# Epileptic Seizure Classification of EEG time-series using Rational Discrete Short Time Fourier Transform

Kaveh Samiee*, Péter Kovács, and Moncef Gabbouj, *Fellow, IEEE*

***Abstract*—A system for epileptic seizure detection in Electroencephalography (EEG) is described in the paper. One of the challenges is to distinguish rhythmic discharges from non-stationary patterns occurring during seizures. The proposed approach is based on an adaptive and localized time-frequency representation of EEG signals by means of rational functions. The corresponding rational Discrete Short Time Fourier Transform (DSTFT) is a novel feature extraction technique for epileptic EEG data. A Multilayer Perceptron (MLP) classifier is fed by the coefficients of the rational DSTFT in order to separate seizure epochs from seizure-free epochs. The effectiveness of the proposed method is compared with several state-of-art feature extraction algorithms used in off-line epileptic seizure detection. The results of the comparative evaluations show that the proposed method outperforms competing techniques in terms of classification accuracy. In addition, it provides a compact representation of EEG time-series.**

***Index Terms*—EEG, Seizure classification, Rational functions, Malmquist–Takenaka system, Time-frequency analysis.**

## I. INTRODUCTION

ELECTROENCEPHALOGRAPHY (EEG) is one of the most common techniques used for monitoring the brain activities. Generally, expert neurologists analyze the records visually. This is time-consuming. In particular, the noisy characteristics of the EEG recording make it difficult to separate seizures from artifacts with similar time-frequency patterns. To address this problem, machine learning algorithms have been widely used for automatic detection or prediction of epileptic seizures in raw EEG signals. Generally, pre-processing and feature extraction steps are performed, followed by (un)supervised data mining methods. Lately, a wide range of general signal processing and machine learning algorithms were adapted for seizure detection and EEG classification. Then the most important step is to extract discriminative features in the time and/or frequency domains. Pre-processing of the raw EEG data is commonly used to remove major artifacts caused by muscle activities or eye blinking.

For instance Polat and Güneş [1], [2] used the fast Fourier transform (FFT) based feature extraction to detect rhythmic discharges in epileptic seizures, while Gabor et al. [3], [4] used FFT to calculate the spectrograms in the frequency band 0 to 32 Hz for each EEG channel. Namely, a two dimensional FFT spectrogram of size $8 \times 32$ was computed for 8 EEG channels. A Self-Organizing Map (SOM) was then used for classification [5]. However, wavelet transforms are more appropriate for analyzing non-stationary EEG signals since they provide a simultaneous representation in the time-frequency domain. For instance, Adeli et al. [6], [7] applied Daubechies wavelet transform to analyze the epileptic seizure pattern in five frequency sub-bands: delta (0-4 Hz), theta (4-7 Hz), alpha (8-12 Hz), beta (13-30 Hz), and gamma (30-60 Hz). Using these five sub-bands and chaos theory, nonlinear dynamics of the EEG are simultaneously quantified by a classifier. Then the correlation dimension (CD) is used to represent the complexity of the system, and the chaoticity is measured by the largest Lyapunov exponent (LLE) in the five sub-bands. Finally, each EEG signal is classified into three classes: healthy, inter-ictal and ictal. It is also worth mentioning that the special Short Time Fourier Transform (STFT), namely the Gabor transform, has been found effective in epileptic seizure analysis. Chen et al. [8] described the EEG signal patterns during epilepsy periods by means of the so-called Frequency Band Relative Intensity Ratio (FBRIR). In this method, a filter bank consisting of a number of Gabor filters, with different parameters such as orientations and central frequencies, was used to detect seizure and seizure-free epochs. The advantage of this method is that it works simultaneously on both scalp EEG and electrocardiography (ECG) and so improves information extraction [9].

The proposed method relies on rational functions which were efficiently used to represent one period of an ECG in [10] and [11]. Furthermore, these systems were successfully applied for modeling the QRS complex [12] and for compressing the heart beats as well. We note that the basic concept of this method was originally proposed in [13]. In this paper, we extend the previous work by providing a detailed description of the time-frequency model, adding feature analysis and more comparisons with the state-of-the-art methods. In particular, we use results from the recently developed theory of signal models based on rational orthogonal polynomials (see e.g.,

Manuscript received November 8, 2013; revised July 1, 2014; accepted September 11, 2014. Date of publication Sep xx, 2014; date of current version September xx, 2014.

*K. Samiee is with the Department of Signal Processing, Tampere University of Technology, Tampere, 33720 Finland (e-mail: kaveh.samiee@tut.fi).

P. Kovács is with the Department of Numerical Analysis, Eötvös Loránd University, Budapest, 1117 Hungary (e-mail: kovika@inf.elte.hu).

M. Gabbouj is with the Department of Signal Processing, Tampere University of Technology, Tampere, 33720 Finland (e-mail: moncef.gabbouj@tut.fi).



0000–0000/00$00.00©2013 IEEE

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TBME.2014.2360101, IEEE Transactions on Biomedical Engineering



[10]–[12]) to construct an efficient epileptic seizure classification algorithm utilizing such functions. Namely, we introduce a novel time-frequency representation, which can be interpreted as a generalization of the classical STFT. Furthermore, we compare the accuracy of both methods by means of the reconstruction error and the classification performance. The main advantages of the proposed feature extraction algorithm are that it is adaptive, scalable and the model is compact in both time and frequency domains. Moreover, we show that our method outperforms conventional time-frequency feature extraction techniques commonly used in epilepsy research in terms of robustness and efficiency. The experiments are available and completely reproducible by using the MATLAB implementation in [14].

The paper is organized as follows. Section II, contains the relevant part of the theory of rational function systems. Section III presents the method for time-frequency representation. Additionally, Section IV provides a novel feature extraction technique for EEG signals. Section V presents the experiments on the recordings of the Bonn EEG database [15] while Section VI discusses the results and the comparisons with state-of-the-art methods. Finally, Section VII concludes the paper.

## II. Rational functions

In this section we give a brief introduction of the theory of rational functions. Let $\mathbb{C}$ represent the set of complex numbers, $\mathbb{D} := \{ z \in \mathbb{C} : |z| < 1 \}$ the open unit disk, $\mathbb{N} := \{ 1, 2, 3, \dots \}$ the set of natural numbers, and $\mathbb{T} := \{ z \in \mathbb{C} : |z| = 1 \}$ the unit circle (or torus).

The basic rational functions (RF) are defined as follows

$$r_{a,k}(z) = \frac{1}{(1-\overline{a}z)^k}\,, \qquad (a \in \mathbb{D}, k \in \mathbb{N})\,.$$

Parameter $a$ is referred to as the *inverse pole* (because $1/\overline{a}$ is a pole in the standard sense), $k$ is said to be the *order* of the basic function. Using a terminology similar to that used in trigonometric, the value $k = 1$ corresponds to the *fundamental tone* and $k > 1$ the *overtones*. Let us denote the rational functions that are analytic on the closed unit disk by $\mathfrak{R}$. Then, it can be shown that $\mathfrak{R} = \operatorname{span}\{r_{a,k} : k \in \mathbb{N}, a \in \mathbb{D}\}$, i.e. any function $f \in \mathfrak{R}$ can be written as

$$f = \sum_{k=0}^{\infty} c_k r_{a,k}\,, \qquad (a \in \mathbb{D}), \tag{1}$$

for an appropriate set of complex coefficients $c_k \in \mathbb{C}$.

However, the basic rational functions $\{r_{a,k} : k \in \mathbb{N}, a \in \mathbb{D}\}$ are linearly independent, but do not form an orthogonal set, so it is difficult to compute the $c_k$ coefficients in Eq. (1). On the other hand, we can easily solve this problem by applying Gram–Schmidt orthogonalization to the basic rational functions. The corresponding rational function system is the so-called Malmquist–Takenaka (MT) system. A handy property of the MT system is that the elements can be explicitly expressed as Blaschke products. Namely, taking the basic functions for a given $n \in \mathbb{N}$ and the sequence of poles $a_0, \dots, a_n \in \mathbb{D}$ the orthogonalized MT system can be written as:

$$\Phi_k(z) = \frac{\sqrt{1-|a_k|^2}}{1-\overline{a}_k z} \prod_{j=0}^{k-1} B_{a_j}(z)\,,$$

with $0 \le k \le n$, where $B_a(z)$ is the so-called Blaschke function defined by

$$B_a(z) := \frac{z-a}{1-\overline{a}z} \qquad (z \in \mathbb{C} \setminus \{ 1/\overline{a} \})\,.$$

Although we have an orthonormal set of functions, the time localization property of the basic rational form has been lost. Fortunately, biorthogonal rational functions (BRF) cure this problem by keeping the orthogonality while avoiding the drawbacks of the MT system. This type of biorthogonal systems can be defined by taking $n+1$ different poles $a_0, \dots, a_n$ with multiplicities $m_0, \dots, m_n$ and the corresponding modified rational base functions (MRF)

$$\varphi_{k,i}(z) = \frac{z^{i-1}}{(1-\overline{a}_k z)^i} \qquad (k = 0, \dots, n,\ i = 1, \dots, m_k)\,.$$

We note that, the system of $r_{a,k}$ and $\varphi_{a,k}$ span the same subspaces of $\mathfrak{R}$. For the definition of a biorthogonal system, the following functions are needed:

$$\Omega_{\ell n}(z) = \frac{1}{(1-\overline{a}_\ell z)^{m_\ell}} \prod_{i=0, i\neq\ell}^{n} \left( \frac{z-a_i}{1-\overline{a}_i z} \right)^{m_i}\,,$$

$$\omega_{\ell n}(z) = \frac{\Omega_{\ell n}(a_\ell)}{\Omega_{\ell n}(z)}$$

where $(0 \le \ell \le n)$.

By Theorem 1 in [16] the (BRF) functions

$$\Psi_{\ell,j}(z) = \frac{\Omega_{\ell n}(z)(z-a_\ell)^{j-1}}{\Omega_{\ell n}(a_\ell)} \sum_{s=0}^{m_\ell - j} \frac{\omega_{\ell n}^{(s)}(a_\ell)}{s!} (z-a_\ell)^s$$

$(0 \le \ell \le n,\ 1 \le j \le m_\ell)$ are biorthogonal to $\varphi_{k,i}$ with respect to the scalar product:

$$\langle F, G \rangle = \frac{1}{2\pi} \int_{-\pi}^{\pi} F(e^{it}) \overline{G}(e^{it})\, \mathrm{d}t \quad (F, G \in H^2(\mathbb{T}))\,. \tag{2}$$

More precisely,

$$\langle \Psi_{\ell r}, \varphi_{ks} \rangle = \delta_{k\ell}\delta_{rs}\,,$$

where $(1 \le r \le m_\ell,\ 1 \le s \le m_k,\ 0 \le k, \ell \le n)$ and $\delta_{ij}$ is the well-known Kronecker delta symbol.

We note that, the previously defined rational function systems are complete in the Hardy space $H^2(\mathbb{D})$, if and only if the so-called Blaschke condition is satisfied [17]:

$$\sum_{n=0}^{\infty} (1 - |a_n|) = \infty\,.$$

Then for a finite set of poles, the MT and the biorthogonal systems span the $N$ and $n$ dimensional subspaces of $H^2(\mathbb{D}))$, respectively:

$$\begin{aligned} P_\Psi^N f &= \sum_{k=0}^{n} \sum_{i=1}^{m_k} \langle f, \Psi_{ki} \rangle\, \varphi_{ki}\,, \\ P_\Phi^n f &= \sum_{k=0}^{n} \langle f, \Phi_k \rangle\, \Phi_k\,, \end{aligned} \tag{3}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TBME.2014.2360101, IEEE Transactions on Biomedical Engineering



where $N = m_0 + m_1 + \cdots + m_n$ .

In this work, we take the restriction of the functions $f \in H^2(\mathbb{D})$ on the unit circle by applying the mapping $t \to e^{it}$, $t \in [-\pi, \pi)$. Besides, we consider the discrete time signals as functions $f \in H^2(\mathbb{T})$ which are uniformly sampled at the unit circle. Then, the scalar product can be approximated by replacing the integral with a sum in Eq. (2).

The advantages of the use of rational function systems over the other transformation methods is as follows

- flexibility in the sense that not only the coefficients but also the system itself can be varied which means the system can be adapted to the EEG signal;
- the coefficients give a compressed representation of the signal, so in further processing steps they can be used as features;
- the elementary waves are localized in time and the basic functions can carry time-frequency information;
- this is a simple analytic representation of the original signal, which means the whole set of analytic tools can be applied on the representation;
- only a couple of arithmetic operations are required to recover the signal.

We note that the MT and the biorthogonal systems $\Phi$ and $\Psi$ with the basic rational functions $\varphi$ are referred as the rational orthogonal basis (ROB) in the literature. The construction of these generalized orthogonal bases were introduced by Heuberger et al. [17].

## III. Generalized Short Time Fourier Transform

Fourier transform is a well-known tool for analyzing the frequency distribution of a signal. Unfortunately, the time information is lost after this transformation. As a consequence, sudden changes of the signal cannot be localized in time. However, there is a wide range of applications where both time and frequency information are needed. On the other hand, several techniques such as short time Fourier transform (STFT), Wigner–Ville transform (WVT), Choi–Williams distribution (CWD), etc. are attempted to solve this problem. Most of these algorithms operate on short segments of the signal, which make it possible to localize the frequencies in time. On the other hand, there is a large trade-off between the resolutions in the two domains (time and frequency). That is any attempt to increase the time resolution causes a decrease in the frequency resolution, and vice versa. The well-known continuous wavelet transform (CWT) provides a good solution for this problem. In the case of CWT, the frequency domain is logarithmically scaled by dynamically changing the window size. For this reason, the lower frequencies have a fine frequency distribution while the higher ones are well localized in time. Generally, the orthogonal wavelet basis is not an adaptive system. Hence, the required number of coefficients can be high for achieving an acceptable level of accuracy.

Let us consider a signal $f \in L^2(\mathbb{R})$ and define its STFT in the form

$$\mathcal{F}_g f(t,\omega) = \int_{\mathbb{R}} f(\tau - t)\overline{g}(\tau) e^{-2\pi i \tau \omega} d\tau \quad (t, \omega \in \mathbb{R}), \quad (4)$$

where $\overline{g}$ is the complex conjugate of $g \in L^2(\mathbb{R})$. In practical applications, the so-called window function $g$ is compactly supported. In this case, the integral in Eq. (4) can be computed over the support of $g$. Furthermore, this algorithm can be interpreted as a successive evaluation of Fourier transforms over short segments of the whole signal. Additionally, the frequencies can be visually represented by displaying the squared magnitude of the Fourier coefficients at each segment. The resulting diagram is called the spectrogram of signal $f$.

Let us denote the uniformly sampled $f(t)$ and $g(t)$ functions by $f[n]$ and $g[n]$], respectively. The discrete (D) STFT over the compactly supported $g$ window function can be written as

$$\mathcal{F}_g f[n,k] = \sum_{m=0}^{M-1} f[m-n] g[m] \epsilon_k[m] \quad (n,k \in \mathbb{N}), \quad (5)$$

where $\epsilon_k[m] = e^{-2\pi k \frac{m}{M}}$ $(0 \le k \le N-1)$, and $M$ is the window length. Perfect reconstruction is possible if we choose $N \ge M$ which is the number of equally spaced frequency bins (see e.g., Chapter 5. in [18]).

Using the same terminology as in Eq. (5) we can define a similar representation of the signal by replacing the trigonometric bases $\epsilon_k$ with the elements of the ROB. More precisely, let us consider $n+1$ different poles $a_0, \ldots, a_n$ with multiplicities $m_0, \ldots, m_n$ and the uniformly sampled functions $f, g \in H^2(\mathbb{T})$. In order to define a generalized time-frequency representation, we set $N$ to be equal to the sum of the multiplicities and $M$ to be a proper window size. Then the rational DSTFT can be written as

$$\mathcal{R}_\phi \mathcal{F}_g f[n,k] = \sum_{m=0}^{M-1} f[m-n] g[m] \phi_k[m], \quad (6)$$

with $0 \le k \le N-1$, where $\phi_k[m] = \Phi_k(e^{-2\pi \frac{m}{M}})$, but $\psi_k[m] = \Psi_{0,k+1}(e^{-2\pi \frac{m}{M}})$ or $\varphi_k[m] = \varphi_{0,k+1}(e^{-2\pi \frac{m}{M}})$ can also be used. The corresponding inverse transform can be written in a similar form

$$\begin{aligned} f[m-n] &\approx \frac{1}{M g[m]} \sum_{k=0}^{N-1} \mathcal{R}_\psi \mathcal{F}_g f[n,k] \varphi_k[m], \\ f[m-n] &\approx \frac{1}{M g[m]} \sum_{k=0}^{N-1} \mathcal{R}_\varphi \mathcal{F}_g f[n,k] \psi_k[m], \\ f[m-n] &\approx \frac{1}{M g[m]} \sum_{k=0}^{N-1} \mathcal{R}_\phi \mathcal{F}_g f[n,k] \phi_k[m], \end{aligned} \quad (7)$$

where $0 \le m \le M-1$. These formulas follow from the discretization of Eqs. (2) and (3) as a consequence of the orthogonal and biorthogonal properties of the rational function systems $\Phi_k$, $\Psi_{l,k}$ and $\varphi_{i,j}$. This procedure can also be interpreted as a windowed Fourier transform, but now we are using a different basis. In addition, if we set $a_k = 0$ for all $k$ then $\varphi_k[m] = \phi_k[m] = \epsilon_k[m]$ so we get back the classical DSTFT as a special case which can be seen in Figs. 2 and 3. We note that the $\approx$ token was used in Eq. (7). The main reason for this is that there are no proper inversion formulas for these types of rational systems at the uniform discretization of the unit disk. On the other hand, perfect reconstruction is

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TBME.2014.2360101, IEEE Transactions on Biomedical Engineering



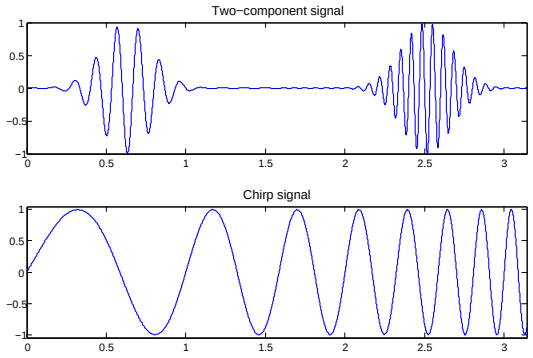



Fig. 1. Schematics of the test signals.

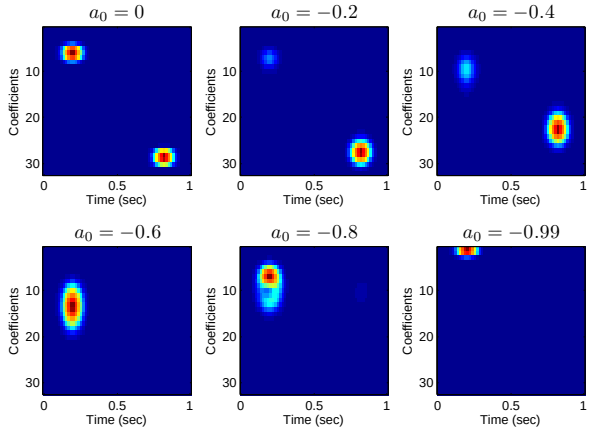



Fig. 2. Rational DSTFT spectra of the function $f = f_1 + f_2$.

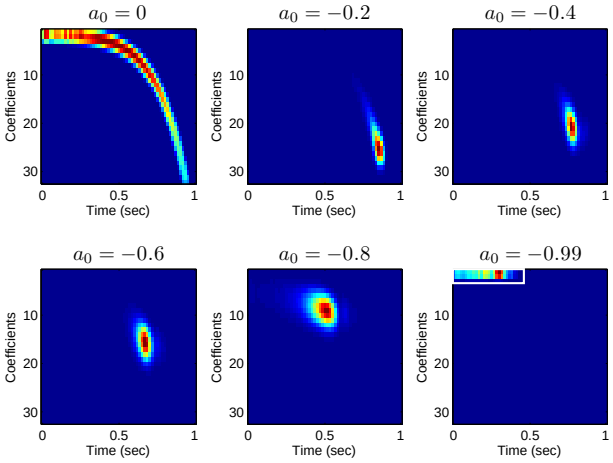



Fig. 3. Rational DSTFT spectra of the chirp signal $f_3$.

also possible if we resample the original signal $f$ over an appropriate non- uniform discrete grid. For further details, the reader is referred to the proof of Theorem 2 in [16].

Let us set the number of poles and their multiplicities such that $N = M$ in Eq. (7). Then the following statement holds for the corresponding non-uniform discretization [19], [20]:

$$\sum_{m=0}^{M-1} (f[m-n]g[m])^2 = \frac{1}{M} \sum_{k=0}^{M-1} \left| \mathcal{R}_\phi \mathcal{F}_g f[n,k] \right|^2 .$$

This can be interpreted as Parsevals formula of the generalized transformation. Hence, we have the same energy preservation property as in the case of the classical DSTFT.

Now, the generalized rational DSTFT and its inverse can be easily computed for $f \in H^2(\mathbb{T})$ using Eqs. (6) and (7). Unfortunately, in most cases the real signals $f \in L^2(\mathbb{T})$ have nonzero negative Fourier coefficients, so they cannot be interpreted as the elements of $H^2(\mathbb{T})$. Nevertheless, we can handle this issue by using the analytic representation of real signals. Namely, we should compute the function $F = f + i\mathcal{H}f$ where $\mathcal{H}$ denotes the well-known Hilbert transformation. For instance the `hilbert` operator can evaluate $F$ in MATLAB and then the generalized rational DSTFT can be applied on the result. Moreover, the coefficients and the expansions in Eq. (3) can be easily calculated by using the `biort_coeffs` and `mt_coeffs` commands of the RAIT MATLAB toolbox [21].

Let us consider some examples with the following functions

$$\begin{aligned} f_1(t) &= e^{-20(t-\frac{1}{5}\pi)^2} \cdot \sin(100\pi t) \ , \\ f_2(t) &= e^{-20(t-\frac{4}{5}\pi)^2} \cdot \sin(2000\pi t) \ , \\ f_3(t) &= \sin(2\pi 6^t), \\ f(t) &= f_1(t) + f_2(t) \, , \\ g(t) &= \sin^2(\pi t) \ , \end{aligned}$$

where $t \in [0, \pi]$ and $g$ is the so-called Hanning window. In our first experiment we apply the rational DSTFT transformation to $f$ which was uniformly sampled with the sampling frequency $f_s = 4000$ Hz. Additionally, we use only one pole $a_0$ with multiplicity $N = m_0 = 32$ by setting the window length $M = 128$. Hence, we compute 32 frequency components at each windowed segment. For the sake of simplicity, we use the uniform distribution on the real interval $[-1, 0]$ as the poles of the biorthogonal rational function system. Then the absolute square values of the $\mathcal{R}_\psi \mathcal{F}_g f[n,k]$ coefficients are displayed in Figs. 2 and 3. We note that for $a_0 = 0$ we get back the classical DSTFT where the two frequency components $f_1$ and $f_2$ are well separated.

## IV. Feature Extraction using Rational DSTFT

### A. Dataset

In this study, we use an EEG database which has been provided by the University of Bonn and is freely available online [15]. This database has been widely used for EEG feature extraction and classification in the literature [22]. The EEG database consists of five sets (A-E). Each set contains 100 single-channel EEG segments, each with a duration of $23.6s$. Sets A and B have been recorded using the standard international 10-20 system for surface EEG recording. Five healthy volunteers were participated in these tests with eyes open (A) and eyes closed (B). For sets C, D and E five epileptic patients were selected for presurgical evaluation of epilepsy by using intracranial electrodes. Depth electrodes were implanted symmetrically to record EEG from the epileptogenic zone (D) and from hippocampal formation of the opposite hemisphere

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TBME.2014.2360101, IEEE Transactions on Biomedical Engineering



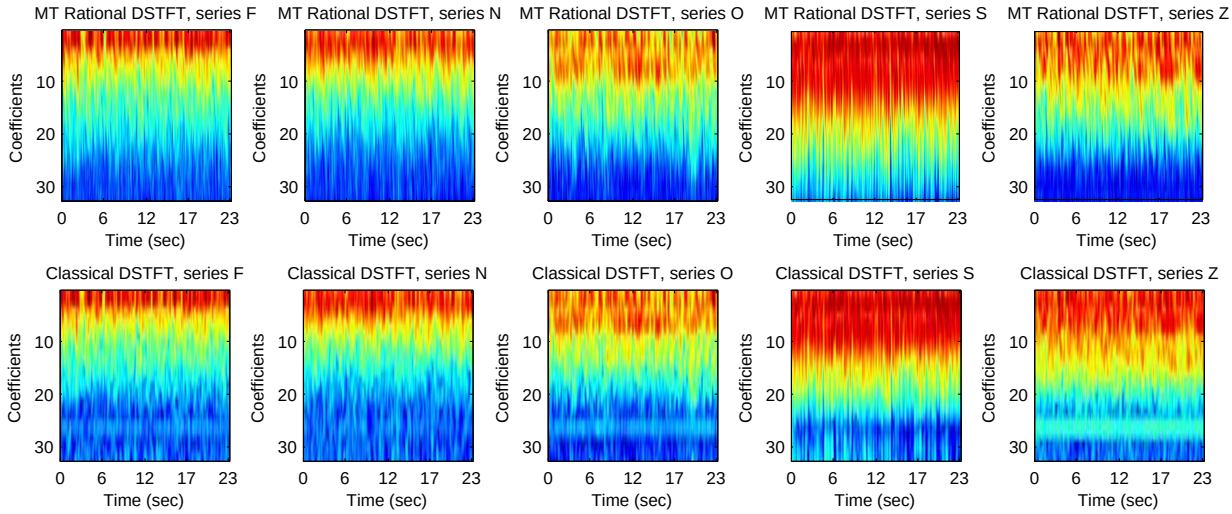



Fig. 4. Spectrogram of sets A to E for different selection of poles. First row) rational DSTFT spectra for $a_0 = -0.1 + 0.1i$, Second row) classical STFT spectra.

of the brain (C). Segments of set E were taken from contacts of all electrodes. In sets C and D, segments contain inter-ictal intervals while seizure activities occur in set E. Each epoch was sampled at 173.61 Hz resulting in a total of 4096 samples.

### *B. Rational DSTFT coefficients*

In the present work, we apply the generalized rational DSTFT to EEG time-series. The seizure epochs in set E contain low frequency oscillations in the theta frequency sub-band along with high frequency rhythmic spikes and high amplitude changes in the alpha and beta frequency sub-bands. These characteristics of the seizure patterns can be used to distinguish them from seizure-free EEG segments. Thus, we want to achieve a better resolution by increasing the number of coefficients. In this work, we use one pole $a_0$ with $m_0$ multiplicity. Then we represent each EEG epoch of length $M$ by $N = m_0$ coefficients of the rational DSTFT. The optimal parameters $a_0$, $M$, $N$ will be determined in Section V. As features, the absolute values of the coefficients and five statistical measures are considered:

- absolute mean value;
- absolute median value;
- absolute standard deviation;
- absolute maximum value;
- absolute minimum value of the coefficients.

The resulting feature vector contains $N + 5$ feature elements for each $M$ sample long EEG epoch. For instance, in Fig. 4, one can see the rational DSTFT spectra for all the five sets and the comparison with the classical STFT spectra. Here we used the MT system $\Phi_k$ with 32 coefficients and a fixed window size of 0.5 second (i.e., $N = 32$, $M = 88$). The figure shows that the frequency responses can be similar to each other as we move the pole closer to zero.

## V. Classification

In order to evaluate the efficiency and the robustness of the proposed feature extraction method, we perform different binary classification tasks. We mainly investigate the detection of epileptic seizure epochs (Set E) in the presence of seizure-free epochs (Sets A, B, C and D). Thus, we define six classification tasks as:

- (E - A): classification of set E in the presence of set A;
- (E - B): classification of set E in the presence of set B;
- (E - C): classification of set E in the presence of set C;
- (E - D): classification of set E in the presence of set D;
- (E - A, C): classification of set E in the presence of sets A and C;
- (E - A, B, C, D): classification of set E in the presence of sets A, B, C and D.

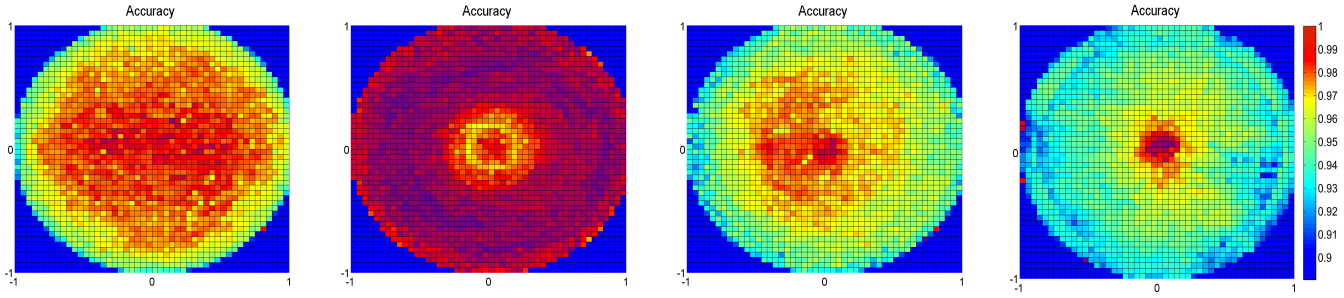
In all of these classification tasks the rational DSTFT features were tested by using various classifiers such as Naive Bayes, Logistic Regression, Support Vector Machine (SVM), K-Nearest Neighbors (KNN) and Multilayer Perceptron (MLP) architectures. After a series of numerical experiments, the feed forward MLP has been selected as the optimal classifier which is trained by the back-propagation algorithm. More precisely, the MLP architecture consists of $N$+5 neurons in the input layer, ($N$+6)/2 neurons in the hidden layer and two neurons in the output layer representing seizure and seizure-free classes.

### *A. Impact of the pole on the classification accuracy*

Once the poles and multiplicities are given, the orthogonal and biorthogonal rational systems are determined. Then the coefficients of the representation can be easily calculated by using the discretization of the scalar product in Eq. (2). Although the system is adaptive due to the free parameters provided by the pole(s), it has an effect on the coefficients as well. Namely, the shape of the base functions varies along with the position of the pole $a_0$. For this reason, the classification results can also change. One of the main interest of this section is to analyze the impact of the position of the pole $a_0$ on the classification accuracy. More precisely, we compare the

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TBME.2014.2360101, IEEE Transactions on Biomedical Engineering






(a) MT (b) RF (c) MRF (d) BRF

Fig. 5. Classification results obtained by different rational coefficients for each pole in the $50 \times 50$ grid .

classification results of the simple (E - A) task for a set of poles which are chosen from a $50 \times 50$ grid of the unit circle. Half of the dataset is used to train the classifier and the other half was used for testing. Additionally, for each 1 second long epoch $M = 176$, we extract $N = 16$ coefficients with respect to four different rational systems (MT, RF, MRF and BRF). Then we apply MLP architecture as the classifier which was described in the previous section.

According to our experiments, the number of zero coefficients increases in the rational DSTFT as the absolute value of the pole increases. This phenomenon is caused by the terms in the rational series expansions in Eq. (6) as they represent higher frequencies while the poles tend closer to the torus. These zero coefficients can reduce the accuracy of the A-E classification task up to $6\%$. However, for the poles which are close enough to the center, the classification accuracy is similar to the rational STFT which is a special case of the proposed model for $a_0 = 0$. This can be seen in Fig. 5 where the classification results of the four rational systems have been displayed at each pole. During this experiment we extracted different rational coefficients for all EEG records while we used the same pole for every segment. Then the classification accuracy at each pole is indicated by a color code. One can see that the results are similar near the zero pole. Unfortunately, stepping away from zero can degrade the results by introducing more zero coefficients to the rational expansion. All the rational function systems defined in Section II share the same behavior because they span the same subspaces of $\mathfrak{R}$. Table I shows the average and the highest classification accuracies for each rational system. As it can be seen, MT representation can outperform all the other rational series expansions. Therefore, we will consider only MT rational systems in our further experiments. Moreover, the proposed algorithm can outperform the classical STFT technique as well by fixing the pole at $a_0 = -0.1 + 0.1i$. It means that the performance of the classical STFT based seizure detection methods can be improved by using rational function systems. Furthermore, the results highlight the importance of choosing the right pole as well.

For this reason, we decided not to fix the pole for the whole signal. More precisely, an optimal pole was chosen by using the well-known particle swarm optimization (PSO) algorithm [23]–[25] to minimize the mean square error (MSE) of the reconstructed signal. Taking advantage of this adaptive behavior of the rational function systems the projection error can be minimized. The purpose of the optimization procedure is to make a compact representation of each segment. Consequently, the coefficients can carry more information and they can be used as features. Flexibility is one of the main advantages of the rational DSTFT comparing to other STFT representations. Namely, the rational systems can be varied from segment to segment in contrast with the uniform representations such as STFT, WVT or even the CWT where the shapes of the base functions are fixed for all segments. For this reason, we expect an improvement in the classification accuracies comparing to STFT based features.

TABLE I
COMPARISON OF AVERAGE AND BEST CLASSIFICATION ACCURACIES OBTAINED BY DIFFERENT RATIONAL COEFFICIENTS IN THE $50 \times 50$ GRID

| Rational System | Avg Acc (%) | Best Acc (%) (non-zero poles) |
|---|---|---|
| MT | **97.9** | **99.4** |
| RF | 94.6 | 97.3 |
| MRF | 96.8 | 99.2 |
| BRF | 94.4 | 99.3 |
| rational STFT | - | 99.0 |

### B. *Impact of the time-frequency resolution*

The classification accuracy of EEG epochs mostly depends on two primary regularizations of the time-frequency transform: 1) time resolution, which is equal to the size of each epoch $M$. In real applications, it is generally more than 1 second; 2) frequency resolution, which is related to the number of coefficients $N$ (i.e., features). Hence, in case of multichannel EEG data, high dimensional feature vectors should be avoided. For this reason, we investigate the trade-off between time and frequency resolutions to reduce the computational complexity.

In order to find the best choice for the epoch size and the optimal number of coefficients, we consider all the possible combinations for $M = 173, 256, 512, 1024$ samples long segments with $N = 8, 16, 32$ number of MT coefficients. We evaluated the average classification accuracy of the six classification tasks for each combination. As it can be seen in Table II, the best average classification accuracy using MT rational

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TBME.2014.2360101, IEEE Transactions on Biomedical Engineering



TABLE II
CLASSIFICATION RESULTS OF MT RATIONAL DSTFT FOR DIFFERENT EPOCH SIZES AND NUMBER OF COEFFICIENTS

| **Classification task** | | **Window size + Number of coefficients** | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | **173** | | | **256** | | | **512** | | | **1024** | | |
| | | **8** | **16** | **32** | **8** | **16** | **32** | **8** | **16** | **32** | **8** | **16** | **32** |
| E-A | **Sens%** | 99.5 | 99.0 | 99.0 | 99.9 | 100.0 | 99.9 | 99.5 | 100.0 | 100.0 | 97.9 | 96.9 | 100.0 |
| | **Spec%** | 99.1 | 99.9 | 99.9 | 98.1 | 99.4 | 99.6 | 98.4 | 98.4 | 99.5 | 96.6 | 99.0 | 97.6 |
| | **Acc%** | 99.3 | 99.5 | 99.5 | 99.0 | 99.7 | 99.8 | 99.0 | 99.3 | 99.8 | 97.3 | 98.0 | 98.8 |
| E-B | **Sens%** | 92.5 | 98.9 | 98.9 | 99.1 | 94.9 | 99.6 | 98.5 | 99.8 | 99.0 | 99.5 | 88.1 | 99.0 |
| | **Spec%** | 98.3 | 98.4 | 98.4 | 94.3 | 98.9 | 99.0 | 95.6 | 92.8 | 98.2 | 91.3 | 98.6 | 95.7 |
| | **Acc%** | 95.4 | 98.7 | 98.7 | 96.8 | 96.9 | 99.3 | 97.1 | 96.4 | 98.6 | 95.3 | 93.5 | 97.3 |
| E-C | **Sens%** | 96.9 | 96.8 | 96.8 | 95.5 | 97.3 | 99.3 | 87.9 | 99.5 | 98.3 | 96.4 | 94.8 | 99.5 |
| | **Spec%** | 96.7 | 97.3 | 97.3 | 97.1 | 99.0 | 97.7 | 95.3 | 93.5 | 98.4 | 78.7 | 91.3 | 90.8 |
| | **Acc%** | 96.8 | 97.0 | 97.0 | 96.3 | 98.1 | 98.5 | 91.5 | 96.6 | 98.4 | 87.3 | 93.0 | 95.0 |
| E-D | **Sens%** | 94.5 | 94.1 | 94.1 | 91.3 | 93.7 | 95.6 | 86.0 | 95.9 | 96.9 | 83.4 | 82.9 | 90.7 |
| | **Spec%** | 94.3 | 95.9 | 95.9 | 94.5 | 94.7 | 94.1 | 88.6 | 89.4 | 92.5 | 86.0 | 93.7 | 92.8 |
| | **Acc%** | 94.4 | 95.0 | 95.0 | 92.9 | 94.2 | 94.9 | 87.3 | 92.8 | 94.8 | 84.8 | 88.5 | 91.8 |
| E-A, C | **Sens%** | 98.8 | 98.7 | 98.7 | 98.2 | 99.1 | 99.2 | 94.6 | 99.3 | 97.8 | 94.0 | 97.7 | 100.0 |
| | **Spec%** | 95.0 | 96.6 | 96.6 | 96.7 | 97.6 | 98.9 | 94.7 | 97.7 | 97.5 | 91.6 | 92.6 | 94.1 |
| | **Acc%** | 97.5 | 98.0 | 98.0 | 97.7 | 98.6 | 99.1 | 94.7 | 98.8 | 97.7 | 93.2 | 96.0 | 98.0 |
| E-A, B, C, D | **Sens%** | 98.3 | 99.6 | 99.6 | 97.4 | 97.1 | 99.2 | 98.3 | 97.8 | 98.7 | 97.3 | 98.5 | 98.2 |
| | **Spec%** | 91.6 | 87.6 | 87.6 | 93.7 | 94.3 | 93.8 | 78.7 | 91.1 | 91.6 | 66.8 | 81.3 | 88.3 |
| | **Acc%** | 96.9 | 97.2 | 97.2 | 97.4 | 96.5 | 98.1 | 94.4 | 96.5 | 97.3 | 90.8 | 94.8 | 96.1 |
| Average | **Sens%** | 96.8 | 97.9 | 97.9 | 96.9 | 97.0 | **98.8** | 94.1 | 98.7 | 98.5 | 94.8 | 93.2 | 97.9 |
| | **Spec%** | 95.8 | 96.0 | 96.0 | 95.7 | **97.3** | **97.2** | 91.9 | 93.8 | 96.3 | 85.2 | 92.8 | 93.2 |
| | **Acc%** | 96.7 | 97.6 | 97.6 | 96.7 | 97.3 | **98.3** | 94.0 | 96.7 | 97.8 | 91.5 | 94.0 | 96.2 |

DSTFT is obtained for 32 coefficients with 256 sample size window which is roughly equal to 1.5 seconds. Besides, the highest average Sensitivity (true-positive-ratio) and Specificity (true-negative-ratio) are also obtained for this combination of epoch size and number of coefficients. Moreover, by increasing the size of the epoch, the performance of the MT rational DSTFT is dropped. This phenomenon is due to the fact that only a limited number of base functions were used for the representation. In contrast, the classical STFT operates with an equal number of bases and samples in the epoch. The former method is an approximation while the latter one is an interpolation procedure. For this reason, rational DSTFT may not be able to fit the signal in the presence of a long epoch and high frequency changes. Hence, we fix the window size to $M = 256$ samples and we extract $N = 32$ coefficients in our further experiments.

## VI. RESULTS

In this section, we compare the performance of the MT rational DSTFT system with the classical STFT in terms of classification accuracy. Based on the results in Section V, we decided to use the optimal pole at each segment by setting $N = 32$ and $M = 256$. Furthermore, we applied the classical STFT which is embedded in MATLAB. For all of the six classification tasks defined in Section V, we consider $50\%$ of the data, chosen randomly, as the training set and the remaining data as the test set. More precisely, for the MT rational DSTFT, we extract feature vectors containing 37 elements (32 coefficients + 5 statistical values). For the classical STFT, we choose 32 coefficients with the largest absolute values from each 256 sample long epoch. Table III shows the classification results for the six classification tasks. As it can be seen, MT rational DSTFT yields the highest classification accuracy. It means that the proposed method is able to achieve a higher sensitivity while keeping the same specificity of the classical STFT. Recall that, sets A, B and sets C, D, and E are recorded using scalp and intracranial electrodes, respectively. However, this has no influence on the performance of the proposed method as it can be seen in Table III. Therefore, the classical STFT the classical STFT based EEG feature extraction algorithms can be improved by using the proposed rational *t-f* distribution. Although we achieved a small amount of enhancement in the sense of classification accuracy, the reconstruction error has been significantly decreased comparing to the classical STFT. Furthermore, the representation is more compact thanks to the adaptive nature of the rational function systems which can be seen in Fig. 6. The inverse discrete (ID) STFT of the signal was computed by using the optimal parameters $N = 32$ and $M = 256$. Then the average mean square error was evaluated for all EEG records of the Bonn dataset. The overall MSEs are as follows: $2.52 \times 10^3$ and $5.5 \times 10^2$. In this sense, the accuracy of the representation is about 4.5 times better than STFT. As a consequence, the proposed representation is more robust than the classical STFT, i.e., one can achieve a better classification result with our model using a smaller set of coefficients or features. In other words, the proposed method provides a sparse representation of the signal while the components remain orthogonal.

### A. Time-frequency distributions

In quantum mechanics, the well-known Heisenberg uncertainty principle states that the position and momentum of a particle cannot be measured at the same time with arbitrarily high precision. In signal processing, this theorem is referred to

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TBME.2014.2360101, IEEE Transactions on Biomedical Engineering



TABLE III
COMPARISON OF CLASSIFICATION RESULTS OF MT RATIONAL DSTFT COEFFICIENTS WITH CLASSICAL STFT COEFFICIENTS.

| **Classification task** | | **Feature extraction** | |
|---|---|---|---|
| | | Classical STFT | MT rational DSTFT |
| E-A | Sens% | 98.2 | 99.9 |
| | Spec% | 100.0 | 99.6 |
| | **Acc%** | 99.1 | **99.8** |
| E-B | Sens% | 98.3 | 99.6 |
| | Spec% | 93.3 | 99.0 |
| | **Acc%** | 95.8 | **99.3** |
| E-C | Sens% | 94.0 | 99.3 |
| | Spec% | 98.1 | 97.7 |
| | **Acc%** | 96.1 | **98.5** |
| E-D | Sens% | 94.2 | 95.6 |
| | Spec% | 94.5 | 94.1 |
| | **Acc%** | 94.3 | **94.9** |
| E-A. C | Sens% | 98.9 | 99.2 |
| | Spec% | 98.9 | 98.9 |
| | **Acc%** | 97.7 | **99.1** |
| E-A. B. C. D | Sens% | 97.7 | 99.2 |
| | Spec% | 93.1 | 93.8 |
| | **Acc%** | 96.8 | **98.1** |

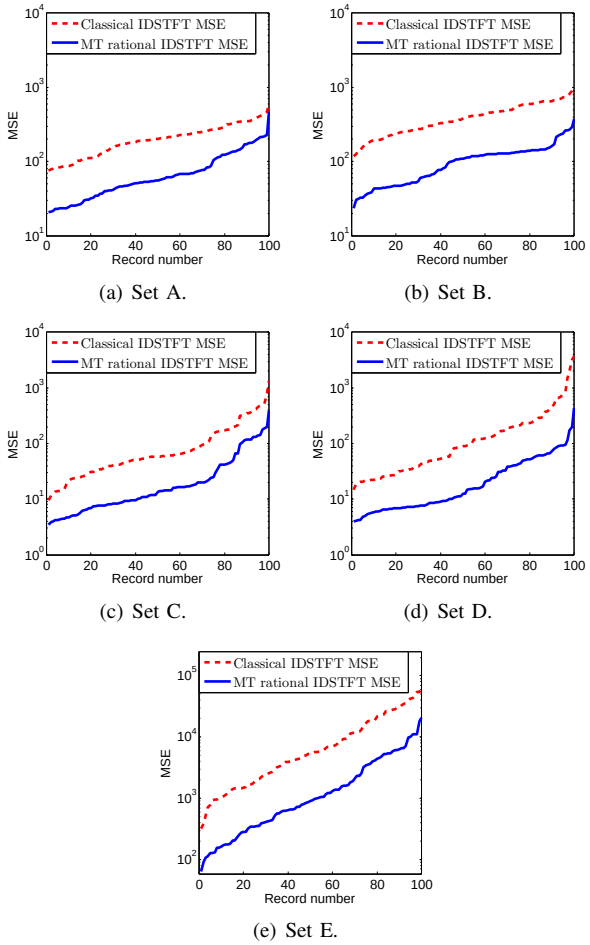
as Gabor limit which is related to the fact that a function and its Fourier transform cannot have compact support at the same time. As a consequence, the higher level is for the frequency resolution, the lower it is for the time localization, and vice versa (see e.g., Chapter 1. in [26]).

Although several procedures have been constructed to cure this problem, there is no clear winner among them. Unfortunately, all time-frequency (*t-f*) distributions suffer from this problem to some extent. In the previous sections, we have already compared the proposed method with the classical STFT where a particular window $g(t)$ had been used for time localization. Here the length of the function $g(t)$ controls the trade-off between the *t-f* resolutions. The WVT was another solution, which does not spread the time or frequency support of sinusoids (see Theorem 4.7. in [26]). Although it gives a very good *t-f* resolution, it can hardly be used in the case of multi-component signals because of the so-called cross terms. However, there is a family of *t-f* representations, which may solve this problem. The Cohens class of distributions [27] is a generalization of the WVT where most of the cross terms are eliminated using different kernels. The related power spectrum can be written in the form

$$P(t,\omega) = \int_{\mathbb{R}} R_t(\tau) e^{-2\pi i \tau \omega} d\tau$$

where $R_t(\tau)$ denotes the convolution of the autocorrelation and the kernel function defined as

$$R_t(\tau) = \int_{\mathbb{R}} r(u-t,\tau)\bar{s}(u-\frac{1}{2}\tau)s(u+\frac{1}{2}\tau)du\,,$$
$$r(u,\tau) = \int_{\mathbb{R}} \phi(\theta,\tau)e^{2\pi i\theta u}d\theta\,.$$




(a) Set A. (b) Set B. (c) Set C. (d) Set D. (e) Set E.

Fig. 6. Mean Square Errors for the Bonn Database.

There is a wide range of applications of these distributions. Additionally, a recent study [28] has already introduced a feature synthesis for epileptic seizure detection in EEG based on these *t-f* distributions. For the sake of completeness, we also tested the proposed algorithm against 13 Cohen's class distributions with different kernels. In order to extract features from each *t-f* representation, we follow the same procedure as suggested in [28]. More precisely, a grid is constructed for each one second long EEG epoch to divide its *t-f* plane into regions. Each region points out a specific frequency sub-band in a certain time window. The frequency sub-bands are: 0-2.5Hz, 2.5-5.5Hz, 5.5-10.5Hz, 10.5-21.5Hz and 21.5-43.5Hz. Moreover, we divide the time axis into three equal windows resulting in a total 15 *t-f* regions. Then the energy of the signal is computed in each region to extract feature vectors. Furthermore, the energy of the signal over the entire segment is also computed as an additional feature resulting in a feature vector with 16 elements. Finally, the feature vectors are fed into the MLP classifier, which was trained in the same way by using $50\%$ of the data.

Table IV shows the results of the six classification tasks using different *t-f* distributions and the comparison with two sets of MT rational DSTFT features, one with similar epoch size and $N$ = 16 coefficients and the other with the optimal setting ($M$ = 256 samples and $N$ = 32 coefficients) determined

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TBME.2014.2360101, IEEE Transactions on Biomedical Engineering



TABLE IV
COMPARISON OF CLASSIFICATION RESULTS OF DIFFERENT *t-f* DISTRIBUTIONS.

| **Distribution** | **Classification task** | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **E-A** | | | **E-B** | | | **E-C** | | | **E-D** | | | **E-A,C** | | | **E-A,B,C,D** | | |
| | **Sens%** | **Spec%** | **Acc%** | **Sens%** | **Spec%** | **Acc%** | **Sens%** | **Spec%** | **Acc%** | **Sens%** | **Spec%** | **Acc%** | **Sens%** | **Spec%** | **Acc%** | **Sens%** | **Spec%** | **Acc%** |
| Born-Jordan | 71.5 | 100.0 | 85.8 | 71.7 | 100.0 | 85.9 | 71.4 | 100.0 | 85.8 | 67.2 | 98.6 | 83.0 | 66.7 | 100.0 | 88.8 | 74.4 | 100.0 | 94.8 |
| Butterworth | 72.0 | 100.0 | 86.1 | 72.1 | 100.0 | 86.1 | 71.9 | 100.0 | 86.0 | 68.0 | 98.6 | 83.4 | 66.8 | 100.0 | 88.8 | 68.7 | 99.4 | 93.2 |
| Choi-Williams | 72.5 | 100.0 | 86.3 | 72.7 | 100.0 | 86.4 | 72.4 | 100.0 | 86.3 | 68.1 | 98.6 | 83.4 | 66.8 | 100.0 | 88.8 | 68.6 | 99.4 | 93.2 |
| Generalized | 72.2 | 100.0 | 86.2 | 73.0 | 100.0 | 86.6 | 72.0 | 100.0 | 86.1 | 67.3 | 98.5 | 83.0 | 67.7 | 88.8 | 89.1 | 67.8 | 99.3 | 93.0 |
| Margenau-Hill | 70.1 | 100.0 | 85.1 | 70.0 | 100.0 | 85.1 | 70.0 | 100.0 | 85.1 | 69.2 | 98.0 | 83.7 | 72.5 | 100.0 | 90.8 | 71.0 | 99.2 | 93.5 |
| Pseudo Margenau-Hill | 100.0 | 0.0 | 49.8 | 100.0 | 0.0 | 49.8 | 100.0 | 0.0 | 49.8 | 69.0 | 95.6 | 82.4 | 70.4 | 100.0 | 90.0 | 69.6 | 98.8 | 92.9 |
| Wigner-Ville | 66.5 | 100.0 | 83.3 | 66.7 | 100.0 | 83.4 | 66.3 | 100.0 | 83.3 | 66.1 | 99.3 | 82.8 | 69.3 | 100.0 | 89.7 | 74.5 | 98.9 | 94.0 |
| Pseudo Wigner-Ville | 71.5 | 100.0 | 85.8 | 71.6 | 100.0 | 85.9 | 71.5 | 100.0 | 85.8 | 66.7 | 98.4 | 82.6 | 68.6 | 100.0 | 89.4 | 69.7 | 99.3 | 93.3 |
| Zhao-Atlas-Mark | 66.0 | 100.0 | 83.1 | 66.0 | 100.0 | 83.1 | 66.0 | 100.0 | 83.1 | 60.3 | 99.1 | 79.8 | 67.9 | 100.0 | 89.2 | 71.4 | 99.5 | 93.8 |
| Rihaczek | 70.1 | 100.0 | 85.1 | 70.0 | 100.0 | 85.1 | 70.0 | 100.0 | 85.1 | 69.2 | 98.0 | 83.7 | 72.5 | 100.0 | 90.8 | 71.0 | 99.2 | 93.5 |
| Page | 70.1 | 100.0 | 85.1 | 70.0 | 100.0 | 85.1 | 70.0 | 100.0 | 85.1 | 69.2 | 98.0 | 83.7 | 72.5 | 100.0 | 90.8 | 71.0 | 99.2 | 93.5 |
| Pseudo Page | 98.1 | 100.0 | 99.1 | 97.2 | 96.1 | 96.7 | 97.6 | 96.6 | 97.1 | 71.6 | 95.4 | 83.6 | 70.4 | 100.0 | 90.0 | 73.1 | 98.7 | 93.6 |
| Pseudo interference | 71.2 | 100.0 | 85.7 | 71.6 | 100.0 | 85.9 | 67.7 | 100.0 | 89.1 | 67.0 | 98.5 | 82.8 | 67.7 | 100.0 | 89.1 | 68.1 | 99.4 | 93.1 |
| MT (16 coeff. 173p) | 99.0 | 99.9 | **99.5** | 98.9 | 98.4 | **98.7** | 96.8 | 97.3 | **97.0** | 94.1 | 95.9 | **95.0** | 98.7 | 96.6 | **98.0** | 99.6 | 87.6 | **97.2** |
| MT (32 coeff. 256p) | 99.9 | 99.6 | **99.8** | 99.6 | 99.0 | **99.3** | 99.3 | 97.7 | **98.5** | 95.6 | 94.1 | **94.9** | 99.2 | 98.9 | **99.1** | 99.2 | 93.8 | **98.1** |

in Section V-B. The classification results of the Cohen's class *t-f* distributions do not differ significantly from each other except for the Pseudo Page distribution. Furthermore, we could achieve the best accuracy in all the binary classification tasks, and the difference is more than $8.3\%$ and $3.3\%$ for last two classification tasks, respectively. Moreover, the proposed method shows a significant improvement in terms of sensitivity and accuracy for the first four classification tasks. We note that the database is very small compared to long term recordings used in practical applications [29]. Additionally, our algorithm uses only a few coefficients of the rational decomposition in contrast to *t-f* features in [28] where the sums of several coefficients were used from particular partitions of the power spectra.

Wavelet based power spectra has been also developed to maintain the trade-off between the *t-f* domains. In this case, the frequency distribution is scaled in a non-linear way. Namely, higher levels possess a fine time resolution, but poor frequency resolutions. This has proved to be a very useful property in many practical applications [30]. Although various types of wavelet functions exist, the decomposition procedure itself generally works with a fixed orthogonal basis. For this reason, the wavelet packets (WP) were introduced by Coifman et al. [31] to form an adaptive wavelet representation. In this case, one can define a particular tiling of the *t-f* plane by selecting the optimal coefficients related to the most suitable basis. Here the optimization means the minimization of the Shannon entropy of the coefficients. We note that both the adaptive representation [10] and the non-uniformly spaced frequency scale [32] are also available with rational functions. Additionally, in our case the adaptation can be further improved by varying the number of different poles at each segment. For localizing the best poles, we can use the dimensional optimization process, i.e., the multi-dimensional (MD) PSO [24]. So our algorithm provides a time varying power spectra, where the density of the frequency distribution changes dynamically.

### *B. The most significant feature element*

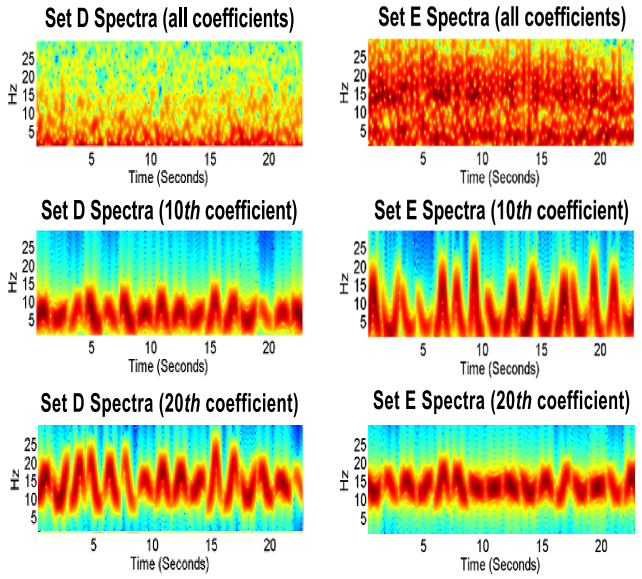
Since our extracted features are coefficients of the *t-f* distribution, finding the most discriminative elements in the feature vector indicates the frequency range where epileptic seizure patterns are mostly active. In order to determine the most significant components in the feature vector, we sort the feature elements in a supervised manner via SVM. Using the whole feature vector of the training set, the elements can be ranked by using a recursive leave-one-out procedure according to the weights of the support vectors [33]. One can see the top five most significant elements for each classification task in Table V. For five classification tasks, the most discriminative component of the feature vector is the median value of the coefficients. In addition, the most common elements in the top five ranking list are the $10th$, $11th$ and $12th$ coefficients. Thus, it is expected that the seizure patterns of set E should be more active in a specific frequency sub-band. In order to illustrate this phenomenon the spectrograms of sets E and D were displayed in Fig. 7. As it can be seen, the spectrum of set E possesses a higher magnitude for the theta and alpha frequency

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TBME.2014.2360101, IEEE Transactions on Biomedical Engineering



TABLE V
TOP FIVE MOST SIGNIFICANT FEATURE ELEMENTS FOR EACH CLASSIFICATION TASK

| | Classification task | | | | | |
|---|---|---|---|---|---|---|
| | E-A | E-B | E-C | E-D | E-A,C | E-A,B,C,D |
| # top 5 | 36 | 36 | 36 | 20 | 36 | 36 |
| | 12 | 5 | 18 | 19 | 10 | 10 |
| | 10 | 8 | 11 | 21 | 37 | 11 |
| | 11 | 7 | 16 | 18 | 12 | 32 |
| | 5 | 11 | 19 | 31 | 16 | 9 |




(a) Spectograms of set D. (b) Spectograms of set E.

Fig. 7. Spectrograms of sets D and E by using different number of rational components.

sub-bands in contrast with the spectrum of set D. Furthermore, the power spectrum of the $10th$ rational component $(c_{10}\Phi_{10})$ shows a noticeable visual difference in the theta sub-band. Namely, seizure patterns of set E are mostly active in a wider frequency range comparing to set D. Comparison of the power spectra of the $20th$ rational component $(c_{20}\Phi_{20})$ of sets D and E shows that the dominant seizure activity of set E is concentrated in the alpha frequency sub-band. This explains the reason why the $20th$ coefficient is ranked as the most significant feature element in the discrimination of sets E and D.

### C. Comparison with state-of-the-art methods

As a final performance assessment, we compare our method with other state-of-the-art feature extraction techniques proposed for seizure classification in the literature. These methods includes: 1) Wavelet analysis which decomposes an EEG signal into five frequency sub-bands by using second-order Daubechies (db2) wavelet transform. The feature vector consists of the maximum, minimum, average and standard deviation of the wavelet coefficients of each sub-band [34], [35]; 2) Entropy based feature extraction using Shannon entropy, log-energy entropy and sample entropy [36]; 3) Spectral analysis applying Welch FFT Power Spectral Density (PSD) [1]; 4) Approximate entropy based feature extraction [37]. For all of these methods, the same MLP classifier was used by feeding $50\%$ of the data in the learning phase. As it can be seen in Table VI, our method is mostly ranked as the first or the second best approach in terms of classification accuracy. It can outperform both the entropy and the approximate entropy based feature extraction methods in all classification tasks. Besides, due to consideration of all coefficients in Welch PSD based method, a large dimensional feature vector is obtained in contrast with our method which uses only $32 + 5$ features. For this reason, training the classifier by the proposed compact features is faster. Our approach is ranked first together with the wavelet based method on average. Moreover, the high accuracy level obtained by the proposed method in the classification task 6 demonstrates its ability to discriminate between epileptic seizure and seizure-free epochs in the presence of an imbalanced EEG data.

Finally, we compare the proposed algorithm with the results of some state-of-the-art epilepsy detection methods reported in the literature using the same dataset. We consider the classification accuracies of the same classification tasks, which are summarized in Table VII. For the classification tasks (E-B) and (E-C), our method achieved the best results with a significant improvement comparing to the others. It is ranked second best for the classification task (E-D) where the difference is about $0.6\%$. In addition, the proposed method is close to the best scores in terms of the highest classification accuracy where the difference is $0.2\%$ and $2\%$ for the classification tasks (E-A) and (E-A,B,C,D), respectively. The deviation is caused by the high variance between the pole(s) and the coefficients. More precisely, in each segment the optimal value of the pole is selected via a stochastic search method using the hyperbolic modification of the basic PSO algorithm [25]. Furthermore, altering the pole(s) changes the basis functions as well. This procedure can affect the variance of the extracted coefficients, which can cause variations in the classification results. In order to address this problem, a higher number of coefficients can be extracted and fed to the classifier which will attempt to fit better the underlying distribution. In addition, a more sophisticated classification approach such as ensemble learning with bagging techniques can be considered. Such techniques may reduce the variance between the poles and coefficients by using, e.g. bootstrap replicates.

## VII. CONCLUSIONS

The main contribution of this paper is to provide an effective and novel feature extraction algorithm for epileptic seizure detection in EEG signals. We proposed the rational DSTFT which is an adaptive generalization of the classical STFT. Additionally, we described the epileptic seizure patterns in the *t-f* domain by using different types of rational functions. In order to increase the efficiency, we investigated both the optimal number of coefficients and the best window size. Moreover, we analyzed the proposed features in order to determine the most significant ones. The performance of different rational systems was compared with the classical

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TBME.2014.2360101, IEEE Transactions on Biomedical Engineering



TABLE VI
CLASSIFICATION RESULTS OBTAINED BY DIFFERENT FEATURE EXTRACTION METHODS USING MLP CLASSIFIER.

| **Classification task** | | **Feature extraction** | | | | |
|---|---|---|---|---|---|---|
| | | Wavelet Analysis | Entropy Based | Welch PSD | Approximate Entropy | MT rational DSTFT |
| E-A | Sens% | 100.0 | 92.1 | 98.7 | 87.3 | 99.9 |
| | Spec% | 100.0 | 99.9 | 100.0 | 89.6 | 99.6 |
| | **Acc%** | **100.0** | 96.0 | 99.4 | 88.4 | 99.8 |
| E-B | Sens% | 99.2 | 90.5 | 98.2 | 91.1 | 99.6 |
| | Spec% | 99.1 | 99.4 | 100.0 | 94.0 | 99.0 |
| | **Acc%** | 99.2 | 94.9 | 99.1 | 92.6 | **99.3** |
| E-C | Sens% | 98.7 | 89.6 | 92.3 | 0.0 | 99.3 |
| | Spec% | 97.9 | 97.6 | 99.9 | 100.0 | 97.7 |
| | **Acc%** | 98.3 | 93.6 | 96.1 | 50.3 | **98.5** |
| E-D | Sens% | 95.6 | 85.2 | 86.7 | 2.8 | 95.6 |
| | Spec% | 97.4 | 94.4 | 99.6 | 99.4 | 94.1 |
| | **Acc%** | **96.5** | 89.8 | 93.2 | 51.3 | 94.9 |
| E-A, C | Sens% | 97.5 | 93.6 | 98.3 | 0.0 | 99.2 |
| | Spec% | 99.7 | 98.0 | 99.7 | 100.0 | 98.9 |
| | **Acc%** | 99.0 | 96.5 | **99.3** | 66.3 | 99.1 |
| E-A, B, C, D | Sens% | 96.0 | 90.3 | 95.0 | 0.0 | 99.2 |
| | Spec% | 98.7 | 96.7 | 98.7 | 100.0 | 93.8 |
| | **Acc%** | **98.1** | 95.5 | 97.9 | 79.8 | **98.1** |

TABLE VII
THE COMPARISON OF THE PROPOSED ALGORITHM WITH STATE-OF-THE-ART METHODS FOR THE SAME CLASSIFICATION TASKS PERFORMED ON THE BONN EEG DATASET

| **Author(s)** | **Classification task** | **Accuracy (%)** |
|---|---|---|
| Srinivasan et al. (2005) [38] | E-A | 99.6 |
| Subasi (2007) [39] | E-A | 95 |
| Parvin Kumar et al. (2010) [40] | E-A | 96 - 99.8 |
| Guo et al. (2009) [41] | E-A | 95.2 |
| Nicoletta and Georgiou (2012) [42] | E-A | 93.5 |
| Xie and Krishnan (2013) [43] | E-A | 100 |
| Kaya et al. (2014) [44] | E-A | 99.5 |
| **Proposed method** | E-A | 99.8 |
| Nicoletta and Georgiou (2012) [42] | E-B | 82.9 |
| **Proposed method** | E-B | 99.3 |
| Nicoletta and Georgiou (2012) [42] | E-C | 88.0 |
| **Proposed method** | E-C | 98.5 |
| Nicoletta and Georgiou (2012) [42] | E-D | 79.9 |
| Kaya et al. (2014) [44] | E-D | 95.5 |
| **Proposed method** | E-D | 94.9 |
| Tzallas et al. (2007) [45] | E-A, B, C, D | 97.7 |
| Ocak (2008) [46] | E-A, B, C, D | 96.2 |
| Sheng-Fu Liang et al. (2010) [47] | E-A, B, C, D | 97.8 - 98.5 |
| Guo et al. (2010) [48] | E-A, B, C, D | 98.3 |
| Xie and Krishnan (2013) [49] | E-A, B, C, D | 100 |
| **Proposed method** | E-A, B, C, D | 98.1 |

STFT in terms of classification accuracy. Furthermore, we increased the adaptability of our method by applying the stochastic hyperbolic PSO algorithm to determine the optimal position of the pole for each EEG epoch. For this reason, the proposed method gives a compact *t-f* representation of the signal. We also showed that the inverse rational DSTFT achieves a reconstruction signal with a smaller mean square error compared to the classical STFT for the same number of coefficients. Finally, we compared the performance of our feature extraction procedure with 13 Cohen's class *t-f* distributions. Several state-of-the-art feature extraction methods were taken into account, which are commonly used in epileptic seizure detection and classification. The comparative study shows that the proposed rational DSTFT can be used as a robust feature extraction method in off-line epilepsy analysis and detection. In addition, the proposed algorithm uses results from the recently developed theory of signal models for rational orthogonal systems. As confirmed by experiments, the MT rational DSTFT can be successfully applied to perform epileptic seizure pattern detection in EEGs.

We note that, the adaptability of the system can be further improved by using multi-dimensional (MD) PSO [24], [50] to determine the optimum number of unique poles. MDPSO can explore multi-dimensional search spaces while it will converge

to the best pole configuration in the optimal dimension. In this latter case, we can describe each segment with a set of poles and multiplicities.

By taking the advantage of the adaptive behavior of the rational representations, we expect that our method can perform well on longer EEG recordings as well. Hence, future work will focus on the use of rational DSTFT in multichannel long-term epileptic seizure detection in order to demonstrate the suitability of the proposed method in the presence of high frequency EEG activities (i.e., high gamma $>$50Hz) as was suggested in [51]. For this purpose, we shall test our future algorithms on larger datasets such as Physionet [29] that contains more variability between seizure and seizure-free epochs. Furthermore, using long-term EEG records, we can evaluate the performance of future methods for real-time seizure detection by measuring the delay of detected seizures onset.

## REFERENCES


[1] K. Polat and S. Güneş, "Classification of epileptiform EEG using a hybrid system based on decision tree classifier and fast Fourier transform," *Applied Mathematics and Computation*, vol. 187, no. 2, pp. 1017–1026, Apr. 2007.

[2] K. Polat and S. Güneş, "Artificial immune recognition system with fuzzy resource allocation mechanism classifier, principal component analysis and FFT method based new hybrid automated identification system for classification of EEG signals," *Expert Systems with Applications*, vol. 34, no. 3, pp. 2039–2048, Apr. 2008.

[3] A. J. Gabor, R. R. Leach, and F. U. Dowla, "Automated seizure detection using a self-organizing neural network," *Electroencephalography and Clinical Neurophysiology*, vol. 99, no. 3, pp. 257–266, Sept. 1996.

[4] A. J. Gabor, "Seizure detection using a self-organizing neural network: validation and comparison with other detection strategies," *Electroencephalography and Clinical Neurophysiology*, vol. 107, no. 1, pp. 27–32, 1998.

[5] T. Kohonen, *Self-organization and associative memory*, Springer-Verlag, New York, USA, 3rd edition, 1989.

[6] S. Ghosh-Dastidar, H. Adeli, and N. Dadmehr, "Mixed-band wavelet-chaos-neural network methodology for epilepsy and epileptic seizure detection," *IEEE Transactions on Biomedical Engineering*, vol. 54, no. 9, pp. 1545–1551, Sept. 2007.

[7] H. Adeli, Z. Zhou, and N. Dadmehr, "Analysis of EEG records in an epileptic patient using wavelet transform," *Journal of Neuroscience Methods*, vol. 123, no. 1, pp. 69–87, Feb. 2003.

[8] L. Chen, E. Zhao, D. Wang, Z. Han, S. Zhang, and C. Xu, "Feature extraction of EEG signals from epilepsy patients based on Gabor transform and EMD decomposition," in *Proceedings of the 6th International Conference on Natural Computation (ICNC)*, Aug. 2010, vol. 3, pp. 1243–1247.

[9] S. Nasehi and H. Pourghassem, "Real-time seizure detection based on EEG and ECG fused features using Gabor functions," in *Proceedings of the International Conference on Intelligent Computation and Bio-Medical Instrumentation (ICBMI)*, 2011, pp. 204–207.

[10] S. Fridli, L. Lócsi, and F. Schipp, "Rational function system in ECG processing," in *Computer Aided Systems Theory–EUROCAST 2011: Part I*, R. Moreno-Díaz et al., Ed., vol. 6927 of *LNCS*, pp. 88–95. Springer-Verlag Berlin, Heidelberg, Germany, 2012.

[11] L. Lócsi and P. Kovács, "Processing ECG signals using rational function systems," in *Proceedings of the 7th IEEE International Symposium on Medical Measurements and Applications (MeMeA)*, 2012, pp. 1–5.

[12] S. Fridli, P. Kovács, L. Lócsi, and F. Schipp, "Rational modeling of multi-lead QRS complexes in ECG signals," *Annales Univ. Sci. Budapest., Sect. Comp*, vol. 37, pp. 145–155, Oct. 2012.

[13] P. Kovács, K. Samiee, and M. Gabbouj, "On application of rational discrete short time fourier transform in epileptic seizure classification," in *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2014, pp. 5839–5843.

[14] "Rational DSTFT Toolbox, optimized for EEG classification," 2014, [Online]. Available: http://www.cs.tut.fi/∼samiee/ [Accessed: Aug. 18, 2014].

[15] 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," *Physical Review E*, vol. 64, no. 6, 1, pp. 1–8, Nov. 2001.

[16] S. Fridli and F. Schipp, "Biorthogonal systems to rational functions," *Annales Univ. Sci. Budapest., Sect. Comp*, vol. 35, pp. 95–105, Nov. 2011.

[17] P. S. C. Heuberger, P. M. J. Van den Hof, and B. Wahlberg, *Modelling and Identification with Rational Orthogonal Basis Functions*, Springer-Verlag, London, UK, 2005.

[18] S. K. Mitra, *Digital Signal Processing: A Computer-Based Approach*, McGraw-Hill, New York, USA, 4th edition, 2011.

[19] L. Lócsi, "Calculating non-equidistant discretizations generated by Blaschke products," *Acta Cybernetica*, vol. 20, pp. 111–123, Jan. 2011.

[20] P. Kovács and V. Vad, "Fast computing of non-uniform sampling positions for real signals," in *Proceedings of the 15th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC)*, 2013, pp. 146–150.

[21] P. Kovács and L. Lcsi, "RAIT: the rational approximation and interpolation toolbox for matlab, with experiments on ECG signals," *International Journal of Advances in Telecommunications, Electrotechnics, Signals and Systems*, vol. 1, no. 2-3, pp. 67–75, Nov. 2012.

[22] A. T. Tzallas, M. G. Tsipouras, D. G. Tsalikakis, E. C. Karvounis, L. Astrakas, S. Konitsiotis, and M. Tzaphlidou, "Automated epileptic seizure detection methods: A review study," in *Epilepsy - Histological, Electroencephalographic and Psychological Aspects*. InTech, Feb. 2012.

[23] J. Kennedy and R. C. Eberhart, "Particle swarm optimization," in *Proceedings of IEEE International Conference on Neural Networks*, 1995, vol. 4, pp. 1942–1948.

[24] S. Kiranyaz, T. Ince, and M. Gabbouj, *Multidimensional Particle Swarm Optimization for Machine Learning and Pattern Recognition*, vol. 15 of *Adaptation, Learning, and Optimization*, Springer-Verlag Berlin, Heidelberg, Germany, 2014.

[25] P. Kovács, S. Kiranyaz, and M. Gabbouj, "Hyperbolic particle swarm optimization with application in rational identification," in *Signal Processing Conference (EUSIPCO), 2013 Proceedings of the 21st European*, Sept. 2013, pp. 1–5.

[26] S. Mallat, *A Wavelet Tour of Signal Processing: The Sparse Way*, Academic Press, Burlington, USA, 3rd edition, 2008.

[27] L. Cohen, "Time frequency distributions-a review," *Proceedings of the IEEE*, vol. 77, no. 7, pp. 941–981, July 1989.

[28] A. T. Tzallas, M. G. Tsipouras, and D. I. Fotiadis, "Epileptic seizure detection in eegs using time–frequency analysis," *IEEE Transactions on Information Technology in Biomedicine*, vol. 13, no. 5, pp. 703–710, Sept. 2009.

[29] A. L. Goldberger, L. A. N. Amaral, L. Glass, J. M. Hausdorff, P. Ch. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C. K. Peng, and H. E. Stanley, "PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, pp. 215–220, June 2000.

[30] P. S. Addison, *The Illustrated Wavelet Transform Handbook – Introductory Theory and Applications in Science, Engineering, Medicine and Finance*, Institute of Physics Publishing, Bristol, UK, 2002.

[31] R. R. Coifman, Y. Meyer, and M. V Wickerhauser, "Wavelet analysis and signal processing," in *Wavelets and Their Applications*, M. B. Ruskai et al., Ed., pp. 153–178. Jones and Bartlett, Boston, USA, 1992.

[32] A. Soumelidis, F. Schipp, and J. Bokor, "Frequency domain representation of signals in rational orthogonal bases," in *Proceedings of the 10th IEEE Mediterranean Conference on Control and Automation (MED)*, 2002.

[33] I. Guyon, J. Weston, S. Barnhill, and V. Vapnik, "Gene selection for cancer classification using support vector machines," *Machine Learning*, vol. 46, no. 1-3, pp. 389–422, Jan. 2002.

[34] İ. Güler and E. D. Übeyli, "Adaptive neuro-fuzzy inference system for classification of EEG signals using wavelet coefficients," *Journal of Neuroscience Methods*, vol. 148, no. 2, pp. 113 – 121, Nov. 2005.

[35] K. A. H. Kulasuriya and M. U. S. Perera, "Forecasting epileptic seizures using EEG signals, wavelet transform and artificial neural networks," in *Proceedings of the International Symposium on IT in Medicine and Education (ITME)*, 2011, vol. 1, pp. 557–562.

[36] S. Aydin, H. M. Saraoğlu, and S. Kara, "Log energy entropy-based EEG classification with multilayer neural networks in seizure," *Annals of Biomedical Engineering*, vol. 37, no. 12, pp. 2626–2630, Dec. 2009.

[37] V. Srinivasan, C. Eswaran, and N. Sriraam, "Approximate entropy-based epileptic EEG detection using artificial neural networks," *IEEE Transactions on Information Technology in Biomedicine*, vol. 11, no. 3, pp. 288–295, May 2007.

[38] V. Srinivasan, C. Eswaran, and N. Sriraam, "Artificial neural network based epileptic detection using time-domain and frequency-domain features," *Journal of Medical Systems*, vol. 29, no. 6, pp. 647–660, Dec. 2005.

[39] A. Subasi, "EEG signal classification using wavelet feature extraction and a mixture of expert model," *Expert Systems with Applications*, vol. 32, no. 4, pp. 1084–1093, May 2007.

[40] S. Pravin Kumar, N. Sriraam, P.G. Benakop, and B.C. Jinaga, "Entropies based detection of epileptic seizures with artificial neural network classifiers," *Expert Systems with Applications*, vol. 37, no. 4, pp. 3284–3291, Apr. 2010.

[41] L. Guo, D. Rivero, J. A. Seoane, and A. Pazos, "Classification of EEG signals using relative wavelet energy and artificial neural networks," in *Proceedings of the First ACM/SIGEVO Summit on Genetic and Evolutionary Computation*, New York, NY, USA, 2009, GEC '09, pp. 177–184, ACM.

[42] N. Nicolaou and J. Georgiou, "Detection of epileptic electroencephalogram based on permutation entropy and support vector machines," *Expert Systems with Applications*, vol. 39, no. 1, pp. 202–209, Jan. 2012.

[43] S. Xie, S. Shan, X. Chen, and W. Gao, "V-LGBP: Volume based local Gabor binary patterns for face representation and recognition," in *Proceedings of the 19th International Conference on Pattern Recognition (ICPR)*, 2008, pp. 1–4.

[44] Y. Kaya, M. Uyar, R. Tekin, and S. Yldrm, "1d-local binary pattern based feature extraction for classification of epileptic EEG signals," *Applied Mathematics and Computation*, vol. 243, pp. 209–219, 2014.

[45] A. T. Tzallas, M. G. Tsipouras, and D. I. Fotiadis, "Automatic seizure detection based on time-frequency analysis and artificial neural networks," *Computational Intelligence and Neuroscience*, vol. 2007, pp. 1–13, 2007.

[46] H. Ocak, "Optimal classification of epileptic seizures in EEG using wavelet analysis and genetic algorithm," *Signal Processing*, vol. 88, no. 7, pp. 1858–1867, July 2008.

[47] C.-P. Shen, C.-M. Chan, F.-S. Lin, M.-J. Chiu, J.-W. Lin, J.-H. Kao, C.-P. Chen, and F. Lai, "Epileptic seizure detection for multichannel EEG signals with support vector machines," in *Proceedings of the 11th IEEE International Conference on Bioinformatics and Bioengineering*, 2011, pp. 39–43.

[48] L. Guo, D. Rivero, and A. Pazos, "Epileptic seizure detection using multiwavelet transform based approximate entropy and artificial neural networks," *Journal of Neuroscience Methods*, vol. 193, no. 1, pp. 156–163, Oct. 2010.

[49] S. Xie and S. Krishnan, "Wavelet-based sparse functional linear model with applications to EEGs seizure detection and epilepsy diagnosis," *Medical & Biological Engineering & Computing*, vol. 51, no. 1–2, pp. 49–60, 2013.

[50] S. Kiranyaz, J. Pulkkinen, A. Yildirim, and M. Gabbouj, "Multi-dimensional particle swarm optimization in dynamic environments," *Expert Systems with Applications*, vol. 38, no. 3, pp. 2212–2223, Mar. 2011.

[51] G. A. Worrell, L. Parish, S. D. Cranstoun, R. Jonas, G. Baltuch, and B. Litt, "Highfrequency oscillations and seizure generation in neocortical epilepsy," *Brain*, vol. 127, no. 7, pp. 1496–1506, July 2004.