

Classification of EEG signals for detection of epileptic seizures based on wavelets and statistical pattern recognition

D. Gajic, Z. Djurovic, S. Di Gennaro and Fredrik Gustafsson

Linköping University Post Print



N.B.: When citing this work, cite the original article.

Original Publication:

D. Gajic, Z. Djurovic, S. Di Gennaro and Fredrik Gustafsson, Classification of EEG signals for detection of epileptic seizures based on wavelets and statistical pattern recognition, 2014, Biomedical Engineering: Applications, Basis and Communications, (26), 2, 1450021.

<http://dx.doi.org/10.4015/S1016237214500215>

Copyright© 2014 World Scientific Publishing Co.

<http://www.worldscientific.com/>

Postprint available at: Linköping University Electronic Press

<http://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-110540>

Classification of EEG Signals for Detection of Epileptic Seizures Based on Wavelets and Statistical Pattern Recognition

Dragoljub Gajic,^{1,2,*} Zeljko Djurovic,¹ Stefano Di Gennaro,² Fredrik Gustafsson³

¹*Department of Control Systems and Signal Processing, School of Electrical Engineering,
University of Belgrade, Serbia*

²*Department of Information Engineering, Computer Science and Mathematics, University of
L'Aquila, Italy*

³*Department of Electrical Engineering, Linköping University, Sweden*

*Correspondence: dragoljubgajic@gmail.com

Abstract: The electroencephalogram (EEG) signal is very important in the diagnosis of epilepsy. Long-term EEG recordings of an epileptic patient contain a huge amount of EEG data. The detection of epileptic activity is, therefore, a very demanding process that requires a detailed analysis of the entire length of the EEG data, usually performed by an expert. This paper describes an automated classification of EEG signals for the detection of epileptic seizures using wavelet transform and statistical pattern recognition. The decision making process is comprised of three main stages: (a) feature extraction based on wavelet transform, (b) feature space dimension reduction using scatter matrices, and (c) classification by quadratic classifiers. The proposed methodology was applied on EEG data sets that belong to three subject groups: a) healthy subjects, b) epileptic subjects during a seizure-free interval, and c) epileptic subjects during a seizure. An overall classification accuracy of 99% was achieved. The results confirmed that the proposed algorithm has a potential in the classification of EEG signals and detection of epileptic seizures, and could thus further improve the diagnosis of epilepsy.

Keywords: Epilepsy Diagnosis, Seizure Detection, Scatter Matrices, Dimension Reduction, Quadratic Classifiers

1. Introduction

Epilepsy is a common brain disorder that, according to an estimate of the World Health Organization, affects almost 60 million people around the world. Approximately one in every 100 persons will experience a seizure at some time in their life [1]. Epilepsy is characterized by the recurrent and sudden incidence of epileptic seizures which can lead to dangerous and possibly life-threatening situations [2]. The seizures are the result of a transient and unexpected electrical disturbance of the brain and excessive neuronal discharge that is evident in the electroencephalogram (EEG) signal representative of the electrical activity of the brain. Consequently, the EEG signal has been the most utilized signal in clinical assessments of the

state of the brain and detection of epileptic seizures, and is very important for a proper diagnosis of epilepsy.

The detection of epileptic seizures by visual scanning of a patient's EEG data usually collected over a few days is a tedious and time-consuming process. In addition, it requires an expert to analyze the entire length of the EEG recordings, in order to detect epileptic activity. A reliable automatic classification and detection system would ensure an objective and facilitating treatment and significantly improve the diagnosis of epilepsy as well as long-term monitoring and treatment of patients. For example, long-term treatment with antiepileptic drugs, which may cause cognitive or other neurological side effects, could be reduced to a targeted short-acting intervention [3]. Therefore, there is a strong demand for the development of such automated systems, due to both huge amounts and increased usage of long-term EEG recordings for proper evaluation and treatment of neurological diseases, including epilepsy. The possibility of the expert misreading the data and failing to make a proper decision would also be narrowed down [4, 5].

Many automated EEG signal classification and seizure detection systems, using different approaches, have emerged in recent years. Among such studies, Gotman [6] presented a computerized system for detecting a variety of seizures, while Qu and Gotman [7] proposed the use of the nearest-neighbor classifier on EEG features extracted in both time and frequency domains to detect the onset of epileptic seizures. Gigola *et al.* [8] applied a method based on the evolution of accumulated energy using wavelet analysis for the prediction of epileptic seizure onset from intracranial epileptic EEG recordings, while Adeli *et al.* [9], Guler *et al.* [10] and Ubeyli *et al.* [11] discussed the potential of nonlinear time series analysis in seizure detection.

Artificial neural network-based detection systems for diagnosis of epilepsy have been proposed by several researchers [11, 12, 13]. The method put forward by Weng and Khorasani [14] uses the features proposed by Gotman and Wang [15], namely, average EEG amplitude, average EEG duration, variation coefficient, dominant frequency and average power spectrum, as inputs to an adaptive structured neural network. The method proposed by Pradhan *et al.* [16] uses a

raw EEG signal as an input to a learning vector quantization network. Nigam and Graupe [17] proposed a new neural network model called LAMSTAR (large memory storage and retrieval) network and two time-domain attributes of EEG; namely, relative spike amplitude and spike rhythmicity have been used as inputs for the purpose of detecting seizures. The algorithm proposed by Kiymiket *al.* [18] uses a back propagation neural network with periodogram and autoregressive (AR) features as inputs for automated detection of epileptic seizures. Ghosh-Dastidar *et al.* [19] discussed a classification methodology based on wavelet analysis and both radial basis function and Levenberg-Marquardt backpropagation neural network. Srinivasan *et al.* [20] presented an algorithm based on approximate entropy as an input to an artificial neural network classifier, while Subasi [21] used wavelet analysis and mixture of experts, in addition to the artificial neural network, to classify EEG signals and detect seizures.

Due to quite a low understanding of the mechanisms underlying the problem, most existing methods suffer from low accuracy, a high rate of false alarms and missed detections [22]. In addition, due to a lack of reliable standardized data, most reported EEG analysis-based algorithms are performed on a small number of datasets, which often demonstrate good accuracy for selected EEG segments but are not robust enough to adjust to EEG variations commonly encountered in a hospital setting [20]. In this research, however, a larger number of EEG data sets, which belong to three subject groups, were used: a) healthy subjects (normal EEG), b) epileptic subjects during a seizure-free interval (interictal EEG), and c) epileptic subjects during a seizure (ictal EEG). The EEG signal classification and seizure detection problem was modeled as a three-group classification problem that could be of great clinical significance. An automated system able to accurately differentiate between normal and interictal EEG signals can be used to diagnose epilepsy, while a system that can accurately differentiate between interictal and ictal EEG signals can be used to detect seizures in a clinical setting. Therefore, the classification algorithm must be able to classify all three groups accurately and at the same time be robust with respect to EEG signal variations across various mental states and subjects. The improvement of the classification accuracy is mainly based on the design of both

an appropriate feature space, by identifying combinations of all extracted features that increase the inter-class separation, and classifiers that can accurately classify all three groups of EEG signals based on the selected and reduced feature space. Real EEG recordings were applied to test algorithm performance and the results indicated that the algorithm has a potential to be applied within an automatic epilepsy diagnosis system.

2. Materials and Methods

2.1. Materials

The EEG data used were a subset of EEG data corresponding to both normal and epileptic subjects, made available by Dr. Ralph Andrzejak from the Epilepsy Centre at the University of Bonn [23]. Three EEG data sets from three different groups were analyzed: healthy subjects with normal EEG data, epileptic subjects during a seizure-free interval with interictal EEG data, and epileptic subjects during a seizure with ictal (epileptic) EEG data. Each data set recorded with a 128-channel amplifier system contained 100 single-channel EEG segments sampled at 173.61 Hz, each of 23.6 sec duration. These segments were selected and cut out from the continuous multi channel EEG recordings after visual inspection for artifacts (e.g. due to muscle activity or eye movement). In addition, the segments had to fulfill a stationarity criterion described in detail in Andrzejak *et al.* [23]. The first EEG data set corresponding to healthy subjects was taken from the surface EEG recordings of five healthy subjects, who were relaxed in an awaken state, using the standardized electrode placement technique. The second and third data sets obtained from five different epileptic subjects during a seizure-free and seizure interval, respectively, were taken from the intracranial EEG recordings during presurgical diagnosis.

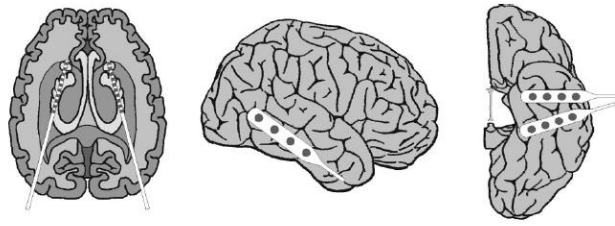


Figure 1 - Implanted intracranial electrodes.

The type of epilepsy was diagnosed as temporal lobe epilepsy with the epileptogenic focus being the hippocampal formation. A schematic of intracranial electrode placement is shown in Fig. 1 [23]. The depth electrodes were implanted symmetrically into the hippocampal formations and the strip electrodes were implanted onto the lateral and basal regions of the neocortex. The EEG segments were selected from all the recording sites exhibiting ictal activity. Each EEG segment was considered as a separate EEG signal resulting in a total of 300 EEG data segments.

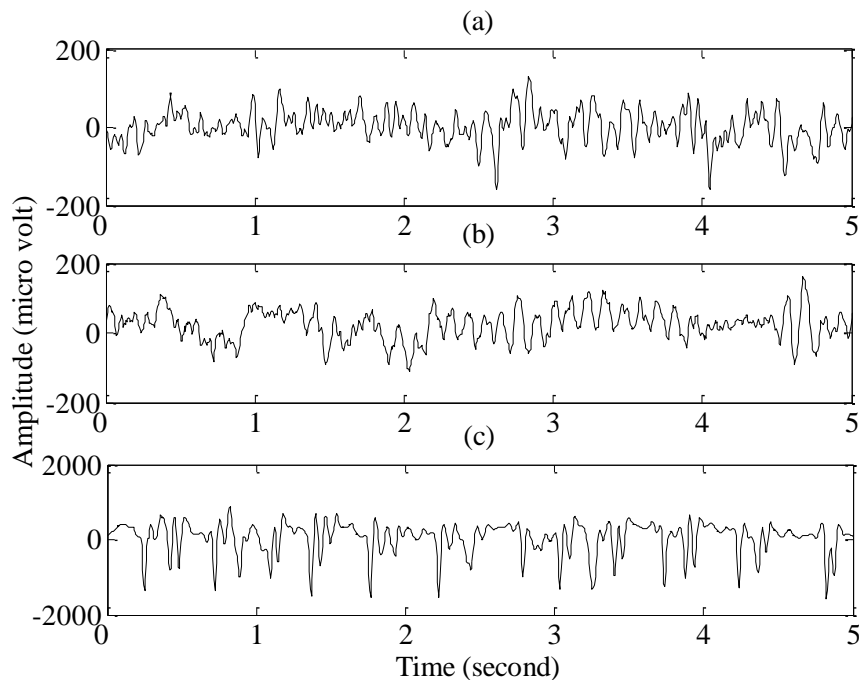


Figure 2 - Segments of EEG data: (a) Normal, (b) Interictal, (c) Ictal.

As an example, the first five seconds of all three different EEG data segments are magnified and shown in Fig. 2. Interictal EEG data can contain only occasional transient waveforms, as isolated spikes, spike trains, sharp waves or spike-wave complexes, while ictal EEG data are composed of a continuous discharge of polymorphic waveforms of variable amplitude and frequency, spike and sharp wave complexes, rhythmic hypersynchrony, or electrocerebral inactivity observed over a duration longer than the average duration of these abnormalities during interictal periods, as shown in Fig. 2. [24].

There are five broad spectral 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). These five frequency sub-bands provide more accurate information about neuronal activities underlying the problem and, 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. That was the basic premise of this research. Most of the features were extracted from each sub-band separately, after wavelet decomposition of the full-spectrum EEG signal, as well as reconstructed in all five sub-bands using the inverse wavelet transform. For example, the difference between normal and interictal EEG data is more apparent in Fig. 3, where only theta sub-bands are presented, than in Fig. 2 where the same but full-spectrum signals are shown. On the other hand, ictal EEG data are easier to distinguish, mainly due to higher amplitudes.

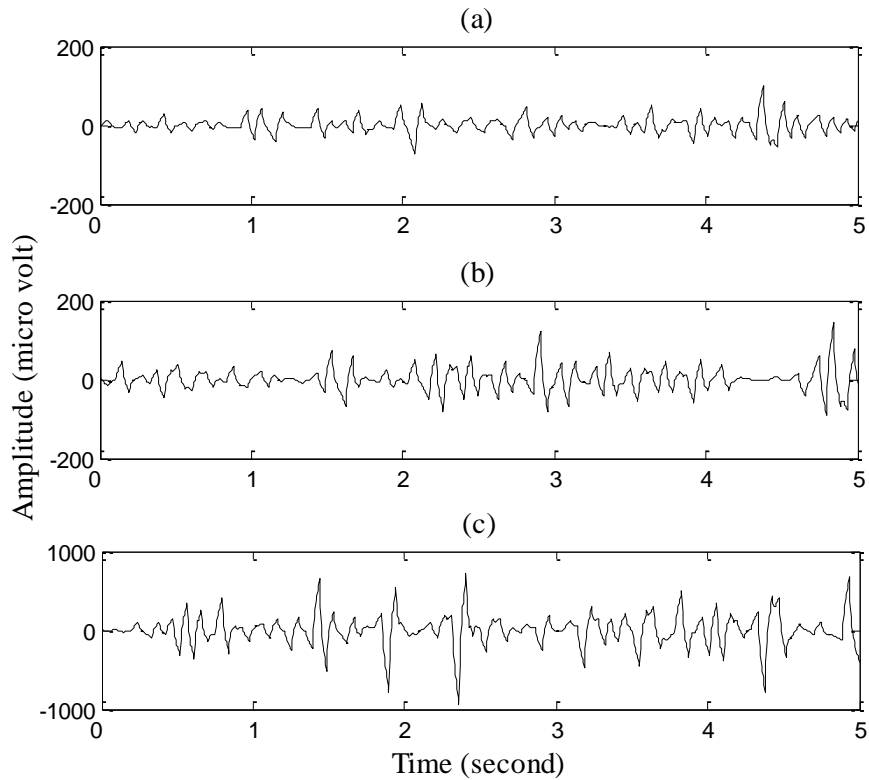


Figure 3 - Theta sub-bands of EEG data: (a) Normal, (b) Interictal, (c) Ictal.

2.2. Methods

An automated classification of EEG signals for the detection of epileptic seizures based on wavelet transform and statistical pattern recognition is proposed. The first step of this method is to obtain a set of features after wavelet transform of EEG data, including energy, entropy, and standard deviation of both wavelet coefficients and the EEG signal in different frequency bands of clinical interest. The second step is to perform dimension reduction of the feature space using scatter matrices. Finally, two quadratic classifiers are designed, which are able to distinguish all three groups of the EEG signals of interest from each other. The entire structure of the algorithm is shown in Fig. 4.

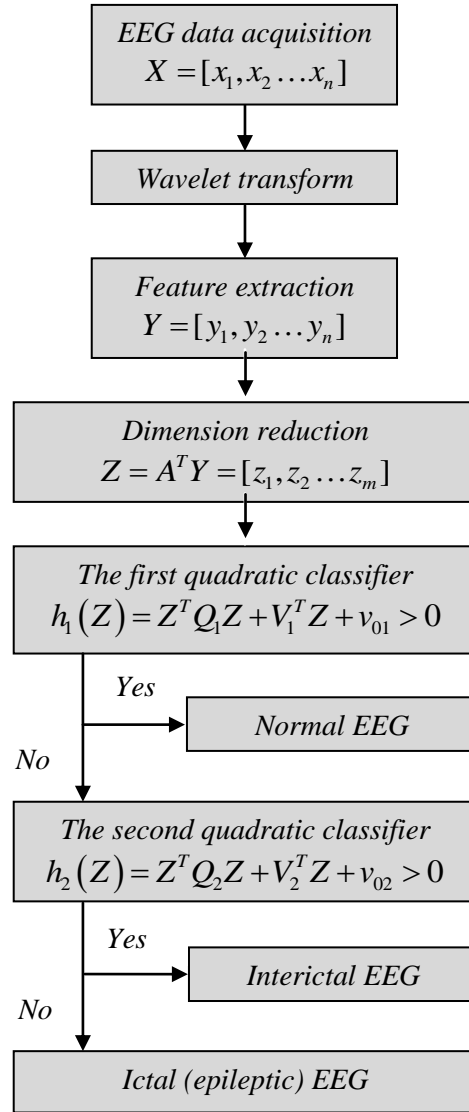


Figure 4 - Proposed classification algorithm.

2.2.1. Wavelet transform

Abnormalities in EEG data during serious neurological diseases such as epilepsy are too subtle to be detected using conventional techniques that usually transform mostly qualitative diagnostic criteria into a more objective quantitative signal feature classification problem. The techniques that have been applied to address this problem include the analysis of EEG signals for the detection of epileptic seizures using the autocorrelation function, time domain features, frequency domain features, time frequency analysis, nonlinear time series analysis, and wavelet

transform. However, the results of various studies have demonstrated that the wavelet transform is the most promising method for extracting features from EEG signals [4, 9, 25, 26, 27, 28, 29, 30]. As such, the wavelet transform was used to extract features from EEG signals.

The wavelet transform, as a linear time-frequency transform, represents an efficient analytical tool in signal processing, pattern recognition and classification, and is suitable for analysis of transient and non-stationary phenomena as well as noise reduction. As a class of functions, it has the ability to localize information in both time and frequency [31]. Therefore, the wavelet transform has been utilized widely in biomedical signal processing [32, 33, 34]. In discrete wavelet analysis, a multi-resolution description is used to decompose a given signal $x(t)$ into increasingly finer detail based on two sets of basis functions [35], the wavelets and the scaling functions, as follows:

$$x(t) = \sum_k 2^{j_0/2} a_{j_0}(k) \varphi(2^{j_0} t - k) + \sum_{j=j_0}^{\infty} \sum_k 2^{j/2} d_j(k) \psi(2^j t - k) \quad (1)$$

where functions $\varphi(t)$ and $\psi(t)$ are the basic scaling and mother wavelet, respectively. In the above expansion, the first summation represents an approximation of $x(t)$ based on the scale index of j_0 , while the second term adds more detail using larger j (finer scales). The coefficients in this wavelet expansion are called the discrete wavelet transform (DWT) of the signal $x(t)$.

When the wavelets are orthogonal, these coefficients can be calculated by

$$a_j(k) = \int_{-\infty}^{\infty} 2^{j/2} x(t) \varphi(2^j t - k) dt \quad (2)$$

$$d_j(k) = \int_{-\infty}^{\infty} 2^{j/2} x(t) \psi(2^j t - k) dt \quad (3)$$

where $a_j(k)$ and $d_j(k)$ are the wavelet approximation and detail coefficients, respectively. In the DWT, the frequency axis is divided into dyadic intervals towards the lower frequencies, while the bandwidth length decreases exponentially. The wavelet packet (WP) transform is a

generalization of the DWT in which decomposition is undertaken in both directions (lower and higher frequencies). This general decomposition offers a greater range of possibilities for signal analysis than the discrete wavelet decomposition. In the WP tree, each node is recognized by the decomposition level (scale) l with respect to the WP tree root and the frequency band f . The ability of the wavelet transform in adaptive time-scale representation and decomposition of a signal into different frequency sub-bands presents an efficient signal analysis method without introducing a calculation burden [36]. Based on wavelet coefficients obtained after the wavelet transform, the signal can be reconstructed in each of the previously derived sub-bands and its time-domain features in different sub-bands can be studied separately.

2.2.2. Feature space reduction

After an appropriate signal analysis (e.g. wavelet transform used in this research), as well as feature extraction, the feature vector $Y = [y_1, y_2 \dots y_n]$ is derived. Its dimension should be reduced since the dimension n is often too large and the design of classifiers for a large dimension suffers from various difficulties. Those are mostly numerical problems that involve operation with high-order matrices. At the same time, a classifier in n -dimensional space is very difficult to analyze and almost impossible to imagine. Thus, it is helpful to define a matrix A whose dimension is $n \times m$ and in which the number of columns m is smaller than the number of rows n , such that the initial vector Y , following linear transformation $Z = A^T Y$, is projected onto the vector Z whose dimension m is significantly smaller (e.g. 2 or 3 when it is possible to visualize classifiers in two- or three-dimensional space). Obviously, such transformation results in a loss of some information contained in the original vector Y but the classification procedure is simplified. The selection of the matrix A is a trade-off between the desired level of simplicity of the classification procedure and the inevitable loss of information due to dimension reduction [37].

The reduction matrix A can be determined in several different ways using different approaches. Among them, the most often applied approach is the Karhunen-Loeve Expansion method [38], which, depending on the area of application, is also referred to as Principal Component Analysis (PCA). The main idea behind these methods is to determine the direction in which the scattering of the random vector is the greatest, through analysis of a covariance matrix. It is assumed that this is the most informative direction and that, in the case of dimension reduction, it should be preserved because it carries the largest amount of information. However, such an approach is not always convenient for some applications including the one discussed here. Figure 5 shows the realizations of a random two-dimensional vector. The Karhunen-Loeve Expansion method would, based on calculated eigenvectors and eigenvalues, determine the principal components z_1 and z_2 . Since the eigenvalue that corresponds to the component z_1 is higher than the corresponding value of the component z_2 , following dimension reduction the dimension of the component z_2 would be sacrificed and z_1 would be retained.

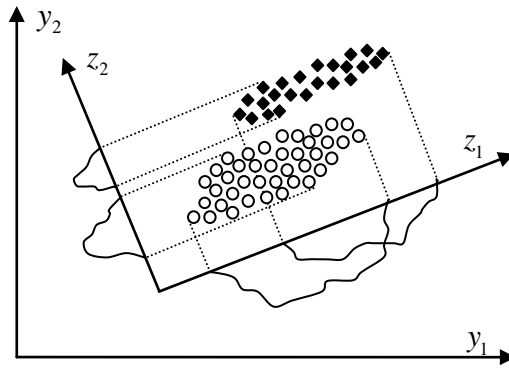


Figure 5 - Two different approaches to dimension reduction.

However, the samples shown in Fig. 5 form two clusters, represent measurement data obtained under different conditions and fall into two different categories. If the entire dimension reduction procedure is only one step of a larger process whose goal is the ultimate classification of measurements, then dimension reduction must address the separability of the categories created after reduction. Figure 5 clearly shows that it is more beneficial to preserve another

principal component, z_2 , even though its eigenvalue is considerably lower, because following the projection of the original vectors onto the axis z_2 there will be no overlapping of corresponding probability density functions. In this way it is possible to achieve good classification even in a space with reduced dimensions.

The procedure that both reduces dimensions and addresses the separability of classes is known as scatter matrices-based dimension reduction [38] and its essence can be described as follows:

It is assumed that N vectors of n -dimensional features $Y_i, i=1, \dots, N$, are available in the form $Y_i = [Y_{i1} Y_{i2} \dots Y_{in}]$. In the present case, n represents the number of extracted features after wavelet analysis of the EEG signal and N is the number of EEG data segments analyzed. It is further assumed that the elements of this data set can be divided into a certain number of classes, c . The number of classes in the present case is three (normal, interictal and ictal EEG data). In other words, the initial data set $\{X_i, i=1, \dots, N\}$ can be divided into c subsets of the form $\{X_i^{(k)}, i=1, \dots, N_k, k=1, \dots, c\}$. It is then assumed that the reduction matrix A of dimensions $n \times m$ is such that linear transformation $Z = A^T Y$ of each of the vectors from the initial data set is projected onto the corresponding m -dimensional vector Z . Consequently, the vector Z set can be classified into c subsets $\{Z_i^{(k)} = A^T Y_i^{(k)}, i=1, \dots, N_k, k=1, \dots, c\}$. Corresponding mathematical expectation vectors and covariance matrices can be added to each of these subsets, or classes:

$$M_k = E\{Z^{(k)}\}, \Sigma_k = E\{(Z^{(k)} - M_k)(Z^{(k)} - M_k)^T\} \quad (4)$$

where $E\{\cdot\}$ is the mathematical expectation operator. In practice, however, because the corresponding joint probability density functions are usually not known, these mathematical expectations are most often approximated by sample estimation:

$$M_k \approx \frac{1}{N_k} \sum_{j=1}^{N_k} Z_j^{(k)} ; \Sigma_k \approx \frac{1}{N_k} \sum_{j=1}^{N_k} (Z_j^{(k)} - M_k)(Z_j^{(k)} - M_k)^T \quad (5)$$

It is also possible to estimate *a priori* probabilities of occurrence of certain classes within the data set:

$$P_k \approx \frac{N_k}{N}, k = 1, \dots, c \quad (6)$$

Based on such estimates, within-class S_w and between-class S_b scatter matrices are generated as follows:

$$S_w = \sum_{k=1}^c P_k \Sigma_k ; S_b = \sum_{k=1}^c P_k (M_k - M_0)(M_k - M_0)^T ; M_0 = \sum_{k=1}^c P_k M_k \quad (7)$$

In statistical discriminant analysis, within-class and between-class scatter matrices are used to formulate criteria of class separability. A within-class scatter matrix shows the scatter of samples around their respective class expected vectors, while a between-class scatter matrix is the scatter of the expected vectors around the mixture mean. If the members of different classes are recognizable and if no mixing of members of different classes occurs, the elements of the within-class matrix should be as small as possible, while the elements of the between-class matrix should be as large as possible. There are different approaches that address these two requirements. The following criterion is adopted in this research:

$$J_1 = tr \{ S_2^{-1} S_1 \} \quad (8)$$

where $tr \{ \cdot \}$ is the trace of the quadratic matrix, while the matrix S_1 is usually S_b , and the matrix S_2 is either S_w or $S_w + S_b$ (as in this paper). This criterion is a function of an unknown reduction matrix A and the proper selection of such matrix A that can maximize it. The criterion is designed to minimize the matrix S_2 and maximize S_1 . The literature shows that its maximum value is obtained if the matrix A is generated as:

$$A = [\Psi_1 \Psi_2 \dots \Psi_m] \quad (9)$$

where Ψ_1, \dots, Ψ_m are the eigenvectors of the matrix $S_2^{-1}S_1$:

$$(S_2^{-1}S_1)\Psi_i = \lambda_i\Psi_i, i=1, \dots, m \quad (10)$$

which correspond to the highest eigenvalues of the same matrix: $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_m \geq \lambda_{m+1} \geq \lambda_n$,

where S_2 is the within-class scatter matrix and S_1 is the between-class scatter matrix of the original feature vector Y . This procedure was applied to obtain the results discussed in Section 3.

An interesting question regarding such a procedure is whether the information lost due to the reduction of the original n -dimensional vector Y into the space of the m -dimensional vector Z can be quantified in some way. The answer can be found in the eigenvalues $\lambda_i, i=1, \dots, n$ of the matrix $(S_2^{-1}S_1)$. Namely, each component of the eigenvectors Ψ_i carries the amount of information that corresponds to the measure of the appropriate eigenvalue λ_i . Since this procedure retains only those components whose eigenvalues are among the m highest values, while the remaining $n-m$ coordinates are cut off, the relative measure of the retained information can be computed as follows:

$$l(m) = \frac{\sum_{i=1}^m \lambda_i}{\sum_{i=1}^n \lambda_i} 100\% \quad (11)$$

In view of its significance, $l(m)$ may be referred to as an informativity index. Based on its definition, the value of the informativity index ranges from 0% for $m=0$ to 100% for $m=n$. This index shows the level of success of dimension reduction, or the extent of information saved following dimension reduction. Experience suggests that the results of dimension reduction with an informativity index higher than 85% is deemed satisfactory [37], as in the present case.

Even though the percentage of information lost during feature dimension reduction might be high (i.e. more than 10%), this does not necessarily mean that we lose that much information needed for successful classification (e.g. as shown in Fig. 5). Proper dimension reduction

improves the separability between classes and thus makes it easier for the classifiers to correctly classify the data at a later stage. In addition, without dimension reduction another very important feature of the proposed algorithm is lost, and that is the visualization of results and thus easier interpretation since any classifier in 25-dimensional space is very difficult to analyze and almost impossible to imagine.

2.2.3. Design of classifiers

As shown in Fig. 4, following dimension reduction, the obtained random vector Z is classified. The use of classifiers with a quadratic discriminatory function is proposed. Since the expected end result of this algorithm is the classification of EEG data into one of three categories (normal, interictal or ictal), the structure of the algorithm shown in Fig. 4 suggests sequential application of two classifiers. The first step is to decide whether the obtained vector Z belongs to an ictal (epileptic) EEG or an EEG which is not ictal. If the EEG is not ictal, the next step is to choose between a normal EEG and an interictal EEG. During the design of a quadratic classifier, the goal is to design the classifier (Z) in the form:

$$h(Z) = Z^T QZ + V^T Z + v_0 \quad (12)$$

where Q, V and v_0 are the unknown $m \times m$ matrix, $m \times 1$ vector and scalar, respectively, which uniquely define the classifier. Based on the sign of the classifier for the given vector Z , the algorithm decides whether this vector originates from the first class, ω_1 (e.g. interictal and ictal EEG data together), or the second class, ω_2 (e.g. normal EEG):

$$h(Z) = Z^T QZ + V^T Z + v_0 > 0 \Rightarrow Z \in \omega_2 \quad (13)$$

$$h(Z) = Z^T QZ + V^T Z + v_0 < 0 \Rightarrow Z \in \omega_1 \quad (14)$$

where ω_1 and ω_2 are the classes under consideration and $h(Z)$ is called the quadratic discrimination function. Q , V and v_0 are the matrix, vector and scalar, respectively. In the case of m -dimensional space, Eq. (13) and Eq. (14) can be interpreted as a linear equation

$$h(Z) = \sum_{i=1}^m \sum_{j=1}^m q_{ij} z_i z_j + \sum_{i=1}^m v_i z_i + v_0 = \sum_{i=1}^{\frac{m(m+1)}{2}} \alpha_i l_i + \sum_{i=1}^m v_i z_i + v_0 \quad (15)$$

where q_{ij} and v_i are the components of Q and V , respectively. Each of the new variables, l_i , represents a product of two z 's, and α is the corresponding q . Since Eq. (15) is a linear discriminant function, one can apply the optimum design procedure for a linear classifier [38], [39], resulting in

$$V_q = \left[\alpha_1 \cdots \alpha_{m(m+1)/2} \quad v_1 \cdots v_m \right]^T = \left[sK_1 + (1-s)K_2 \right]^{-1} (D_2 - D_1) \quad (16)$$

where D_i and K_i are the expected vector and covariance matrix of $W = \left[L^T Z^T \right]^T$ with $\left[m(m+1)/2 \right] + m$ variables. Since the l 's are products of the z 's, K_i includes the third and fourth order moments of Z . The expected values and variances of $h(W)$ are

$$\eta_i = E\{h(W)|\omega_i\} = V_q^T E\{W|\omega_i\} + v_0 = V_q^T D_i + v_0 \quad (17)$$

$$\sigma_i^2 = \text{Var}\{h(W)|\omega_i\} = V_q^T E\{(W - D_i)(W - D_i)^T|\omega_i\} V_q = V_q^T K_i V_q \quad (18)$$

Let $f(\eta_1, \eta_2, \sigma_1^2, \sigma_2^2)$ be any criterion to be minimized or maximized, to determine the optimum V_q and v_0 . The derivatives of f with respect to V_q and v_0 are then

$$\frac{\partial f}{\partial V_q} = \frac{\partial f}{\partial \sigma_1^2} \frac{\partial \sigma_1^2}{\partial V_q} + \frac{\partial f}{\partial \sigma_2^2} \frac{\partial \sigma_2^2}{\partial V_q} + \frac{\partial f}{\partial \eta_1} \frac{\partial \eta_1}{\partial V_q} + \frac{\partial f}{\partial \eta_2} \frac{\partial \eta_2}{\partial V_q} \quad (19)$$

$$\frac{\partial f}{\partial v_0} = \frac{\partial f}{\partial \sigma_1^2} \frac{\partial \sigma_1^2}{\partial v_0} + \frac{\partial f}{\partial \sigma_2^2} \frac{\partial \sigma_2^2}{\partial v_0} + \frac{\partial f}{\partial \eta_1} \frac{\partial \eta_1}{\partial v_0} + \frac{\partial f}{\partial \eta_2} \frac{\partial \eta_2}{\partial v_0} \quad (20)$$

On the other hand, it follows from Eq. (17) and Eq. (18) that

$$\frac{\partial \sigma_i^2}{\partial V_q} = 2K_i V_q, \quad \frac{\partial \eta_i}{\partial V_q} = D_i, \quad \frac{\partial \sigma_i^2}{\partial v_0} = 0, \quad \frac{\partial \eta_i}{\partial v_0} = 1 \quad (21)$$

Equating Eq. (19) and Eq. (20) to zero, after a short calculus the optimum V_q Eq. (16) can be computed, where the parameter s in the range $0 \leq s \leq 1$ needs to be chosen as

$$s = \frac{\partial f / \partial \sigma_1^2}{\partial f / \partial \sigma_1^2 + \partial f / \partial \sigma_2^2} \quad (22)$$

Once the functional form of f is selected, the optimum value of v_0 is derived as the solution of

$$\frac{\partial f}{\partial \eta_1} + \frac{\partial f}{\partial \eta_2} = 0 \quad (23)$$

If we assume a criterion in the form

$$f = \frac{P_1 \eta_1^2 + P_2 \eta_2^2}{P_1 \sigma_1^2 + P_2 \sigma_2^2} \quad (24)$$

which measures the between-class scatter (around zero) normalized by the within-class scatter, optimum V_q and v_0 are derived as

$$V_q = [P_1 K_1 + P_2 K_2]^{-1} (D_2 - D_1) \quad (25)$$

$$v_0 = -V_q^T [P_1 D_1 + P_2 D_2] = -(D_2 - D_1)^T [P_1 K_1 + P_2 K_2]^{-1} [P_1 D_1 + P_2 D_2] \quad (26)$$

It is allowed to add a constant real value to scalar v_0 in order to minimize the number of misclassified patterns.

3. Results

3.1. Wavelet Transform and Feature Extraction

The five EEG sub-bands of clinical interest: delta, theta, alpha, beta and gamma, span the 0-64 Hz frequency range. All higher frequencies are usually considered as noise. Due to the sampling

frequency of 173.61 Hz and in accordance with the Nyquist sampling theorem, the maximum useful frequency of the EEG data in this research is 86.81 Hz, or half the sampling frequency. In comparison to the Fourier transform, wavelet decomposition does not allow the extraction of specific frequency bands without additional filtering. Thus, in order to correlate the wavelet decomposition with the above mentioned five sub-bands, the frequency content of the EEG signal segments was restricted to the 0-64 Hz band by convolving the signal with a low-pass finite impulse response (FIR) filter.

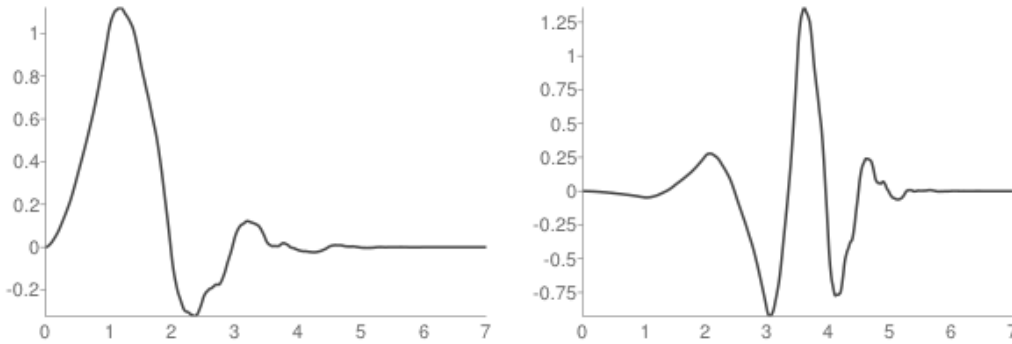


Figure 6 - Scaling function φ and wavelet function ψ of 4th order Daubechies wavelet.

The selection of an appropriate wavelet and the number of decomposition levels is also very important in any analysis of signals using the wavelet transform. After the performance of different wavelets was compared, it was decided to finally implement the 4th order Daubechies wavelet that is orthogonal and has scaling and wavelet functions as shown in Fig. 6. The number of decomposition levels was chosen based on the five sub-bands. The previously filtered EEG signal segments were decomposed into four levels. The wavelet coefficients were computed for all five different sub-bands of clinical interest based on Eqs. (2)-(3). Selection of the right features is among the most important components in the design of proper classifiers since even the best classifier will perform very poorly if the features are not selected well. Given that the computed wavelet coefficients are a good representation of the signal in both

time and frequency, almost all extracted features used later in dimension reduction and classification are based on them.

Apart from the mean and standard deviation of the wavelet coefficients in each sub-band, as typical statistical features, the entropy S (as a statistical measure of randomness) and the relative wavelet energy RWE in each sub-band were also computed based on previously derived wavelet coefficients, which are in the case of the delta sub-band calculated by Eqs. (27)-(28), respectively, as shown in Fig. 7 for all 300 EEG signal segments.

$$S_{\delta} = \sum_k \log(a_4(k)^2) \quad (27)$$

$$RWE_{\delta} = \frac{E_{\delta}}{E_{tot}}, \quad E_{\delta} = \sum_k |a_4(k)|^2, \quad E_{tot} = \sum_{j=1}^4 \sum_k |d_j(k)|^2 + \sum_k |a_4(k)|^2 \quad (28)$$

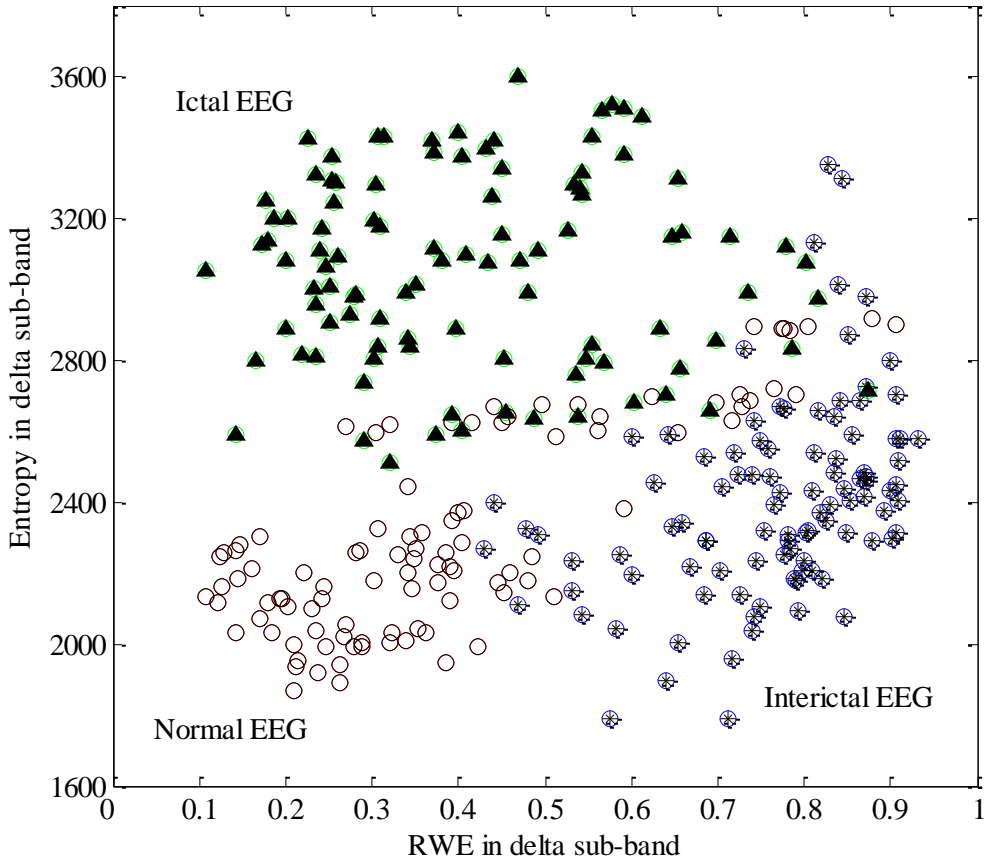


Figure 7 - Relative wavelet energy and entropy of wavelet coefficients in delta sub-band.

In order to make the entire classification algorithm more robust and at the same time increase its accuracy, five additional features in the time domain were computed. Applying the inverse wavelet transform, the EEG signal was reconstructed and its standard deviation in each sub-band computed. A representative example of an interictal EEG signal in all five sub-bands of interest, after inverse wavelet transform, is shown in Fig. 8.

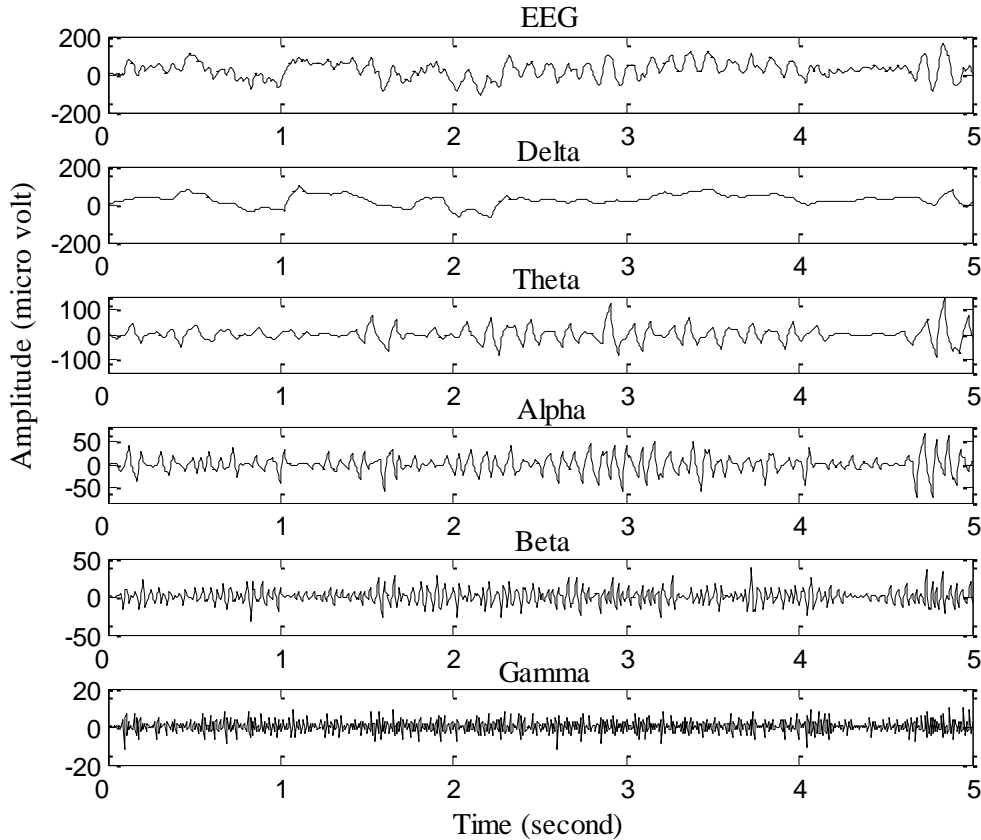


Figure 8 - Decomposition of band-limited EEG into five sub-bands (i.e. 4 levels) using 4th order Daubechies wavelet.

In total, 25 features were extracted from all 300 EEG signal segments using the MATLAB software package and thus a 25-dimensional feature space Y was created. Even though a few of these features carry good information about their class (e.g. the two shown in Fig. 7), none of them in combination with any other is still enough for a successful classification without further processing as described in the following section.

3.2. Feature Space Reduction and Classification

The next step in the proposed algorithm (Fig. 4) is to define the reduction matrix A and suitable classifiers using the set of 300 25-dimensional feature vectors $Y_i = [y_{i1}, y_{i2}, \dots, y_{i25}]^T$, $i = 1, 2, \dots, 300$ previously extracted from all three EEG data groups (i.e. 100 feature vectors from each subject group) that resulted in three subsets $Y_i^{(1)}, Y_i^{(2)}, Y_i^{(3)}$, $i = 1, \dots, 100$. The superscripts (1), (2) and (3) denote normal, interictal and ictal EEG data, respectively. In order to reduce dimensions and according to Eqs. (4)-(7), all the necessary statistics were computed and resulted in (25×1) -dimensional mean vectors M_1, M_2 and M_3 and (25×25) -dimensional within-class scatter matrix -dimensional within-class scatter matrix S_w and between-class scatter matrix S_b . In order to minimize the criterion in Eq. (8), eigenvalues and eigenvectors were computed for the matrix $(S_w + S_b)^{-1} S_b$. The eigenvectors Ψ_1 and Ψ_2 that corresponded to the two highest eigenvalues were selected and based on Eq. (9) the reduction matrix A was created. In that way the initial 25-dimensional vectors were projected onto two-dimensional space and from the initial three subsets three new subsets were derived: $Z_i^{(1)} = A^T Y_i^{(1)}$, $Z_i^{(2)} = A^T Y_i^{(2)}$, $Z_i^{(3)} = A^T Y_i^{(3)}$, $i = 1, \dots, 100$, which will later be used in the design of classifiers. The advantage of such dimension reduction, from 25 to 2, is two-fold. First, all of the obtained data can be represented by points in two dimensional space as shown in Fig. 9 and, as a result, a visual approach to data analysis is made possible. Second, with such dimension reduction, the elements of the matrix A are indicative of the importance of the features obtained from individual EEG signal segments with regard to the ultimate classification. Even though these parameters, and the elements of the reduced vector, have no clear physical meaning, their relative relationship is important. In other words, the coefficients of the matrix A , which is the outcome of a statistical analysis of the distribution of

scatter matrices, are indicative of the relevance of the various features to the classification process. In the present case, after analyzing the elements of the matrix A , the conclusion was that the extracted features related to theta, beta and alpha sub-bands have a larger contribution to the classification accuracy, and thus the entire algorithm as well, than the other sub-bands.

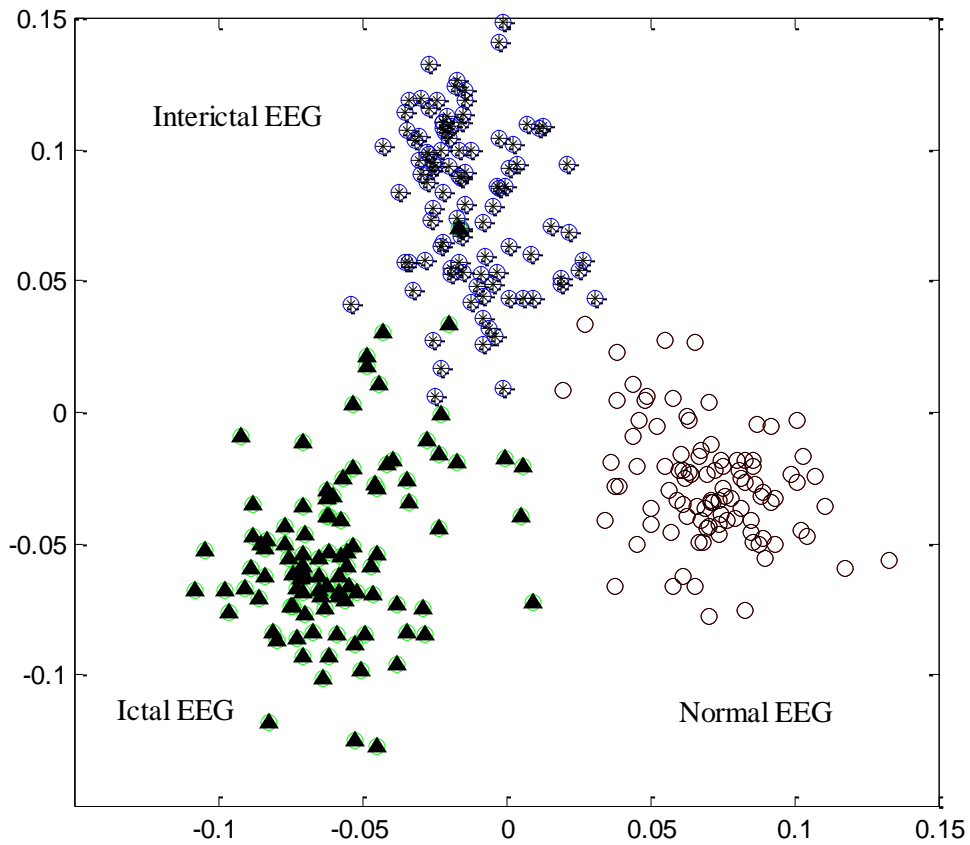


Figure 9 - All 300 EEG signal segments presented in two-dimensional space after feature space reduction.

The final step is the selection of matrices Q_1 and Q_2 , vectors V_1 and V_2 , and scalars v_{01} and v_{02} , which define the quadratic classifiers. The classifiers shown in Fig. 10 are designed using 150 training subsets Z from all three different EEG data groups, following the design procedure explained above.

3.3. Testing and Validation Results

After the design of the classifiers, the rest of the 150 EEG data segments (i.e. 50 EEG segments from each of three classes), were used to test and validate the accuracy of the proposed algorithm for classifying EEG signals. The aim was to assign the input patterns to one of the classes according to the characteristic features selected for that class. There were three classes: normal EEG class (healthy subjects), interictal EEG class (seizure-free intervals of five patients from the epileptogenic zone) and ictal EEG class (epileptic seizure segments). The classification results of the implemented algorithm were displayed by means of a confusion matrix, shown below in Table I, where each cell contains the raw number of exemplars classified for the corresponding combination of desired and actual classification results.

Table I - Confusion matrix

Output/Desired	Normal EEG	Interictal EEG	Ictal EEG
Normal EEG	50	0	0
Interictal EEG	0	49	1
Ictal EEG	0	1	49

According to the confusion matrix, all 50 normal EEG data segments were properly classified by the algorithm. One EEG data segment from the interictal EEG class was classified incorrectly, as a segment from the ictal EEG class. Additionally, one segment from the ictal EEG class was classified incorrectly, as a segment from the interictal EEG class.

The performance of the classifiers was determined by the calculation of sensitivity, specificity and total classification accuracy, which are defined as:

Sensitivity: total number of correctly classified positive patterns/total number of actual positive patterns; a positive pattern indicates an EEG data segment from one of these three classes.

Specificity: number of correctly classified negative patterns/total number of actual negative patterns; a negative pattern indicates a segment from one class classified as a member of one of the other two classes.

Total classification accuracy: total number of correctly classified patterns/total number of applied patterns; a pattern indicates an EEG data segment from all three classes.

The values of these statistical parameters are shown in Table II. The classification algorithm classified normal, interictal and ictal EEG data sets with an accuracy of 100%, 98% and 98%, respectively. All the sets were classified with an accuracy of 99%, which is the total classification accuracy. Taking into account other reported results that span from 85% to 99% (summarized in Table III), it is safe to say that the total classification accuracy of the algorithm proposed for this application is quite high and thus has potential for a real clinical setting.

Table II - Statistical parameters

EEG data sets	Statistical parameters		
	Sensitivity (%)	Specificity (%)	Accuracy (%)
Normal EEG	100	100	99
Interictal EEG	98	99.99	
Ictal EEG	98	99.99	

Table III - Performance of other methods applied on the same EEG data sets

Author(s)	Year	Method	Accuracy (%)
Guler <i>et al.</i> [10]	2005	Lyapunov exponents, recurrent neural network	97
Ubeyli <i>et al.</i> [11]	2006	Lyapunov exponents, artificial neural network	95
Sadati <i>et al.</i> [28]	2006	Wavelet transform, adaptive neuro-fuzzy network	86
Adeli <i>et al.</i> [9]	2007	Wavelet transform, chaos analysis	-
Dastidar <i>et al.</i> [19]	2007	Wavelet transform, chaos analysis, K-means clusters	97
Tzallas <i>et al.</i> [12]	2007	Time-frequency analysis, artificial neural network	99
Chua <i>et al.</i> [40]	2008	Higher order spectra, Gaussian mixture model	93
Dastidar <i>et al.</i> [13]	2008	Principal component analysis, artificial neural network	99
Ubeyli <i>et al.</i> [29]	2008	Wavelet transform, mixture of expert model	93
Guo <i>et al.</i> [41]	2011	Genetic programming, K-nearest neighbor classifier	93
Orhan <i>et al.</i> [30]	2011	Wavelet transform, K-nearest neighbor classifier	97

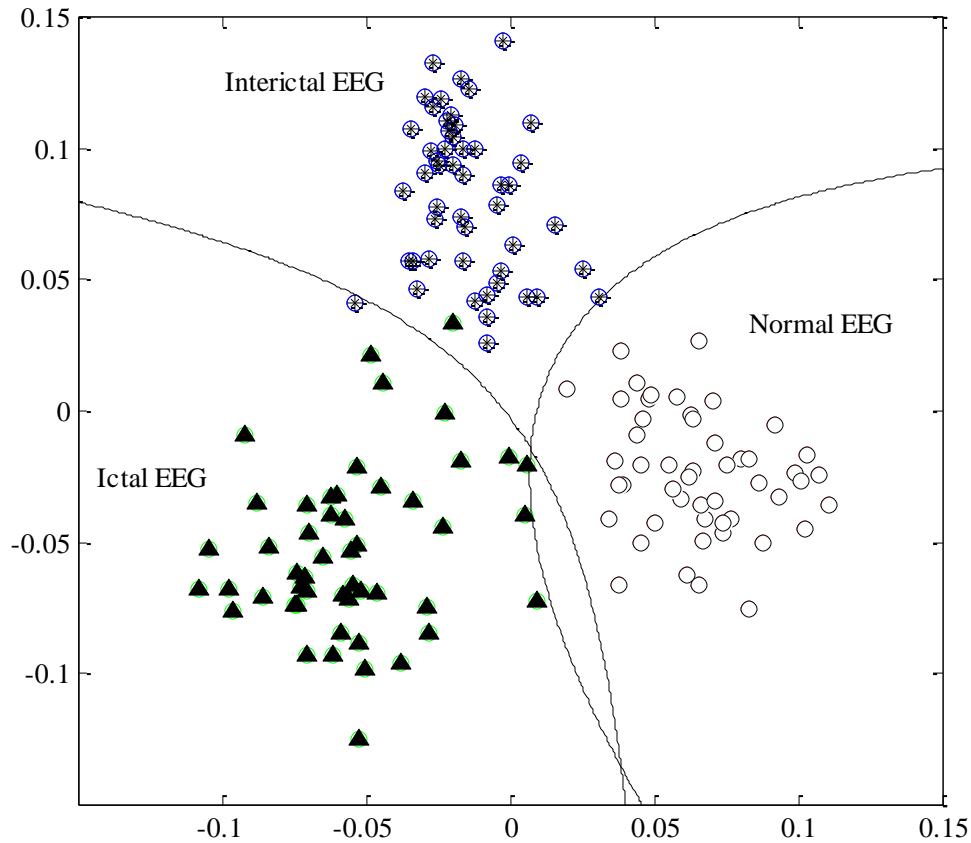


Figure 10 - Reduced two-dimensional feature space with two quadratic classifiers which separate the space into three separable sub-spaces that correspond to appropriate EEG data set.

The classification results are shown in Fig. 10, where it is possible to see the positions of the classifiers that separate the two-dimensional feature space into three classes, as well as all 150 testing data segments mapped into the reduced feature space.

4. Discussion

As shown in Fig. 10, the proposed approach is not important only because it provides an automatic, objective procedure that addresses all available features in a specific way and makes

a decision based on these data, but because it also allows insight into the severity of the brain state. Figure 10 shows that in a two-dimensional representation some of the points are located very close to the classification line that differentiates between interictal and ictal data, and that a few points are even in the wrong class, while a number of points are far from this discrimination function. The distance of these points is indicative of the severity of the brain state and the most distant point in the diagram (the first coordinate greater than 0.1) represents the truly healthiest brain state.

5. Conclusions

This paper presented an EEG data classification algorithm which, based on a large number of features extracted after wavelet transform and statistical pattern recognition, makes an objective decision about the type of the EEG data processed and thus the brain state of a patient. The main advantages of the algorithm are: (a) the ability of the algorithm to run robustly in a clinical setting with noised EEG; (b) feature extractions with highly meaningful wavelet transform because hidden EEG information can be revealed and the noise effort reduced as certain data under some scales are omitted; (c) simplicity and low computational cost guaranteeing real clinical application; (d) very good sensitivity and specificity as well as an overall classification accuracy of 99%; and (e) patient-independent algorithm that does not require any specific prior knowledge of each subject. Therefore, the conclusion is that the proposed algorithm can be used to classify EEG signals and detect seizures in a clinical setting.

Apart from the quite high overall classification accuracy achieved, there are still two directions for further improvement of the proposed algorithm. One is to include additional features in the feature vector, e.g. by nonlinear series analysis (i.e. chaos analysis) of EEG data. The other pertains to the selection of more sophisticated pattern recognition methods that will definitely result in a more complex but also a more accurate classification algorithm. In addition, the

algorithm's on-line seizure detection capabilities should be examined by analyzing long-term continuous EEG recordings, as well as its ability to detect other changes in EEG (e.g. those caused by cognitive tasks).

REFERENCES

1. Iasemidis LD, Epileptic seizure prediction and control, *IEEE Trans Biomed Eng* **50**:549, 2003.
2. Buck D, Baker GA, Jacoby A, Smith DF, Chadwick DW, Patient's experiences of injury as a result of epilepsy, *Epilepsia* **38**:439, 1997.
3. Elger CE, Future trends in epileptology, *Curr Opin Neurol* **14**:185, 2001.
4. Guler I, Ubeyli ED, Adaptive neuro-fuzzy inference system for classification of EEG signals using wavelet coefficients, *J Neurosci Methods* **148**:113, 2005.
5. Webber WR, Litt B, Lesser RP, Fisher RS, Bankman I, Automatic EEG spike detection: what should the computer imitate, *Electroen Clin Neuro* **87(6)**:364, 1993.
6. Gotman J, Automatic recognition of epileptic seizures in the EEG, *Electroen Clin Neuro* **54**:530, 1982.
7. Qu H, Gotman J, A patient-specific algorithm for the detection of seizure onset in long-term EEG monitoring: Possible use as a warning device, *IEEE Trans Biomed Eng* **44(2)**:115, 1997.
8. Gigola S, Ortiz F, Attellis CE, Silvaand W, Kochen S, Prediction of epileptic seizures using accumulated energy in a multiresolution framework, *J Neurosci Methods* **38**:107, 2004.
9. Adeli H, Ghosh-Dastidar S, Dadmehr N, A wavelet-chaos methodology for analysis of EEGs and EEG sub-bands to detect seizure and epilepsy, *IEEE Trans Biomed Eng* **54(2)**:205, 2004.
10. Guler I, Ubeyli ED, Guler I, Recurrent neural networks employing Lyapunov exponents in EEG recordings, *Expert Syst Appl*, **29(3)**, 2005.
11. Ubeyli ED, Analysis of EEG signals using Lyapunov exponents, *Neural Netw World*, **16(3)**:257, 2006.

12. Tzallas AT, Tsipouras MG, Fotiadis DI, Automatic seizure detection based on time-frequency analysis and artificial neural networks, *Comput Intell Neurosci*, 80510, 2007.
13. Ghosh-Dastidar S, Adeli H, Dadmehr N, Principal component analysis enhanced cosine radial basis function neural network for robust epilepsy and seizure detection, *IEEE Trans Biomed Eng*, **55(2)**:512, 2008.
14. Weng W, Khorasani K, An adaptive structure neural network with application to EEG automatic seizure detection, *Neural Networks* **9**:1223, 1996.
15. Gotman J, Wang L, State-dependent spike detection: Concepts and preliminary results, *Electroen, Clin Neuro* **79**:11, 1991.
16. Pradhan N, Sadasivan PK, Arunodaya GR, Detection of seizure activity in EEG by an artificial neural network: A preliminary study, *Comput Biomed Res* **29**:303, 1996.
17. Nigam VP, Graupe D, A neural-network-based detection of epilepsy, *Neurol Res* **26**:55, 2004.
18. Kiyimik MK, Subasi A, Ozcalik HR, Neural networks with periodogram and autoregressive spectral analysis methods in detection of epileptic seizure, *J Med Syst* **28**:511, 2004.
19. Ghosh-Dastidar S, Adeli H, Dadmehr N, Mixed wavelet-chaos-neural network methodology for epilepsy and epileptic seizure detection, *IEEE Trans Biomed Eng* **54(9)**:1545, 2006.
20. Srinivasan V, Eswaranand C, Sriraam N, Approximate entropy-based epileptic EEG detection using artificial neural networks, *IEEE Trans Inf Technol Biomed* **11**:288, 2007.
21. Subasi A, EEG signal classification using wavelet feature extraction and a mixture of expert model, *Expert Syst Appl* **32**:1084, 2007.
22. Iasemidis LD, Epileptic seizure prediction and control, *IEEE Trans Biomed Eng* **50**:549, 2003.
23. Andrzejak RG, Lehnertz K, Mormann F, Rieke C, David P, Elger CE, 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)**:1, 2001.

24. McGrogan N, *Neural Network Detection of Epileptic Seizures in the Electroencephalogram*, Oxford University, Oxford, 2001.
25. Hazarika N, Chen JZ, Tsoi AC, Sergejew A, Classification of EEG signals using the wavelet transform, *Signal Process* **59(1)**:61, 1997.
26. Rosso OA, Figliola A, Cresoand J, Serrano E, Analysis of wavelet-filtered tonic-clonic electroencephalogram recordings, *Med Biol Eng Comput* **42(40)**:516, 2004.
27. Glover JR, Raghaven N, Ktonas PY, Frost JD, Context-based automated detection of epileptogenic sharp transients in the EEG: elimination of false positives, *IEEE Trans Biomed Eng* **36(5)**:519, 1989.
28. Sadati N, Mohseni HR, Magshoudi A, Epileptic seizure detection using neural fuzzy networks, in *Proc. IEEE Intern Conf on Fuzzy Syst*, Canada, 2006.
29. Ubeyli ED, Wavelet/mixture of experts network structure for EEG classification, *Expert Syst Appl*, **37**:1954, 2008.
30. Orhan U, Hekim M, Ozer M, EEG signals classification using the K-means clustering and a multilayer perceptron neural network model, *Expert Syst Appl*, **38**:13475, 2011.
31. Tafreshi R, Dumont G, Gross D, Ries CR, Puil E, MacLeod BA, Seizure Detection by a novel wavelet packet method, in *Proc. 28th IEEE EMBS Annu Intern Conf* 6141-6144, 2006.
32. Khan YU, Gotman J, Wavelet based automatic seizure detection in intracerebral electroencephalogram, *Clin Neurophysiol* **114**:898, 2003.
33. Saab ME, Gotman J, A system to detect the onset of epileptic seizures in scalp EEG, *Clin Neurophysiol* **116**:427, 2005.
34. Zandi AS, Moradi MH, Quantitative evaluation of a wavelet based method in ventricular late potential detection, *Pattern Recogn* **39**:1369, 2006.
35. Burrus CS, Gopinath RA, Guo H, *Introduction to wavelets and wavelet transforms: A primer*, Prentice-Hall, Upper Saddle River, NJ, 1998.
36. Sweldens W, The lifting scheme: a construction of second generation wavelets, *SIAM J Math Anal* **29**:511, 1998.

37. Stricevic R, Djurovic N, Djurovic Z, Drought classification in northern Serbia based on SPI and statistical pattern recognition, *Meteorol Appl* **18**:60, 2010.
38. Fukunaga K, *Introduction to Statistical Pattern Recognition*, Chap. 4, Academic Press, Boston, 1990.
39. Theodoridis S, Koutroumbas K, *Pattern Recognition*, Academic Press, San Diego, 2003.
40. Chua KC, Chandran V, Acharya R, Lim CM, Automatic identification of epilepsy by HOS and power spectrum parameters using EEG signals: a comparative study, *in Proc. IEEE Eng Med Biol Soc*, 3824-3827, 2008.
41. Guo L, Rivero D, Dorado J, Munteanu CR, Pazos A, Automatic feature extraction using genetic programming: An application to epileptic EEG classification, *Expert Syst Appl*, **38**:1042, 2011.

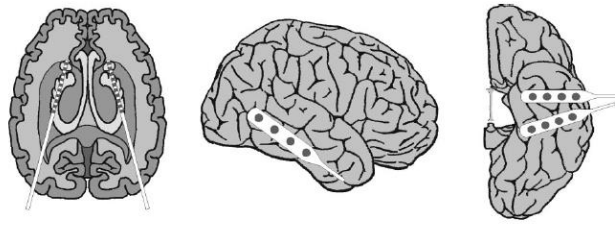


Figure 1 - Implanted intracranial electrodes.

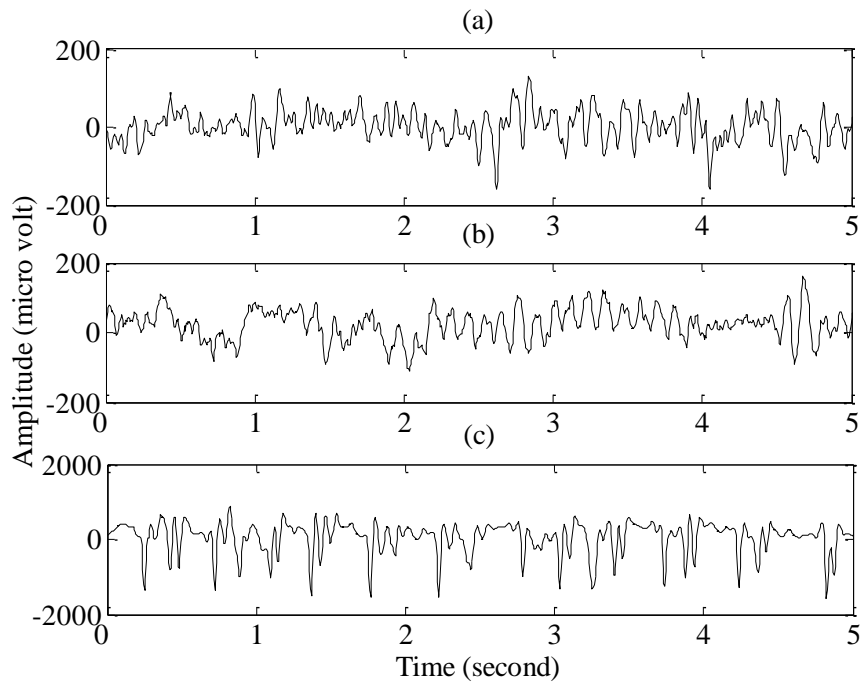


Figure 2 - Segments of EEG data: (a) Normal, (b) Interictal, (c) Ictal.

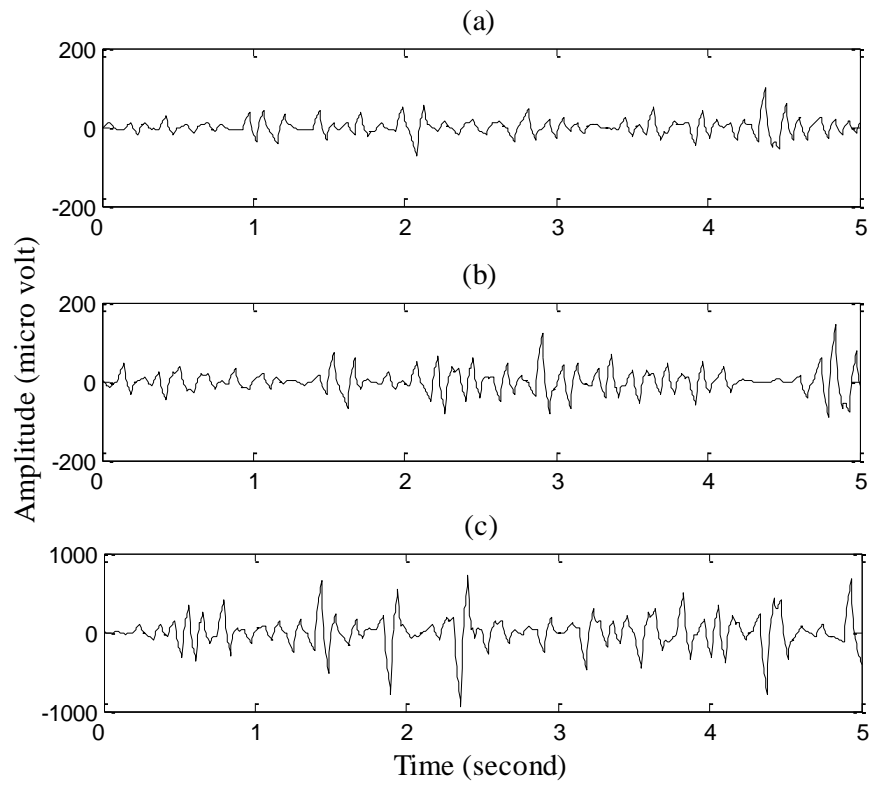


Figure 3 - Theta sub-bands of EEG data: (a) Normal, (b) Interictal, (c) Ictal.

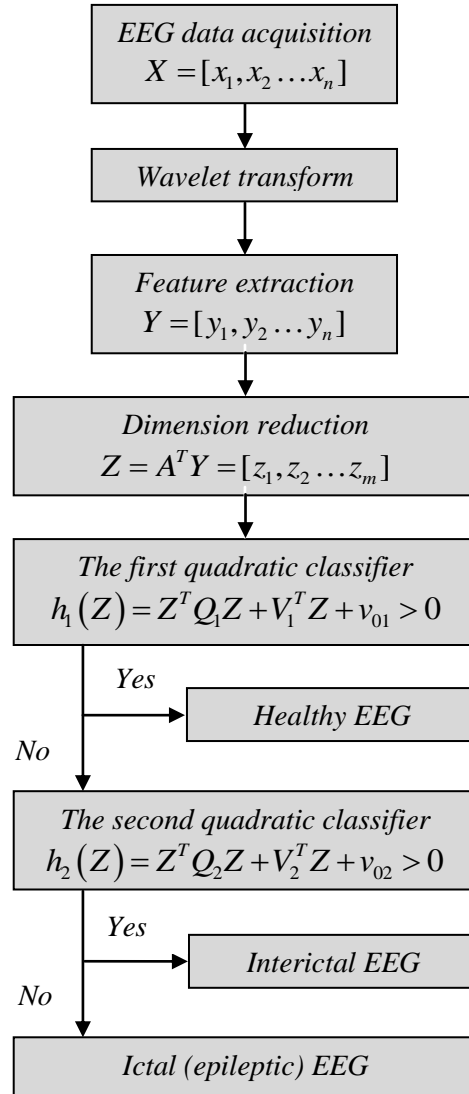


Figure 4 - Proposed classification algorithm.

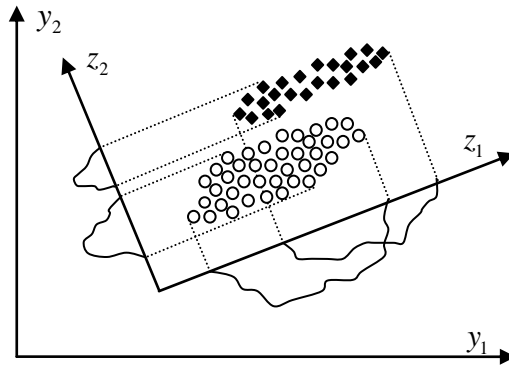


Figure 5 - Two different approaches to dimension reduction.

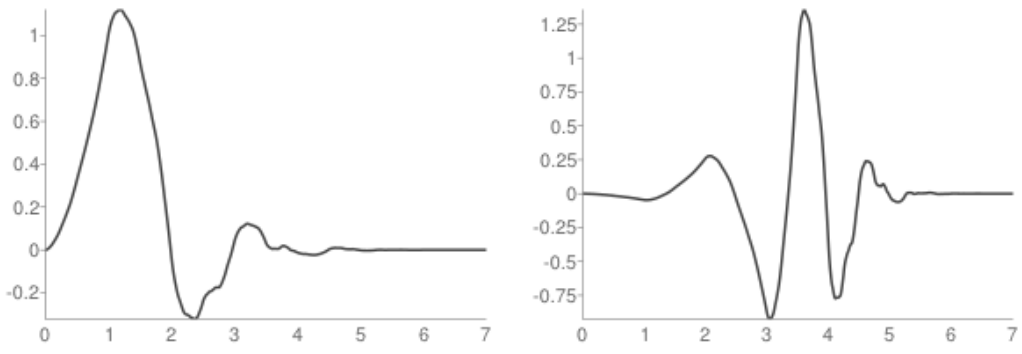


Figure 6 - Scaling function φ and wavelet function ψ of 4th order Daubechies wavelet.

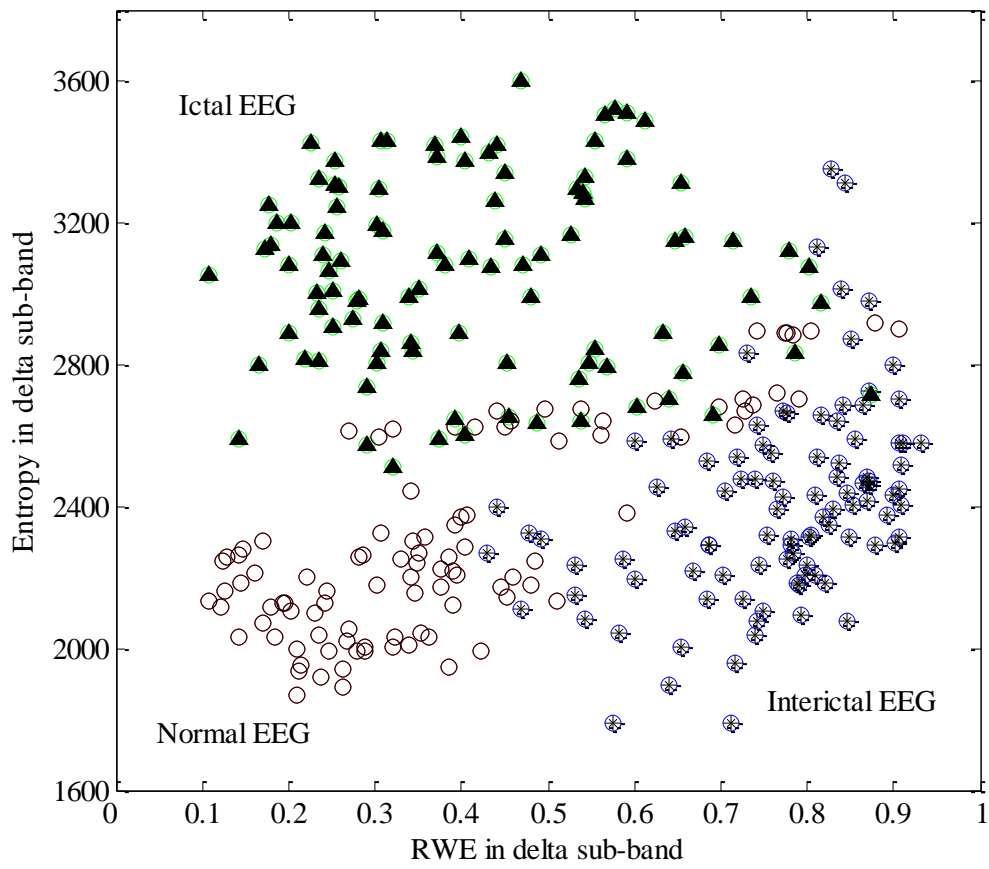


Figure 7 - Relative wavelet energy and entropy of wavelet coefficients in delta sub-band.

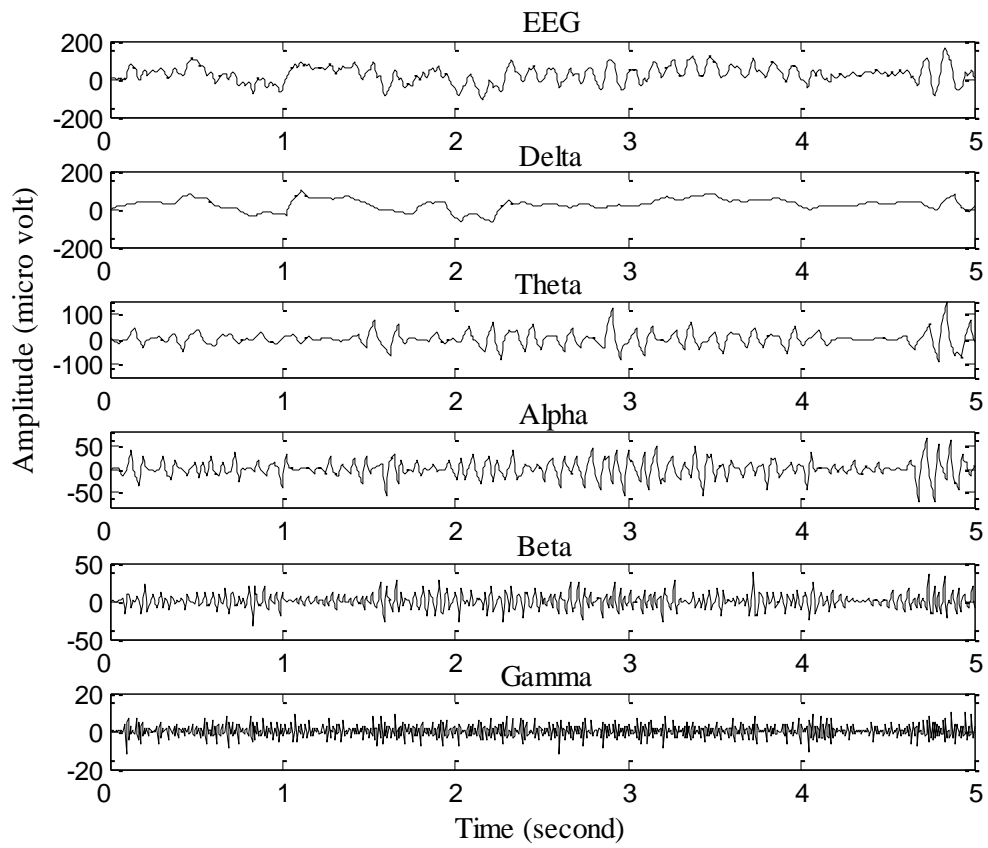


Figure 8 - Decomposition of band-limited EEG into five sub-bands (i.e. 4 levels) using 4th order Daubechies wavelet.

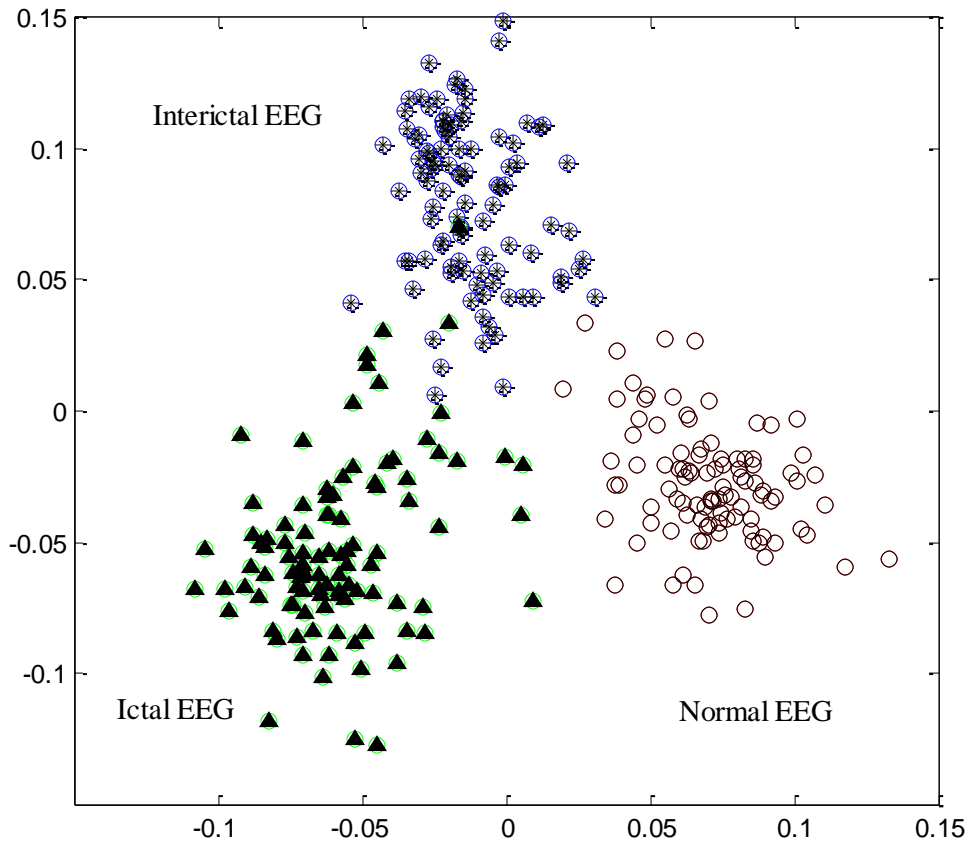


Figure 9 - All 300 EEG signal segments presented in two-dimensional space after feature space reduction.

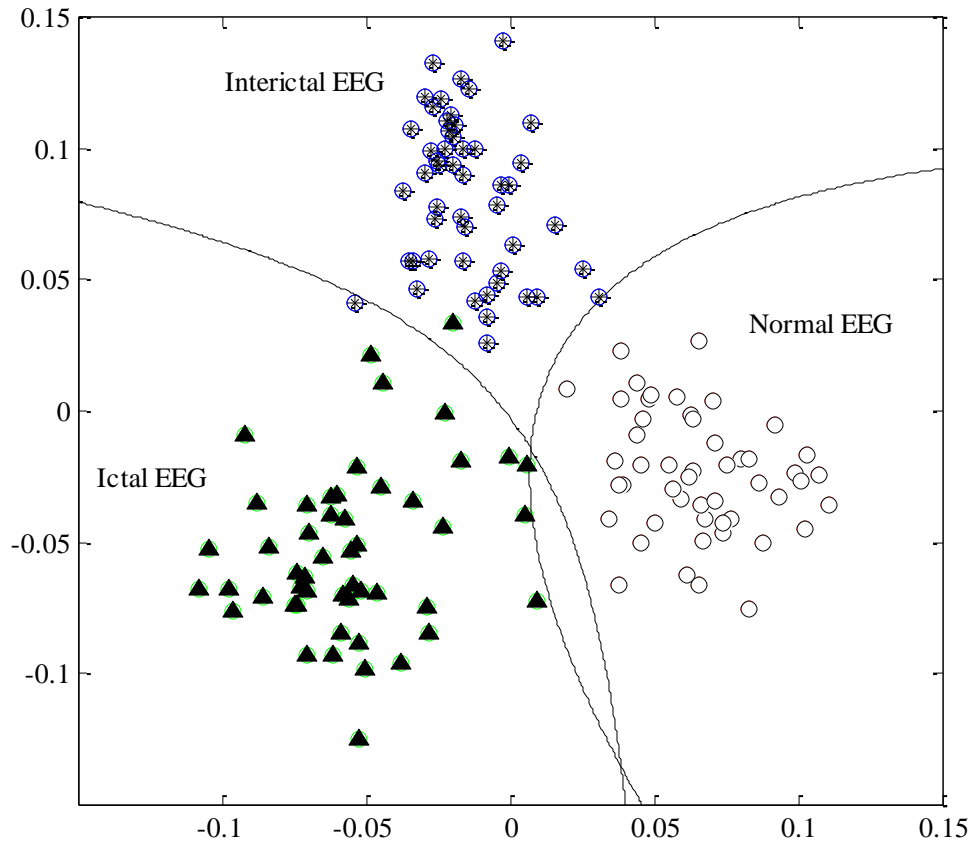


Figure 10 - Reduced two-dimensional feature space with two quadratic classifiers which separate the space into three separable sub-spaces that correspond to appropriate EEG data set.

Table I - Confusion matrix

Output/Desired	Normal EEG	Interictal EEG	Ictal EEG
Normal EEG	50	0	0
Interictal EEG	0	49	1
Ictal EEG	0	1	49

Table II - Statistical parameters

EEG data sets	Statistical parameters		
	Sensitivity (%)	Specificity (%)	Accuracy (%)
Normal EEG	100	100	99
Interictal EEG	98	99.99	
Ictal EEG	98	99.99	

Table III - Performance of other methods applied on the same EEG data sets

Author(s)	Year	Method	Accuracy (%)
Guler <i>et al.</i> [10]	2005	Lyapunov exponents, recurrent neural network	97
Ubeyli <i>et al.</i> [11]	2006	Lyapunov exponents, artificial neural network	95
Sadati <i>et al.</i> [28]	2006	Wavelet transform, adaptive neuro-fuzzy network	86
Adeli <i>et al.</i> [9]	2007	Wavelet transform, chaos analysis	-
Dastidar <i>et al.</i> [19]	2007	Wavelet transform, chaos analysis, K-means clusters	97
Tzallas <i>et al.</i> [12]	2007	Time-frequency analysis, artificial neural network	99
Chua <i>et al.</i> [40]	2008	Higher order spectra, Gaussian mixture model	93
Dastidar <i>et al.</i> [13]	2008	Principal component analysis, artificial neural network	99
Ubeyli <i>et al.</i> [29]	2008	Wavelet transform, mixture of expert model	93
Guo <i>et al.</i> [41]	2011	Genetic programming, K-nearest neighbour classifier	93
Orhan <i>et al.</i> [30]	2011	Wavelet transform, K-nearest neighbour classifier	97