue in green) segments of EEG signals where a shift in the EEG signal power from lower to higher frequencies in the presence of epileptiform activity is evident.

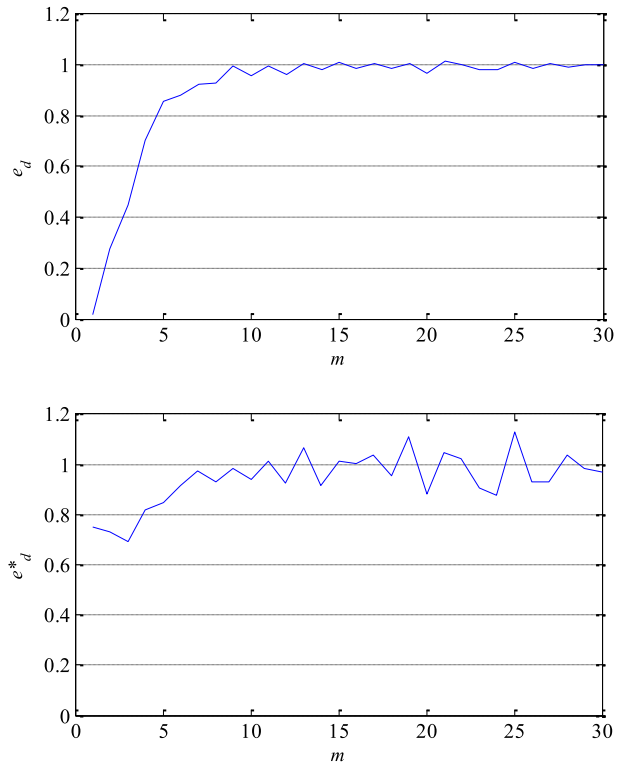

FIGURE 6 | Embedding function e_d (upper) which approaches 1 and thus confirms a presence of a certain level of chaos in EEG signals and redefined embedding function e_d^* (lower) which is not constant for all values of the embedding dimension *m* confirming that chaos is not random but deterministic.

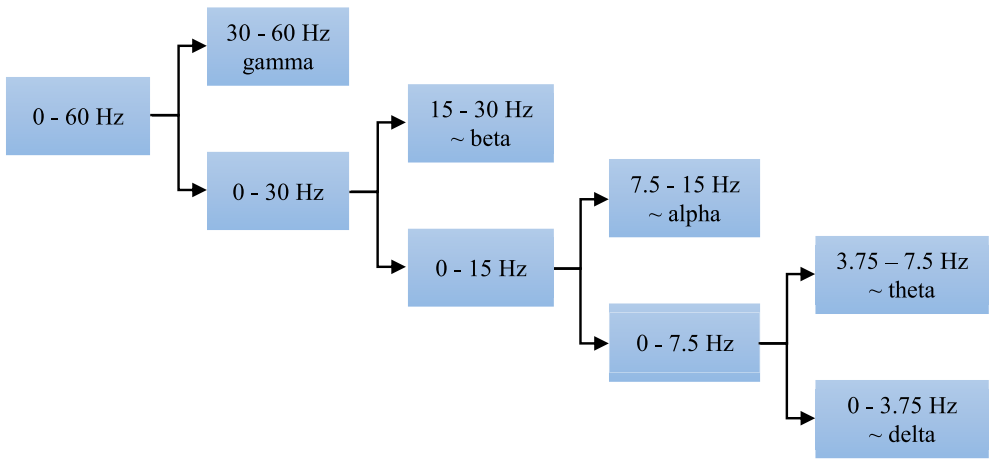

FIGURE 5 | Four-level decomposition of EEG signal that corresponds to five sub-bands of clinical interest which have better discriminatory characteristics compared with the entire frequency band of 0–60 Hz.

the Taken's estimator. After a few tests the value of radius ε in the phase space was set to 5% of the total size of the attractor since the higher values resulted into to many points, and the smaller ones into insufficient number of points for a good estimation of the correlation dimension. From **Table 1**, it can be concluded that the correlation dimension as a non-linear feature has a potential for detection of epileptiform activity in EEG signals. It is also obvious that the attractor complexity, i.e., the chaotic behavior of the EEG signals, is lower in presence of epileptiform activity. The values of the correlation dimension in all cases were higher than the embedding dimension of the lagged phase space, which is in accordance with the chaos theory.

The largest Lyapunov exponent as a measure of signal predictability was estimated using Sato's technique. At first, the prediction error as a function of number of samples k was determined as shown in **Figure 7** in the case of one segment. Then, the largest Lyapunov exponent was estimated based on the function's slope in its medium part. As it can be seen in **Table 1**, the largest Lyapunov exponent has smaller discrimination ability compared with the correlation dimension. Additionally, it can be also noticed that the presence of epileptiform activity reduces the predictability of the EEG signals since the largest Lyapunov exponent is slightly higher in that case.

Dimension Reduction in Feature Space

After the feature extraction from all the segments of the EEG signals, obviously none of the individually extracted features is sufficiently reliable for detection of epileptiform activity in EEG signals. This fact represents the main reason to perform the feature extraction in a few different domains of interest, i.e., time, frequency, time-frequency domain and non-linear analysis. The assumption is that the each of them contains some new information about the EEG signal, i.e., the information which is not present in any other domain and thus later contributes to more accurate classification and detection. Therefore, a better separability between the classes of epileptic and non-epileptic segments is expected after an optimal combination of the features from different domains than in the case of using only features from one domain as it is the case with almost all the literature in the field.

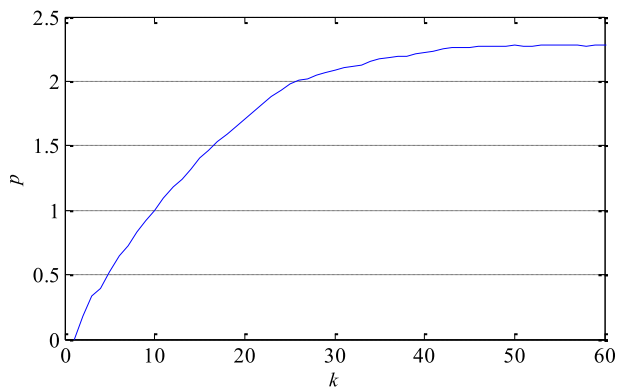

FIGURE 7 | Prediction error *p* of one segment of EEG signal as a function of the number of samples *k*. Its slope in the middle part determines the largest Lyapunov exponent as a measure of the exponential divergence of nearby phase space trajectories.

Both the KL expansion technique and the dimension reduction technique based on the scatter matrices were tested on the features from all the domains. The obtained results, i.e., adequate separability indexes before and after the dimension reduction in the feature space are presented in **Table 2**. The reduction technique based on the scatter matrices gives better results in all the domains of interest and also results into the separability index that is, as expected, greater than any individual separability index given in **Table 1**.

In **Table 2**, one can see that out of all the analyzed features, the highest separability index and the best discrimination characteristics between epileptic and non-epileptic segments have the features obtained in time-frequency domain after the DWT. However, the other features despite their lower separability indexes are also useful for later classification that is concluded based on an additional analysis whose results are presented in **Table 3**. It can be noticed that starting from the features in time domain the separability index increases by a gradual inclusion of the features from other domains.

Unlike the previous figures, **Figure 8** shows 50 original nineteen-dimensional feature vectors X, which correspond to 50 segments from each of the three classes of the EEG signals, mapped into their new reduced two-dimensional feature space. All these 150 two-dimensional vectors Y will be later used in the next section for the design of appropriate classifiers while the rest of 150 segments and their corresponding feature vectors will be used to test the performance of the designed classifiers as well as the total accuracy of the new technique for detection of epileptiform activity in EEG signals.

TABLE 2 | Separability indexes after application of two different techniques for dimension reduction in feature space.

Features analyzed	Dimension		Separability index	
	Before	After	KL expansion	By the scatter matrices
Time domain (x_{1-5})	5	2	1.93	2.13
Frequency domain (x_{6-10})	5	2	1.25	2.16
Time-frequency domain (x_{11-15})	10	2	1.40	4.78
Non-linear analysis (x_{16-20})	10	2	1.07	1.15

TABLE 3 | Separability indexes after the reduction based on the scatter matrices and gradual involvement of features from different domains.

Features analyzed	Dimension		Separability index
	Before	After	
Time domain (x_{1-5})	5	2	2.13
Including frequency domain (x_{1-10})	10	2	3.52
Including time-frequency domain (x_{1-20})	20	2	6.74
Including non-linear analysis (x_{1-30})	30	2	8.78

Classification

After the reduction of the feature space dimension to two, the next step is the design of appropriate classifiers that can separate epileptic from non-epileptic segments of the EEG signals in the reduced feature space shown in **Figure 8**. This represents the last step in design of the new technique for detection of epileptiform activity in EEG signals. Having in mind the nature of the EEG signals and possible changes in their statistical properties it is very desired to use robust classifiers. Based on **Figure 8** it can be concluded that quadratic classifiers represent quite logical choice for classification even though these three classes of the EEG signals are also piecewise linearly separable but with a much higher classification error. In total two quadratic classifiers were designed following the procedure described in Section Design of Quadratic Classifiers.

As it can be seen in **Figure 9**, the first classifier separates the non-epileptic segments of the EEG signals of healthy brain tissue (in green) from the non-epileptic segments of unhealthy tissue (in blue) as well as from the epileptic segments (in red). This classifier is defined using the following equation:

$$h(Y) = \sum_{i=1}^{2}\sum_{j=1}^{2} q_{ij} y_i y_j + \sum_{i=1}^{2} \nu_i y_i + \nu_0 \tag{44}$$

where the unknown parameters are $q_{11} =- 4870.8$, $q_{12} = q_{21} =- 239.9$, $q_{22} =- 174.9$, $\nu_1 =- 29.2$, $\nu_2 =- 174.9$ and $\nu_0 =- 2.3$. After that, the second classifier which separates the remaining two unseparated classes of the EEG signals segments, i.e., the epileptic and the non-epileptic segments of unhealthy brain tissue, was designed. The parameters of the Equation (44) for this classifier are $q_{11} =- 436.7$, $q_{12} = q_{21} =- 128.2$, $q_{22} = 444.6$, $\nu_1 =- 237.9$, $\nu_2 =- 57.2$ and $\nu_0 = 0.5$ while the classifier itself is shown in **Figure 10**.

The performance of the designed classifiers and thus the new technique for detection of epileptiform activity in EEG signals was tested by classification of the remaining 150 segments which were not previously used during the design procedure. The obtained results are presented in **Figure 11**, where the piecewise quadratic classifier is just a combination of two quadratic classifiers.

The classification results can also be represented by a confusion matrix that is given in **Table 4**, where its each cell contains number of classified features for each combination of three classes of the EEG signals segments. Based on the confusion matrix and **Figure 11**, it can be concluded that all the non-epileptic segments of healthy tissue were correctly classified.

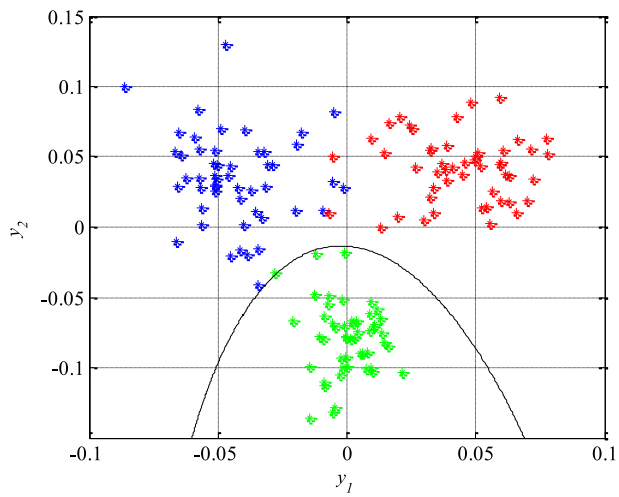

FIGURE 9 | The first quadratic classifier which separates non-epileptic EEG signals of healthy tissue (in green) from non-epileptic (in blue) and epileptic EEG signals of unhealthy tissue (in red) during the design and training phase.

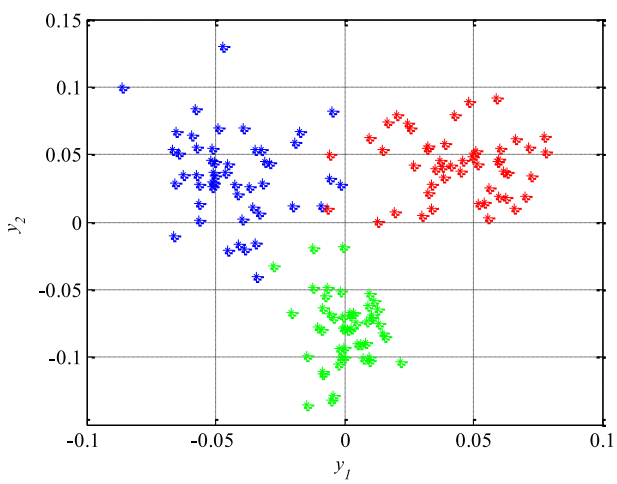

FIGURE 8 | Epileptic (in red) and non-epileptic (unhealthy in blue and healthy tissue in green) EEG signals in a new two-dimensional feature space after dimension reduction based on scatter matrices.

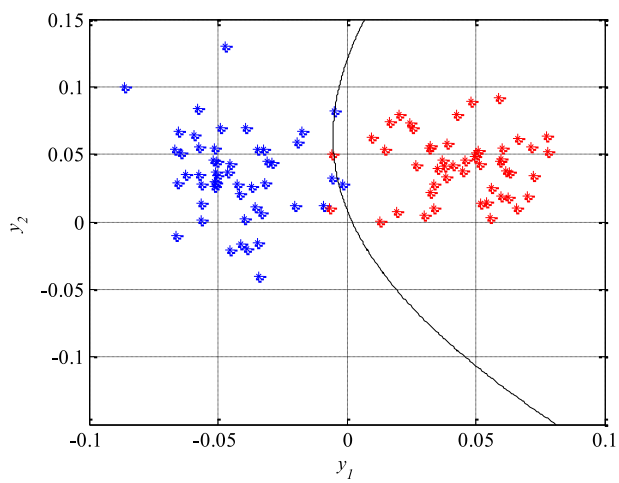

FIGURE 10 | The second quadratic classifier which separates epileptic (in red) from non-epileptic EEG signals of unhealthy tissue (in blue) during the design and training phase.

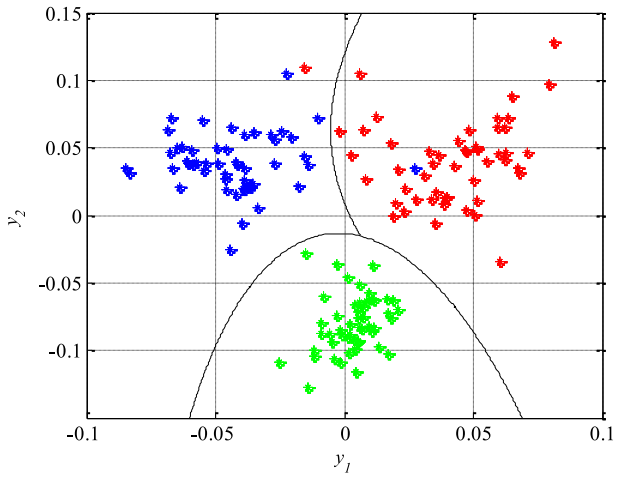

FIGURE 11 | Piecewise quadratic classifier which separates epileptic (in red) from non-epileptic (unhealthy in blue and healthy in green) EEG signals of the test set.

TABLE 4 | Confusion matrix.

EEG signals (input/output)	Non-epileptic		Epileptic
	Healthy	Unhealthy	
Non-epileptic of healthy brain tissue	50	0	0
Non-epileptic of unhealthy brain tissue	0	49	1
Epileptic	0	1	49

TABLE 5 | Statistical performances.

EEG signals	Statistical performances [%]		
	Sensitivity	Specificity	Accuracy
Non-epileptic of healthy brain tissue	100	100	98.7
Non-epileptic of unhealthy brain tissue	98	99	
Epileptic	98	99	

However, the remaining two classes contained one segment each which was incorrectly classified, i.e., classified as it belongs to the other class. The statistical performances such as sensitivity, specificity and accuracy, of the designed piecewise quadratic classifiers are presented in **Table 5**. As it can be seen, the total accuracy of the new technique for detection of epileptiform activity in EEG signals is 98.7%. Typically, quadratic classifiers are robust and do not exhibit overtraining when the number of parameters to be estimated is much less than the number of samples as in this case. Anyway, it is a good practice to cross validate this piecewise classifier in order to ensure its stability. A five-fold cross validation was performed and it resulted in the cross-validation loss, i.e., the error of the out-of-fold samples, of 1.7%. Even though it is slightly higher than the classification error of 1.3% it gives a confidence that the classifier is reasonably accurate.

Discussion

Having in mind the results of other techniques available in the literature, presented in **Table 6** and tested on the identical segments of the EEG signals, the new technique demonstrated a very good performance. The accuracy of the other techniques varied between 85 and 99%. In addition to high accuracy achieved, it should also be emphasized that all the segments of the analyzed EEG signals were normalized before the feature extraction. In that way we managed to overcome one of the main disadvantages of the techniques from **Table 6** in terms of real clinical application, i.e., those techniques rely on the amplitude of the EEG signals as one of the key discriminatory features. However, the EEG signal amplitude has been found as unreliable in real clinical applications since it varies significantly even with healthy individuals, depending on other brain activities as well as other activities of human body. Also, some other undesired effects, e.g., different electrodes used for recording, different patients and their brain tissues, on the detection technique has also been removed by normalization. Unlike the techniques from **Table 6**, which are mainly based only on features from one of the domains, the new technique relies on carefully extracted features from all the domains of interest including non-linear analysis as well. Because of that, this technique is more robust and less sensitive on changes in the EEG signals that dominantly impact the features from one or two domains while at the same time are invisible in other domains and do not have any relation with a presence of epileptiform activity in EEG signals to be detected.

In order to further increase the detection accuracy of the new technique during its real clinical application, a previous elimination of artifacts is very desirable immediately after acquisition of the EEG signals, i.e., before any further processing and feature extraction. The artifacts removal can be performed very reliably using some of already developed and available techniques (Hyvarinen et al., 2001; Rosso et al., 2002). In addition, it is also necessary to make a certain compromise in terms of duration of the segments to be sequentially analyzed in real time. The segment duration should be subsequently adjusted depending on both application and patient. Not only during the feature extraction and the dimension reduction in the feature space, but also during the design of classifiers, a special attention has been paid to the robustness of the detection technique. This resulted in the choice of quadratic classifiers which in addition to their simplicity are known for a high level of robustness in the applications of this type. Quadratic classifiers have also one more important feature that is possibility of visualization of the classification results in two-dimensional space. Namely, despite the fact that the mapped features y_1 and y_2 as a linear combination of the original features x_i extracted from the different domains cannot be anymore associated to certain properties of the EEG signals, they still can provide some further useful insights. For example, in **Figure 11** it can be noticed that the feature y_1 can help during determination of the damage level of the brain tissue, while the feature y_2 indicates either presence or absence of epileptic EEG signal.

As part of our future work we plan an additional testing on other bigger and mainly commercially available data bases of the EEG signals (e.g., http://epilepsy-database.eu) containing much

TABLE 6 | Other techniques for detection of epileptic EEG signals.

Authors and year	Feature extraction	Classification	Accuracy
Nigam and Graupe, 2004	Non-linear filter	Diagnostic neural networks	97.2
Kannathal et al., 2005a	Non-linear analysis	Surrogate data analysis	90.0
Kannathal et al., 2005b	Entropy	Adaptive neuro-fuzzy inference system	92.2
Guler and Ubeyli, 2005	Lyapunov exponents	Recurrent neural networks	96.8
Ubeyli, 2006	Lyapunov exponents	Artificial neural networks	95.0
Sadati et al., 2006	Wavelet transform	Adaptive neuro-fuzzy network	85.9
Subasi, 2007b	Wavelet transform	Expert models	95.0
Tzallas et al., 2007	Time-frequency domain analysis	Artificial neural networks	99.3
Chua et al., 2008	Power spectral density	Gaussian mixture model	93.1
Ghosh-Dastidar et al., 2008	Principal component analysis	Artificial neural networks	99.3
Ocak, 2008	Wavelet transform, approximate entropy and genetic algorithm	Learning vector quantization	98.0
Mousavi et al., 2008	Wavelet transform and autoregressive model	Artificial neural networks	96.0
Ubeyli, 2008	Wavelet transform	Expert models	93.2
Chandaka et al., 2009	Crosscorrelation	Support vectro machines	96.0
Ocak, 2009	Wavelet transform and approximate entropy	Surrogate data analysis	96.7
Guo et al., 2009	Wavelet transform and relative wavelet energy	Artificial neural networks	95.2
Naghsh-Nilchi and Aghashahi, 2010	Eigenvector methods	Artificial neural networks	97.5
Guo et al., 2011	Genetic programming	*K*-nearest neighbor classifier	93.5
Orhan et al., 2011	Wavelet transform	Cauterization and artificial neural networks	96.7
Gajić et al., 2014	Wavelet transform and dimension reduction based on scatter matrices	Quadratic classifiers	99.0

more interictal, preictal and ictal EEG data with the aim of further development and adaptation of the new technique for use in a real clinical environment. We will also try to access its potential in the field of emotion detection (e.g., happiness, sadness, depression, alertness, etc.) as well as detection of abnormal activities associated with some other brain disorders such as Alzheimer's disease and schizophrenia.

Acknowledgments

The support from the Marie Curie FP7-ITN InnHF, Contract No: PITN-GA-2011- 289837 and the Erasmus Mundus Action II EUROWEB Project, Contract No: 204625-1-2011-1-SE-ERA MUNDUS-EMA21 is gratefully acknowledged.

References

Adeli, H., and Ghosh-Dastidar, S. (2010). *Automated EEG-Based Diagnosis of Neurological Disorders*. Boca Raton, FL: CRC Press.

Altunay, S., Telatar, Z., and Erogul, O. (2010). Epileptic EEG detection using the linear prediction error energy. *Expert. Syst. Appl.* 37, 5661–5665. doi: 10.1016/j.eswa.2010.02.045

Andrzejak, R., Lehnertz, K., Mormann, F., Rieke, C., David, P., and Elger, C. (2001). 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* 64:061907. doi: 10.1103/PhysRevE.64.061907

Cao, L. (1997). Practical method for determining the minimum embedding dimension of a scalar time series. *Phys. D* 110, 43–50. doi: 10.1016/S0167-2789(97)00118-8

Casson, A., Yates, D., Smith, S., Duncan, J., and Rodriguez-Villegas, E. (2010). Wearable electroencephalography. What is it, why is it needed, and what does it entail? *IEEE Eng. Med. Biol. Mag.* 29, 44–56. doi: 10.1109/MEMB.2010.936545

Chandaka, S., Chatterjee, A., and Munshi, S. (2009). Cross-correlation aided support vector machine classifier for classification of EG signals. *Expert Syst. Appl.* 36, 1329–1336. doi: 10.1016/j.eswa.2007.11.017

Chua, K. C., Chandran, V., Acharya, R., and Lim, C. M. (2008). Automatic identification of epilepsy by HOS and power spectrum parameters using EEG signals: a comparative study. *Conf. Proc. IEEE Eng. Med. Biol. Soc.* 2008, 3824–3827. doi: 10.1109/IEMBS.2008.4650043

Djurovic, Z. (2006). *Practicum on Pattern Recognition*. Belgrade: School of Electrical Engineering.

Fukunaga, K. (1990). *Introduction to Statistical Pattern Recognition*. Boston, MA: Academic Press.

Gajić, D. (2007). *M.Sc. Thesis: Detection of Epileptic Seizures using Wavelet Transform and Fuzzy Logic (in Serbian)*. Belgrade: School of Electrical Engineering.

Gajić, D., Djurovic, Z., Di Gennaro, S., and Gustafsson, F. (2014). Classification of EEG signals for detection of epileptic seizures based on wavelets and statistical pattern recognition. *Biomed. Eng. Appl. Basis. Commun.* 26, 1450021. doi: 10.4015/S1016237214500215

Ghosh-Dastidar, S., Adeli, H., and Dadmehr, N. (2008). Principal component analysisenhanced cosine radial basis function neural network for robust epilepsy and seizure detection. *IEEE Trans. Biomed. Eng.* 55(2 Pt 1), 512–518. doi: 10.1109/TBME.2007.905490

Gotman, J. (1999). Automatic detection of seizures and spikes. *J. Clin. Neurophysiol.* 16, 130–140. doi: 10.1097/00004691-199903000-00005

Guler, I., and Ubeyli, E. D. (2005). Adaptive neuro-fuzzy inference system for classification of EEG signals using wavelet coefficients. *J. Neurosci. Methods* 148, 113–121. doi: 10.1016/j.jneumeth.2005.04.013

Guo, L., Rivero, D., Dorado, J., Munteanu, C. R., and Pazos, A. (2011). Automatic feature extraction using genetic programming: an application to epileptic EEG classification. *Expert Syst. Appl.* 38, 10425–10436. doi: 10.1016/j.eswa.2011.02.118

Guo, L., Rivero, D., Seoane, J. A., and Pazos, A. (2009). "Classification of EEG signals using relative wavelet energy and artificial neural networks," in *Proceedings of the 1st ACM/SIGEVO Summit on Genetic and Evolutionary Computation* (Shanghai).

Hazarika, N., Chen, J., Tsoi, A., and Sergejew, A. (1997). Classification of EEG signals using the wavelet transform. *Signal Process* 59, 61–72. doi: 10.1016/S0165-1684(97)00038-8

Hively, L. M., Gailey, P. C., and Protopopescu, V. A. (1999). Detecting dynamical change in nonlinear time series. *Phys. Lett. A* 258, 103–114. doi: 10.1016/S0375-9601(99)00342-4

Hyvarinen, A., Karhunen, J., and Oja, E. (2001). *Independent Component Analysis*. New York, NY: John Wiley and Sons.

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

Iasemidis, L., Shiau, D., Chaovalitwongse, W., Sackellares, J., Pardalos, P., Principe, J., et al. (2003). Adaptive epileptic seizure prediction system. *IEEE Trans. Biomed. Eng.* 50, 616–627. doi: 10.1109/TBME.2003.810689

Iscan, Z., Dokur, Z., and Tamer, D. (2011). Classification of electroencephalogram signals with combined time and frequency features. *Expert Syst. Appl.* 38, 10499–10505. doi: 10.1016/j.eswa.2011.02.110

Jerger, K., Netoff, T., Francis, J., Sauer, T., Pecora, L., Weinstein, S., et al. (2001). Early seizure detection. *J. Clin. Neurophysiol.* 18, 259–268. doi: 10.1097/00004691-200105000-00005

Jerger, K., Weinstein, S., Sauer, T., and Schiff, S. (2005). Multivariate linear discrimination of seizures. *Clin. Neurophysiol.* 116, 545–551. doi: 10.1016/j.clinph.2004.08.023

Kalayci, T., and Özdamar, Ö. (1995). Wavelet preprocessing for automated neural network detection of EEG spikes. *IEEE Eng. Med. Biol.* 14, 160–166. doi: 10.1109/51.376754

Kannathal, N., Acharya, U. R., Lim, C. M., and Sadasivan, P. K. (2005a). Characterization of EEG-a comparative study. *Comput. Methods Programs Biomed.* 80, 17–23. doi: 10.1016/j.cmpb.2005.06.005

Kannathal, N., Choo, M. L., Acharya, U. R., and Sadasivan, P. K. (2005b). Entropies for detection of epilepsy in EEG. *Comput. Methods Programs Biomed.* 80, 187–194. doi: 10.1016/j.cmpb.2005.06.012

Liang, S. F., Wang, H. C., and Chang, W. L. (2010). Combination of EEG complexity and spectral analysis for epilepsy diagnosis and seizure detection. *EURASIP J. Adv. Signal Process* 2010:853434. doi: 10.1155/2010/853434

Mousavi, S. R., Niknazar, M., and Vahdat, B. V. (2008). "Epileptic seizure detection using AR model on EEG signals," in *Proceedings of the International Biomedical Engineering Conference* (Cairo).

Naghsh-Nilchi, A. R., and Aghashahi, M. (2010). Epilepsy seizure detection using eigensystem spectral estimation and Multiple Layer Perceptron neural network. *Biomed. Signal Process* 5, 147–157. doi: 10.1016/j.bspc.2010.01.004

Niederhauser, J., Esteller, R., Echauz, J., Vachtsevanos, G., and Litt, B. (2003). Detection of seizure precursors from depth EEG using a sign periodogram transform. *IEEE Trans. Biomed. Eng.* 51, 449–458. doi: 10.1109/TBME.2003.809497

Nigam, V. P., and Graupe, D. (2004). A neural-network-based detection of epilepsy. *Neurol. Res.* 26, 55–60. doi: 10.1179/016164104773026534

Ocak, H. (2008). Optimal classification of epileptic seizures in EEG using wavelet analysis and genetic algorithm. *Signal Process* 88, 1858–1867. doi: 10.1016/j.sigpro.2008.01.026

Ocak, H. (2009). Automatic detection of epileptic seizures in EEG using discrete wavelet transform and approximate entropy. *Expert Syst. Appl.* 36, 2027–2036. doi: 10.1016/j.eswa.2007.12.065

Orhan, U., Hekim, M., and Ozer, M. (2011). EEG signals classification using the K-means clustering and a multilayer perceptron neural network model. *Expert Syst. Appl.* 38, 13475–13481. doi: 10.1016/j.eswa.2011.04.149

Petrosian, A., Prokhorov, D., Homan, R., Dasheiff, R., and Wunsch, D. (2000). Recurrent neural network based prediction of epileptic seizures in intra- and extracranial EEG. *Neurocomputing* 30, 201–218. doi: 10.1016/S0925-2312(99)00126-5

Polat, K., and Gunes, S. (2007). Classification of epileptiform EEG using a hybrid system based on decision tree classier and fast fourier transform. *Appl. Math. Comp.* 187, 1017–1026. doi: 10.1016/j.amc.2006.09.022

Proakis, J., and Manolakis, D. (1996). *Digital Signal Processing: Principles, Algorithms, and Applications*. Upper Saddle River, NJ: Prentice Hall.

Rao, R., and Bopardikar, A. (1998). *Wavelet Transforms: Introduction to Theory and Applications*. Boston, MA: Addison-Wesley Longman.

Rosenstein, M. T., Collins, J. J., and De Luca, C. J. (1993). A practical method for calculating largest Lyapunov exponents from small data sets. *Phys. D* 65, 117–134. doi: 10.1016/0167-2789(93)90009-P

Rosso, O., Martin, M., and Plastino, A. (2002). Brain electrical activity analysis using wavelet-based informational tools. *Phys. A* 313, 587–608. doi: 10.1016/S0378-4371(02)00958-5

Sadati, N., Mohseni, H. R., and Magshoudi, A. (2006). "Epileptic seizure detection using neural fuzzy networks," in *Conference Proceeding IEEE International Conference on Fuzzy System* (Vancouver, BC). doi: 10.1109/FUZZY.2006.1681772

Srinivasan, V., Eswaran, C., and Sriraam, N. (2007). Approximate entropy-based epileptic EEG detection using artificial neural networks. *IEEE Trans. Inf. Technol. Biomed.* 11, 288–295. doi: 10.1109/TITB.2006.884369

Subasi, A. (2007a). Application of adaptive neuro-fuzzy inference system for epileptic seizure detection using wavelet feature extraction. *Comput. Biol. Med.* 37, 227–244. doi: 10.1016/j.compbiomed.2005.12.003

Subasi, A. (2007b). EEG signal classification using wavelet feature extraction and a mixture of expert model. *Expert Syst. Appl.* 32, 1084–1093. doi: 10.1016/j.eswa.2006.02.005

Takens, F. (1981). *Lecture Notes in Mathematics*. Berlin: Springer.

Tzallas, A. T., Tsipouras, M. G., and Fotiadis, D. I. (2007). Automatic seizure detection based on time-frequency analysis and artificial neural networks. *Comput. Intell. Neurosci.* 2007, 80510–80523. doi: 10.1155/2007/80510

Tzallas, A. T., Tsipouras, M. G., and Fotiadis, D. I. (2009). Epileptic seizure detection in EEGs using time-frequency analysis. *IEEE Trans. Inf. Technol. Biomed.* 13, 703–710. doi: 10.1109/TITB.2009.2017939

Ubeyli, E. D. (2006). Analysis of EEG signals using Lyapunov exponents. *Neural Netwk. World* 16, 257–273.

Ubeyli, E. D. (2007). Modified mixture of experts for analysis of EEG signals. *Conf. Proc. IEEE Eng. Med. Biol. Soc.* 2007, 1546–1549. doi: 10.1109/IEMBS.2007.4352598

Ubeyli, E. D. (2008). Wavelet/mixture of experts network structure for EEG classification. *Expert Syst. Appl.* 37, 1954–1962. doi: 10.1016/j.eswa.2007.02.006

Ubeyli, E. D., and Guler, I. (2007). Features extracted by eigenvector methods for detection variability of EEG signals. *Pattern Recogn. Lett.* 28, 592–603. doi: 10.1016/j.patrec.2006.10.004

Varsavsky, A., Mareels, I., and Cook, M. (2011). *Epileptic Seizures and the EEG: Measurement, Models, Detection and Prediction*. Boca Raton, FL: CRC Press.

Wang, D., Miao, D., and Xie, C. (2011). Best basis-based wavelet packet entropy feature extraction and hierarchical EEG classification for epileptic detection. *Expert Syst. Appl.* 38, 14314–14320. doi: 10.1016/j.eswa.2011.05.096

Waterhouse, E. (2003). New horizons in ambulatory electroencephalography. *IEEE Eng. Med. Biol. Mag.* 22, 74–80. doi: 10.1109/MEMB.2003.1213629

Welch, P. D. (1967). The use of fast Fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms. *IEEE Trans. Audio Electroacoust.* AU-15, 70–73. doi: 10.1109/TAU.1967.1161901

Williams, G. P. (1997). *Chaos Theory Tamed*. Washington, DC: National Academy Press.

Wolf, A., Swift, J. B., Swinney, H. L., and Vastano, J. A. (1985). Determining Lyapunov exponents from a time series. *Phys. D* 16, 285–317. doi: 10.1016/0167-2789(85)90011-9

World Health Organization. (2012). *Epilepsy. Fact Sheet N 999*. Available online at: http://www.who.int/mediacentre/factsheets/fs999/en

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2015 Gajic, Djurovic, Gligorijevic, Di Gennaro and Savic-Gajic. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.