View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by IntechOpen

We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

4,800
Open access books available

122,000
International authors and editors

135M
Downloads

Our authors are among the

154
Countries delivered to

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

CLARIVATE ANALYTICS BOOK CITATION INDEX INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

2

Intelligent Decision Support System for Classification of Eeg Signals using Wavelet Coefficients

Pari Jahankhani and Vassilis Kodogiannis
School of Computer Science, University of Westminster, London HA1 3TP, UK

1. Introduction

The human brain is obviously a complex system, and exhibits rich spatiotemporal dynamics. Among the non-invasive techniques for probing human brain dynamics, electroencephalography (EEG) provides a direct measure of cortical activity with millisecond temporal resolution. Early on, EEG analysis was restricted to visual inspection of EEG records. Since there is no definite criterion evaluated by the experts, visual analysis of EEG signals is insufficient. For example, in the case of dominant alpha activity delta and theta, activities are not noticed. Routine clinical diagnosis needs to analysis of EEG signals. Therefore, some automation and computer techniques have been used for this aim (Guler et al., 2001). Since the early days of automatic EEG processing, representations based on a Fourier transform have been most commonly applied. This approach is based on earlier observations that the EEG spectrum contains some characteristic waveforms that fall primarily within four frequency bands—delta (1–4 Hz), theta (4–8 Hz), alpha (8–13 Hz), and beta (13–30 Hz). Such methods have proved beneficial for various EEG characterizations, but fast Fourier transform (FFT), suffer from large noise sensitivity. Parametric power spectrum estimation methods such as AR, reduces the spectral loss problems and gives better frequency resolution. AR method has also an advantage over FFT that, it needs shorter duration data records than FFT (Zoubir et al., 1998). A powerful method was proposed in the late 1980s to perform time-scale analysis of signals: the wavelet transforms (WT). This method provides a unified framework for different techniques that have been developed for various applications. Since the WT is appropriate for analysis of non-stationary signals and this represents a major advantage over spectral analysis, it is well suited to locating transient events, which may occur during epileptic seizures. Wavelet's feature extraction and representation properties can be used to analyse various transient events in biological signals. (Adeli et al., 2003) gave an overview of the discrete wavelet transform (DWT) developed for recognising and quantifying spikes, sharp waves and spike-waves. Wavelet transform has been used to analyze and characterise epileptiform discharges in the form of 3-Hz spike and wave complex in patients with absence seizure. Through wavelet decomposition of the EEG records, transient features are accurately captured and localised in both time and frequency context. The capability of this mathematical microscope to analyse different scales of neural rhythms is shown to be a

Open Access Database www.intechweb.org

Source: Data Mining in Medical and Biological Research, Book edited by: Eugenia G. Giannopoulou,
ISBN 978-953-7619-30-5, pp. 320, December 2008, I-Tech, Vienna, Austria

powerful tool for investigating small-scale oscillations of the brain signals. A better understanding of the dynamics of the human brain through EEG analysis can be obtained through further analysis of such EEG records.

Numerous other techniques from the theory of signal analysis have been used to obtain representations and extract the features of interest for classification purposes. Neural networks and statistical pattern recognition methods have been applied to EEG analysis. Neural network (NN) detection systems have been proposed by a number of researchers. (Pradhan et al. 1996) used the raw EEG as an input to a neural network while (Weng and Khorasani ,1996) used the features proposed by Gotman with an adaptive structure neural network, but his results show a poor false detection rate. (Petrosian et al., 2000) showed that the ability of specifically designed and trained recurrent neural networks (RNN) combined with wavelet pre-processing, to predict the onset of epileptic seizures both on scalp and intracranial recordings only one-channel of electroencephalogram. In order to provide faster and efficient algorithm, (Folkers et al., 2003) proposed a versatile signal processing and analysis framework for bioelectrical data and in particular for neural recordings and 128-channel EEG. Within this framework the signal is decomposed into sub-bands using fast wavelet transform algorithms, executed in real-time on a current digital signal processor hardware platform. Neuro-fuzzy systems harness the power of the two paradigms: fuzzy logic and NNs by utilising the mathematical properties of NNs in tuning rule-based fuzzy systems that approximate the way human process information. A specific approach in neuro-fuzzy development is the adaptive neuro-fuzzy inference system (ANFIS), which has shown significant results in modelling nonlinear functions. In ANFIS, the membership function parameters are extracted from a data set that describes the system behaviour. The ANFIS learns features in the data set and adjusts the system parameters according to a given error criterion. Successful implementations of ANFIS in EEG analysis have been reported (Guler et al., 2004).

As compared to the conventional method of frequency analysis using Fourier transform or short time Fourier transform, wavelets enable analysis with a coarse to fine multi-resolution perspective of the signal. In this work, DWT has been applied for the time–frequency analysis of EEG signals and NNs for the classification using wavelet coefficients. EEG signals were decomposed into frequency sub-bands using discrete wavelet transform (DWT). A neural network system was implemented to classify the EEG signal to one of the categories: epileptic or normal. The aim of this study was to develop a simple algorithm for the detection of epileptic seizure, which could also be applied to real-time.

In this study, an alternative approach based on the multiple-classifier concept will be presented for epileptic seizure detection. A neural network classifier, Learning Vector Quantisation (LVQ2.1), is employed to classify unknown EEGs belonging to one set of signal.

Here we investigated the potential of statistical techniques, such as Rough Set and Principal Component Analysis (PCA) that capture the second-order statistical structure of the data. Fig. 1 shows overall computation scheme.

2. Data selection and recording

We have used the publicly available data described in (Andrzejak *et al.*). The complete data set consists of two sets (denoted A and E) each containing 100 single-channel EEG segments. These segments were selected and cut out from continuous multi-channel EEG recordings

after visual inspection for artefacts, e.g., due to muscle activity or eye movements. Sets A consisted of segments taken from surface EEG recordings that were carried out on five healthy volunteers using a standardised electrode placement scheme (Fig. 2).

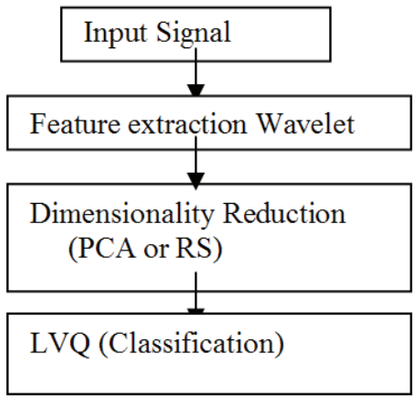


Fig. 1. Computation scheme for classifying signals

Brain Area	Left Hemisphere	Midline	Right Hemisphere
Pre-Frontal	Fp1		Fp2
Frontal	F3		F4
Inferior Frontal	F7		F8
Mid-Frontal		Fz	
Mid-Temporal	T3		T4
Posterior Temporal	T5		T6
Central	C3		C4
Vertex (Mid-Central)		Cz	
Parietal	P3		P4
Mid-Parietal		Pz	
Occipital	O1		O2
Ear (Auricular)	A1		A2
Ground		G	

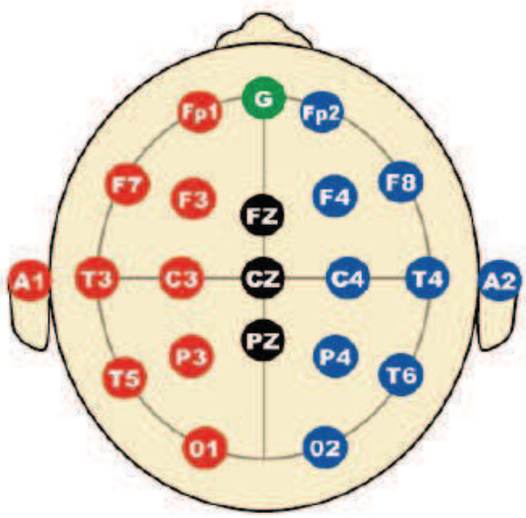


Fig. 2. The 10–20 international system of electrode placement c images of normal and abnormal cases.

Volunteers were relaxed in an awake-state with eyes open (A). Sets E originated from EEG archive of pre-surgical diagnosis. EEGs from five patients were selected, all of who had achieved complete seizure control after resection of one of the hippocampal formations, which was therefore correctly diagnosed to be the epileptogenic zone. Segments, set E only contained seizure activity.

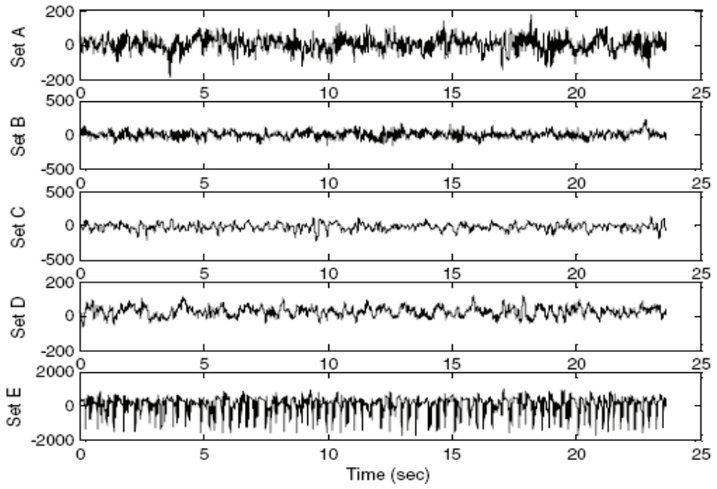


Fig. 3. Examples of five different sets of EEG signals taken from different subjects.

Here segments were selected from all recording sites exhibiting ictal activity. All EEG signals were recorded with the same 128-channel amplifier system, using an average common reference. The data were digitised at 173.61 samples per second using 12-bit resolution. Band-pass filter settings were 0.53–40 Hz (12dB/oct). In this study, we used two dataset (A and E) of the complete dataset (Jahankhani et al 2005) . Typical EEGs are depicted in Fig. 3.

3. Analysis using Discrete Wavelet Transform (DWT)

Wavelet transform is a spectral estimation technique in which any general function can be expressed as an infinite series of wavelets. The basic idea underlying wavelet analysis consists of expressing a signal as a linear combination of a particular set of functions (wavelet transform, WT), obtained by shifting and dilating one single function called a mother wavelet. The decomposition of the signal leads to a set of coefficients called wavelet coefficients. Therefore the signal can be reconstructed as a linear combination of the wavelet functions weighted by the wavelet coefficients.

DWT is a time-frequency analysis technique that is most suited for non-stationary signal. It was chosen as the feature extraction method for these signals, since the important feature may be in the time domain, in the frequency domain, or in both domains. DWT also has other properties, such as providing a substantial amount of data reduction. DWT analyses the signal at different frequency bands with different resolution by decomposing the signal into coarse approximation and detail information. DWT using two sets of functions, called scaling functions and wavelet functions, which are associated with low-pass and high-pass filter. The original signal x[n] is first passed through a half-band high-pass filter g[n] and a low-pass filter h[n].after the filtering, half of the samples can be eliminated according to

Nyquist's rule, since the signal now has a highest frequency of $\pi/2$ radians instead of π. The signal can therefore be sub sampled by 2, simply by discarding every other sample. In general when dealing with stationary signals, whose statistical properties are invariant over time, the ideal tool is the Fourier transform. The Fourier transform is an infinite linear combination of dilated cosine and sine waves. When we encounter non-stationary signals, we ca represent these signals by linear combinations of atomic decompositions known as wavelet.

In order to obtain an exact reconstruction of the signal, adequate number of coefficients must be computed. The key feature of wavelets is the time-frequency localisation. It means that most of the energy of the wavelet is restricted to a finite time interval. Frequency localisation means that the Fourier transform is band limited. When compared to STFT, the advantage of time-frequency localisation is that wavelet analysis varies the time-frequency aspect ratio, producing good frequency localization at low frequencies (long time windows), and good time localisation at high frequencies (short time windows). This produces a segmentation, or tiling of the time-frequency plane that is appropriate for most physical signals, especially those of a transient nature. The wavelet technique applied to the EEG signal will reveal features related to the transient nature of the signal, which are not obvious by the Fourier, transform. In general, it must be said that no time-frequency regions but rather time-scale regions are defined [Subasi, 2005). The operators h and g are called perfect reconstruction or quadrature mirror filters (QMFs) if they satisfy the orthogonality conditions:

$$G(z)G(z^{-1}) + G(-z)G(-z^{-1}) = 1 \tag{1}$$

Where G(z) denotes the z-transform of the filter g. Its complementary high-pass filter can be defined as

$$H(z) = zG(-z^{-1}) \tag{2}$$

A sequence of filters with increasing length (indexed by i) can be obtained

$$\begin{aligned} G_{i+1}(z) &= G(z^{2^i})G_i(z), \\ H_{i+1}(z) &= H(z^{2^i})G_i(z) \end{aligned} \tag{3}$$

With the initial condition $G_0(z) = 1$. It is expressed as a two-scale relation in time domain

$$\begin{aligned} g_{i+1}(k) &= [g]_{\uparrow 2^i}\, g_i(k), \\ h_{i+1}(k) &= [h]_{\uparrow 2^i}\, g_i(k) \end{aligned} \tag{4}$$

where the subscript $[.]_{\uparrow m}$ indicates the up sampling by a factor of m and k is the equally sampled discrete time.

The normalised wavelet and scale basis functions $\varphi_{i,l}(k), \psi_{i,l}(k)$ can be defined as

$$\varphi_{i,l}(k) = 2^{\frac{i}{2}} h_i(k - 2^i l)$$
$$\psi_{i,l}(k) = 2^{\frac{i}{2}} g_i(k - 2^i l) \quad (5)$$

where the factor $2^{i/2}$ is inner product normalization, i and l are the scale parameter and the translation parameter, respectively. The DWT decomposition can be described as:

$$a_i(l) = x(k)\varphi_{i,l}(k)$$
$$d_i(l) = x(k)\psi_{i,l}(k) \quad (6)$$

where $a_i(l)$ and $d_i(l)$ are the approximation coefficients and the detail coefficients at resolution, i, respectively (Daubechies,1990 and 1992), (Solttani, 2002).

One area in which the DWT has been particularly successful is the epileptic seizure detection because it captures transient features and localises them in both time and frequency content accurately. DWT analyses the signal at different frequency bands, with different resolutions by decomposing the signal into a coarse approximation and detail information. DWT employs two sets of functions called scaling functions and wavelet functions, which are related to low-pass and high-pass filters, respectively. The decomposition of the signal into the different frequency bands is merely obtained by consecutive high-pass and low-pass filtering of the time domain signal. The procedure of multi-resolution decomposition of a signal x[n] is schematically shown in Fig 4. Each stage of this scheme consists of two digital filters and two down-samplers by 2. The first filter, h[.] is the discrete mother wavelet, high-pass in nature, and the second, g[.] is its mirror version, low-pass in nature. The down-sampled outputs of first high-pass and low-pass filters provide the detail, D1 and the approximation, A1, respectively. The first approximation, A1 is further decomposed and this process is continued as shown in Fig 4

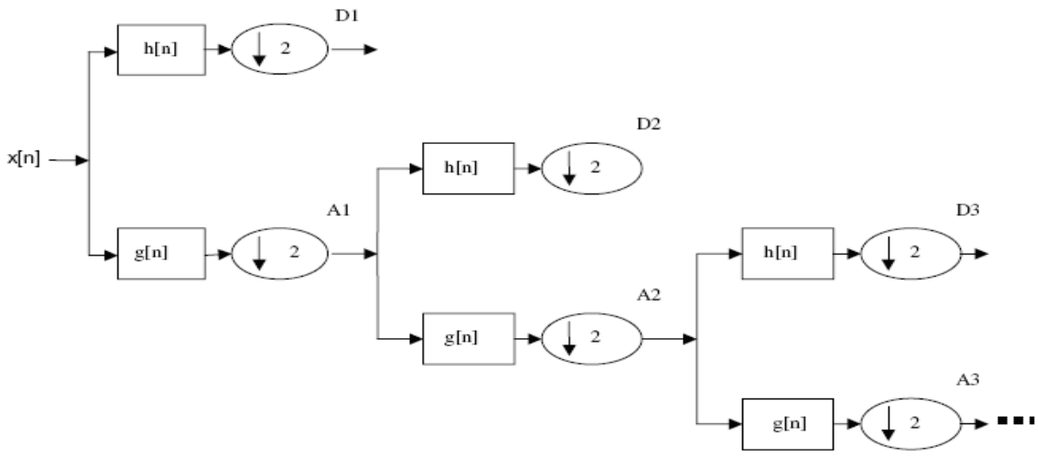


Fig. 4. Sub-band decomposition of DWT implementation; h[n] is the high-pass filter, g[n] the low-pass filter.

Selection of suitable wavelet and the number of decomposition levels is very important in analysis of signals using the DWT. The number of decomposition levels is chosen based on the dominant frequency components of the signal. The levels are chosen such that those parts of the signal that correlates well with the frequencies necessary for classification of the signal are retained in the wavelet coefficients.
It should also be emphasized that the WT is appropriate for analyses f non-stationary signals, and this represents a major advantage over spectral analysis.
One area in which the DWT has been particularly successful is the epileptic seizure detection because it captures transient features and localises them in both time and frequency content accurately. DWT analyses the signal at different frequency bands, with different resolutions by decomposing the signal into a coarse approximation and detail information. DWT employs two sets of functions called scaling functions and wavelet functions, which are related to low-pass and high-pass filters, respectively. The decomposition of the signal into the different frequency bands is merely obtained by consecutive high-pass and low-pass filtering of the time domain signal. The procedure of multi-resolution decomposition of a signal x[n] is schematically shown in Fig. 3. Each stage of this scheme consists of two digital filters and two down-samplers by 2. The first filter, h[.] is the discrete mother wavelet, high-pass in nature, and the second, g[.] is its mirror version, low-pass in nature. The down-sampled outputs of first high-pass and low-pass filters provide the detail, D1 and the approximation, A1, respectively.
Selection of suitable wavelet and the number of decomposition levels is very important in analysis of signals using the DWT. The number of decomposition levels is chosen based on the dominant frequency components of the signal. The levels are chosen such that those parts of the signal that correlates well with the frequencies necessary for classification of the signal are retained in the wavelet coefficients. In the present study, since the EEG signals do not have any useful frequency components above 30 Hz, the number of decomposition levels was chosen to be 4. Thus, the EEG signals were decomposed into details D1–D4 and one final approximation, A4. Usually, tests are performed with different types of wavelets and the one, which gives maximum efficiency, is selected for the particular application. The smoothing feature of the Daubechies wavelet of order 2 (db2) made it more appropriate to detect changes of EEG signals. Hence, the wavelet coefficients were computed using the db4 in the present study. The proposed method was applied on both data sets of EEG data (Sets A and E). Fig. 5 shows approximation (A4) and details (D1–D4) of an epileptic EEG signal (Jahankhani et al (2005a),.

3.1 Feature extraction

The extracted wavelet coefficients provide a compact representation that shows the energy distribution of the EEG signal in time and frequency. Table 1 presents frequencies corresponding to different levels of decomposition for Daubechies order-2 wavelet with a sampling frequency of 173.6 Hz. In order to further decrease the dimensionality of the extracted feature vectors, statistics over the set of the wavelet coefficients was used (Kandaswamy et al, 2004). The following statistical features were used to represent the time frequency distribution of the EEG signals:

- Maximum of the wavelet coefficients in each sub-band.
- Minimum of the wavelet coefficients, in each sub-band.
- Mean of the wavelet coefficients in each sub-band
- Standard deviation of the wavelet coefficients in each sub-band

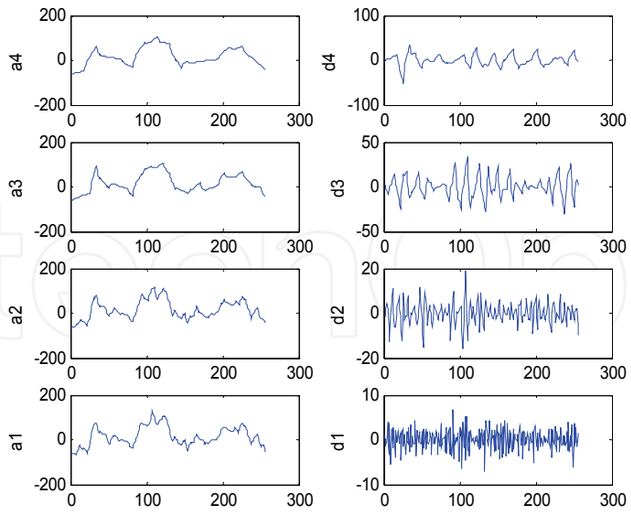
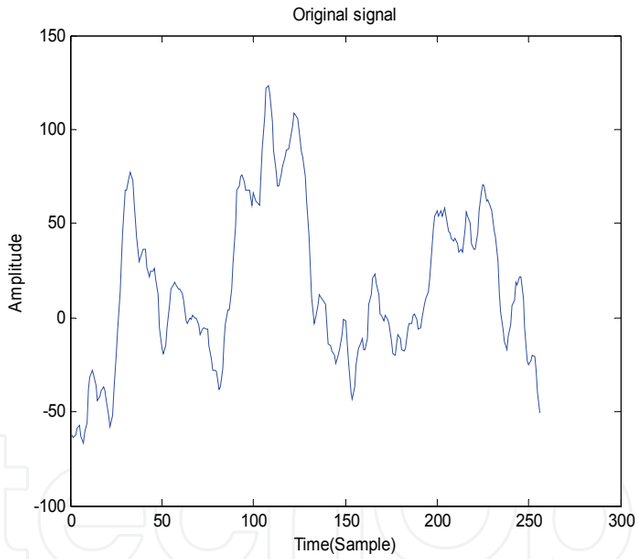


Fig. 5. Approximate and detailed coefficients of EEG signal taken from unhealthy subject (epileptic patient).



Extracted features for two-recorded class A and E shown in Table 2. The data was acquired using a standard 100 electrode net covering the entire surface of the calvarium Fig. 1).

Decomposed signal	Frequency range (Hz)
D1	43.4–86.8
D2	21.7–43.4
D3	10.8–21.7
D4	5.4–10.8
D5	2.7–5.4
A5	0–2.7

Table 1. Frequencies corresponding to different levels of decomposition

The total recording time was 23.6 seconds, corresponding to a total sampling of 4,096 points. To reduce the volume of data, the sample (time points) was partitioned into 16 windows of 256 times points each. From these sub-samples, we performed the DWT and derived measures of dispersion statistics from these windows (each corresponding to approximately 1.5 seconds). The DWT was performed at 4 levels, and resulted in five sub-bands: d1-d4 and a4 (detail and approximation coefficients respectively). For each of these sub-bands, we extracted four measures of dispersion, yielding a total of 20 attributes per sample window. Since our classifiers use supervised learning, we must also provide the outputs, which was simply a class label (for the experiments presented in this paper (Jahankhani et al (2005b), and (Jahankhani et al (2005c), there were 2, corresponding to classes A and E).

Data	Extracted features	Sub-band D1	Sub-band D2	Sub-band D3	Sub-band D4	Approxi-mation
Set A	Max.	28.1094	101.757	131.0846	124.377	114.138
	Min.	-28.4010	-60.813	-149.072	-158.797	-109.521
	Mean	-0.0022	0.0058	-0.0035	0.0388	3.7950
	Std dev.	5.1818	13.6442	23.3685	24.7933	35.1465
Set E	Max	123.3921	278.924	429.6621	375.0564	582.3167
	Min	-90.7055	-238.51	-417.120	-468.064	-361.2154
	Mean	0.0131	-0.0281	-0.0359	-0.0071	-5.5526
	Std dev.	11.8488	35.9941	73.7659	78.1432	180.4493

Table 2. The extracted features of two windows from A & E classes

4. Intelligent classifiers

Recently, the concept of combining multiple classifiers has been actively exploited for developing highly reliable "diagnostic" systems [12]. One of the key issues of this approach is how to combine the results of the various systems to give the best estimate of the optimal result. A straightforward approach is to decompose the problem into manageable ones for several different sub-systems and combine them via a gating network. The presumption is that each classifier/sub-system is "an expert" in some local area of the feature space. The sub-systems are local in the sense that the weights in one "expert" are decoupled from the weights in other sub-networks. In this study, 16 subsystems have been developed, and each of them was associated with the each of the windows across each electrode (16/electrode). Each subsystem was modelled with an appropriate intelligent learning scheme. In our case, two alternative schemes have been proposed: the classic MLP network and the RBF network using the orthogonal least squares learning algorithm. Such schemes provide a degree of certainty for each classification based on the statistics for each plane. The outputs of each of these networks must then be combined to produce a total output for the system

4.1 Learning Vector Quantization (LVQ) network

LVQ algorithm is one of the popular quantization algorithms. LVQ is a method for training competitive layers in a supervised manner. Vector Quantization (VQ) is a form of data compression that represents data vectors by set of codebook vectors. This is the basic idea of

VQ theory, the motivation of which is dimensionality reduction or data compression. Each data vector is then its nearest codebook vector then represents vector. VQ methods are closely related to certain paradigms of self-organising Neural Networks. LVQ can be understood as a special case of an artificial NN, more precisely, it applies a winner-take-all Hebbian learning-based approach.
In effect, the error of the VQ approximation is the total squared distance

$$D = \sum_{x} || x - w_{I(x)} ||^2 \tag{7}$$

Between the input vectors $\{x\}$ and their codebook vectors$\{w_{Ix}\}$, and we clearly wish to minimize this.
Among the many prototype-based learning algorithm proposed recently, LVQ algorithms developed by Kohonen are a family of training algorithms for the nearest-neighbour classifications, which include LVQ1, LVQ2 and its improved versions LVQ2.1, LVQ3 algorithms (Kohonen.T et al, 1988), (Kohonen.T et al, 2001) and (Kohonen.T et al, 2001) The family of LVQ algorithms is widely used for pattern classification, and found a very important role in statistical pattern classification (Kohonen.T et al, 1988), (Kohonen.T et al, 2001) and signal processing achieved satisfactory results than other neural network classifiers in spite of their simple and time efficient training process.
A competitive layer automatically learns to classify input vectors. However, the classes that the competitive layer finds are dependent only on the distance between input vectors. If two input vectors are very similar, the competitive layer probably will put them in the same class. There is no mechanism in a strictly competitive layer design to say whether or not any two input vectors are in the same class or different classes.
A neural network for learning VQ consists of two layers: an input layer and an output layer. It represents a set of reference vectors, the coordinates of which are the weights of the connections leading from the input neuron to an output neuron.
Advantages of LVQ are as follows:

- The model is trained significantly faster than other neural network techniques like Back Propagation.
- Reducing large datasets to a smaller number of codebook vectors for classification
- There is no limit in the number of dimensions in the codebook vectors like nearest neighbour techniques.
- Normalisation of input data is not required.
- Can handle data with missing values.

LVQ has some disadvantages

- Generate useful distance measures for all attributes
- The accuracy is highly depends on the initialisation of the model as well as the learning parameters such as learning rate and training iteration
- Also accuracy is depending on the class distribution in the training dataset.
- It is difficult to determine a good number of codebook vectors for a given problem.

4.1.1 LVQ algorithms

Assume for a one-dimensional case, there are two classes C_1 and C_2 with mean vectors (reference vectors) m_k and an input vector x belong to the same class to which the nearest reference vector belongs.

The learning method of LVQ is often called competition learning, because it works as follows:
For each training pattern the reference vector that is closest to it is determined.
For each data point, the neuron that is closest to its determined called the winner neuron
The weight of the connections to this neuron and winner neuron is then adapted, ie made closer if it correctly classifieds the data point (winner, takes all).
The movement of the reference vector is controlled by a parameter called the learning rate.
The VQ methods are closely related to certain paradigms of self-organising Neural Networks.
LVQ Kohonen ,T,1988), found a very important role in statistical pattern classification (Kohonen ,T,2001).
A detailed description of LVQ training algorithm can be found in (Crammer et al 2002).
LVQ can be included in a broad family of learning algorithms based on Stochastic Gradient Descent In the 1980's, Kohonen proposed a number of improvements in his algorithm generating the LVQ1, LVQ2, LVQ2.1, and LVQ3 (Crammer et al 2002 and Kohonen ,T(2001),
Categorisation of signal patterns is one of the most usual Neural Network (NN) applications.

4.1.2 LVQ2.1

LVQ2.1 is an improved version of LVQ2, which aims at eliminating the detrimental effect described in previous section.
The nearest neighbours, m_1 and m_2, are updated simultaneously; one of them must belong to the correct class and other to wrong class, respectively. Moreover, x must fall into a "window", which is defined around the mid plane of $m_1$and m_2 .

$$m_i \leftarrow m_i + \alpha(t)(x - m_i) \tag{8}$$

$$m_j \leftarrow m_j - \alpha(t)(x - m_j) \tag{9}$$

The "window" is defined using a relative window width w as follows:

$$\min(\frac{d_i}{d_j}, \frac{d_j}{d_i}) > s \tag{10}$$

$$s = \frac{1-w}{1+w} \tag{11}$$

Where $d_i = \| x - m_i \|, d_j = \| x - m_j \|$

4.1.3 LVQ3

The LVQ2.1 has a drawback that reference vectors diverge during learning. In LVQ3 algorithm, corrections are introduced to the LVQ2.1 algorithm to ensure that the reference vectors continue approximating the class distributions.

$$m_i \leftarrow m_i + \alpha(t)(x - m_i) \tag{12}$$

$$m_j \leftarrow m_j - \alpha(t)(x - m_j) \tag{13}$$

Where m_i and m_j are the two nearest neighbours, where by x and m_i belong to the same class, while x and m_j belong to different classes, respectively, also x must be fall into the "window"

$$m_k \leftarrow m_k - \varepsilon\alpha(t)(x - m_k), \varepsilon > 0 \tag{14}$$

For $\kappa\varepsilon\{i, j\}$, if x, m_i and m_j belong to the same class.

In this study we used LVQ2.1 algorithm to locate the nearest two exemplars to the training case.

Step1: Initialise the codebook (centre) vector.

Step2: Find

$$\min(\frac{d_i}{d_j}, \frac{d_j}{d_i}) > s \tag{15}$$

$$s = \frac{1-w}{1+w} \tag{16}$$

where d_i and d_j are the Euclidean distances from 2 classes and window(w) in range 0.2 to 0.3 around the mid-plane of neighbouring codebook vectors m_i and m_j and m_i belong to class C_i and m_j to class C_j respectively.

Step3: Update the centres (codebook vectors) at each step, namely, the "winner" and the "runner-up"

$$m_i(t+1) \leftarrow m_i(t) + \alpha(t)(x(t) - m_i(t)) \tag{17}$$

$$m_j(t+1) \leftarrow m_j(t) - \alpha(t)(x(t) - m_j(t)) \tag{18}$$

where x is the input vector.

5. Rough sets

Rough set theory is a relatively new data-mining technique used in the discovery of patterns within data first formally introduced by (Pawlak, 1982 ,Pawlak, 1991). It's deals with the classificatory analysis of data tables. The data can be acquired from measurements and in principle it must be discrete. The main goal of the rough set theory is the " automated transformation of data into knowledge".

It has a wide range of uses, such as medical data analysis, stock market prediction and financial data analysis, information retrieval systems, voice recognition, and image processing. Rough sets are especially helpful in dealing with vagueness and uncertainty in decision situations.

The overall modelling process typically consists of a sequence of several sub steps that all require various degrees of tuning and fine-adjustments. An important feature of rough sets is that the theory is followed by practical implementations of toolkits that support interactive model development.

The first step in the process of mining any dataset using rough sets is to transform the data into a decision table. In a decision table (DT), each row consists of an observation (also called an object) and each column is an attribute, One attribute is chosen or created for the decision attribute (or dependent attribute). The rest of the attributes are the condition attributes (independent attributes).

Formally, a DT is a pair $A = (U, A \cup \{d\})$ where $d \varpi A$ is the *decision attribute, U* is a finite non-empty set of objects called the *universe* and A is a finite non-empty set of attributes such that $a:U \rightarrow V_a$ is called the value set of a. Once the DT has been produced, the next stage entails cleansing the data.

There are several issues involved in small datasets – such as missing values, various types of data (categorical, nominal and interval) and multiple decision classes. Each of these potential problems must be addressed in order to maximise the information gain from a DT. Missing values is very often a problem in biomedical datasets and can arise in two different ways. It may be that an omission of a value for one or more subject was intentional – there was no reason to collect that measurement for this particular subject (i.e. 'not applicable' as opposed to 'not recorded'). In the second case, data was not available for a particular subject and therefore was omitted from the table. We have 2 options available to us: remove the incomplete records from the DT or try to estimate what the missing value(s) should be. The first method is obviously the simplest, but we may not be able to afford removing records if the DT is small to begin with. So we must derive some method for filling in missing data without biasing the DT. In many cases, an expert with the appropriate domain knowledge may provide assistance in determining what the missing value should be – or else is able to provide feedback on the estimation generated by the data collector. In this study, we employ a conditioned mean/mode fill method for data imputation. In each case, the mean or mode is used (in the event of a tie in the mode version, a random selection is used) to fill in the missing values, based on the particular attribute in question, conditioned on the particular decision class the attribute belongs to. There are many variations on this theme, and the interested reader is directed to (Pawlak, 1982 ,Pawlak, 1991) for an extended discussion on this critical issue. Once missing values are handled, the next step is to discretise the dataset. Rarely is the data contained within a DT all of ordinal type – they generally are composed of a mixture of ordinal and interval data. Discretisation refers to partitioning attributes into intervals – tantamount to searching for "cuts" in a decision tree. All values that lie within a given range are mapped onto the same value, transforming interval into categorical data. As an example of a discretisation technique, one can apply equal frequency binning, where a number of bins *n* is selected and after examining the histogram of each attribute, n-1 cuts are generated so that there is approximately the same number of items in each bin. See the discussion in (Pawlak, 1991) for details on this and other methods of discretisation that have been successfully applied in rough sets. Now that the DT has been pre-processed, the rough sets algorithm can be applied to the DT for the purposes of supervised classification.

Rough sets generates a collection of 'if..then..' decision rules that are used to classify the objects in the DT. These rules are generated from the application of reducts to the decision

table, looking for instances where the conditionals match those contained in the set of reducts and reading off the values from the DT. If the data is consistent, then all objects with the same conditional values as those found in a particular reduct will always map to the same decision value. In many cases though, the DT is not consistent, and instead we must contend with some amount of indeterminism. In this case, a decision has to be made regarding which decision class should be used when there are more than 1 matching conditioned attribute values. Simple voting may work in many cases, where votes are cast in proportion to the support of the particular class of objects. If the rules are too detailed (i.e. they incorporate reducts that are maximal in length), they will tend to overfit the training set and classify weakly on test cases. What are generally sought in this regard are rules that possess low cardinality, as this makes the rules more generally applicable.

6. Principal Component Analysis (PCA)

Dimensionality reduction techniques aim to determine the underlying true dimensionality of a discrete sampling X of an n-dimensional space. That is if X embedded in a subspace of dimensionality m, where m<n, then we can find a mapping F: X→Y such that Y⊂B is a m dimensional manifold. The commonly used method to find such mapping is Principal Component Analysis (PCA).

Reducing the dimensionality of the problem simplifies the task of the classifier, and alleviates the generalization problems due to the curse of dimensionality. Depending on the nature of a given classification problem: the raw data, the chosen features, and classifier. Dimensionality reduction is almost essential when time-frequency representation is used as feature basis. There are numerous books and articles reviewed traditional and current state-of-the art dimension reduction methods published in the statistics and signal processing and machine learning literature (Jollife,2002), (Holtell, 1933), (Lay,2000).

The basic idea is to reduce the dimensionality of dataset D={x_i , i=1..n} from k features (attributes) in D to some m<k, in an optimal way. PCA is a well-known technique in statistics and data compression. PCA is also known as Karhunen_Loeve transformation

The main use of PCA is to reduce the dimensionality of a data set while retaining the most information. PCA's effectiveness in pattern recognition is due to its ability to eliminate linear dependencies and uncorrelated noise in the data.

The importance of PCA is due to several factors.

By capturing directions of maximum variance in the data, the principal component offers a way to compress the data with minimum information loss.

The principal components are uncorrelated, which can aid with interpretation or subsequent statistical analysis.

One limitation of PCA is that it does not model non-linear relationships among variables efficiently.

Computation of the principal components can be presented with the following algorithm:

Calculate the covariance matrix from the input data.

$$Cov(X_j, X_k) = \frac{\sum_{i=1}^{n} (X_{ij} - \overline{X}_j)(X_{ik} - \overline{X}_k)}{(n-1)} \tag{19}$$

Where

$$\bar{X} = \frac{\sum_{i=1}^{n} X_{ij}}{n} \quad \text{and j, k = 1, 2,....., p} \tag{20}$$

The covariance matrix then has the following form:

$$C = \begin{bmatrix} C_{11}C_{12}C_{13}.....C_{1p} \\ C_{21}C_{22}C_{23}.....C_{2p} \\ ..... \\ ..... \\ ..... \\ C_{p1}C_{p2}C_{p3}.....C_{pp} \end{bmatrix}$$

Where C is the covariance matrix, C_{jk} is the covariance of variable X_j and X_k when $j \neq k$ and the diagonal element C_{jk} is the variance of variable X_j when j=k.
There are several properties of covariance matrix Cx:
Cx is a squre symmetric $m \times m$ matrix.
The diagonal terms of Cx are the variance of particular measurement types.
The off-diagonal terms of Cx are the covariance between measurement types.
PCA is finding the eigenvalues and eigenvectors of the sample correlation matrix. Eigenvectors (PCs) and their associated eigenvalues can be calculated from the correlation matrix.
The basic equation in eigen value is

$$Ax=\lambda x \tag{21}$$

This can only hold if

$$\text{Det } |A-\lambda I| = 0 \tag{22}$$

Where λ is an eigenvalue of the matrix A and x the corresponding eigenvector.
Equation 21 can be expressed in matrix form with a matrix V whose columns contain the eigenvectors and diagonal matrix D with the eigenvalues in the diagonal:

$$AV=VD$$

Compute the eigenvalues and eigenvectors and then sort them in a descending order with respect to eigenvalues. Each eigenvalue represents the amount of variance that has been captured by one component.
The first principal component PC_1 is then a linear combination of the original variables $X1$, X_2, ……..X_p

$$PC_1 = a_{11}X_1 + a_{12}X_2 + a_{13}X + ...a_{1p}X_p = \sum_{j=1}^{p} a_{ij}X_j \tag{23}$$

That varies as much as possible for the individuals, subject to the condition that

$$a_{11}{}^2 + a_{12}{}^2 + a_{13}{}^2 + ...a_{1p}{}^2 = 1 \quad (24)$$

where $a_{11}, a_{12}, ..., a_{1p}$ are coefficients assigned to the original p variables for PC_1 . Therefore, the eigenvalue of PC_1 is as large as possible given this constrain on the constant a. Constrain must be imposed in order to avoid the increasing of the eigenvalue of PC_1 by simply increasing one or more of the values.
Similarly, the second principal component,

$$PC_2 = a_{21}X_1 + a_{22}X_2 + a_{23}X + ...a_{2p}X_p \quad (25)$$

Is such that eigenvalues of PC_2, is as large as possible subject to the constraint that:

$$a_{21}{}^2 + a_{22}{}^2 + a_{23}{}^2 + ...a_{2p}{}^2 = 1 \quad (26)$$

and also on the condition that PC2 is uncorrelated with PC1. Other principal components can be expressed in a similar way. There can be up to p principal components if there are p variables.
If our p variables share considerable variance, several of the p components should have large eigenvalues and others have small eigenvalues. To decide how many components to retain, one of the rule is to retain only components with eigenvalues of one or more, and drop any component that account for less variance than does a single variable. Another method for deciding on the number of components to retain is the screen plot. This is a plot with eigenvalues on the ordinate and component number on the abscissa. The plot provides a visual aid for deciding at what point including additional components no longer increases the amount of variance accounted for by a nontrivial amount. Fig. 6 shows the screen plot produced by SPSS.

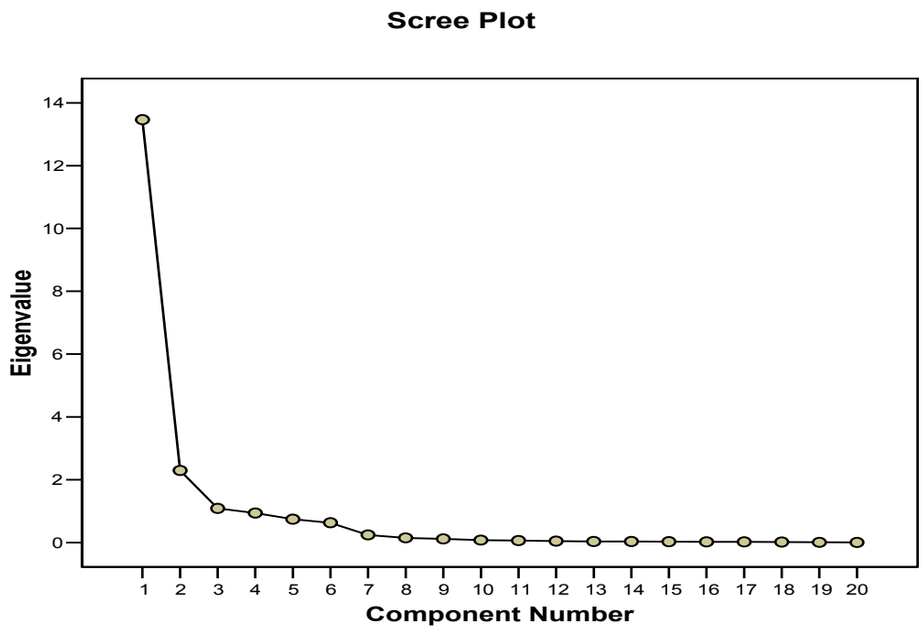


Fig. 6. Screen plot of the correlation matrix, using SPSS.

The lower-dimensional (reduced) matrix can be obtained by multiply the original feature space with the obtained transition matrix (number of component).
The PCA dimensionality reduction method is fast to compute, simple to implement, and since the optimisation do not involve local minima. One of the limitations of PCA is that their effectiveness is limited by the fact that PCA is globally linear method.
Computation of the principal components can be presented with the following algorithm:
Calculate the covariance matrix from the input data.
Compute the eigenvalues and eigenvectors and then sort them in a decending order with respect to the eigenvalues.
From the actual transition matrix by taking the predefined number of components (eigenvectors)
The lower-dimensional matrix can be obtained by multiply the original feature space with the obtained transition matrix.
Table 3 shows how the variation is partitioned between the 8 factors.

7. Results

The proposed diagnostic system consists of a pre-processing /feature selection and one classifier subsystem. Duabechies Wavelets order-2 with 4 levels has been used for pre-processing in order to achieve the same dimensionality reduction of wavelet coefficients. In this work, the 100 time series of 4096 samples for each class partitioned by a rectangular window composed of 256 discrete data and then training and test sets were formed by 3200 vectors (1600 vectors from each class) of 20 dimensions (dimension of the extracted feature vectors). The proposed multi-classifier scheme consists of 16 sub-systems/classifiers. For each one of these sub-systems, LVQ network structure has been utilized. The average concept of combining the individual output of the 16 classifiers has been adopted in this study. The architecture of LVQ is based on straightforward approach with 20 input and two outputs, with 2000 epochs training. The 20 inputs correspond to the four features times the number of wavelet decomposition (D1-D4 & A4).

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	13.404	67.019	67.019	13.404	67.019	67.019
2	2.302	11.509	78.528	2.302	11.509	78.528
3	1.125	5.624	84.152	1.125	5.624	84.152
4	.905	4.525	88.677	.905	4.525	88.677
5	.818	4.092	92.769	.818	4.092	92.769
6	.620	3.098	95.867	.620	3.098	95.867
7	.213	1.066	96.933	.213	1.066	96.933
8	.148	.741	97.674	.148	.741	97.674

Table 3. This table displays the total variance for the first 8 principle components contained within the original dataset.

	Component Matrix (a)							
	1	2	3	4	5	6	7	8
D1	.926	-.008	-.040	-.028	.244	-.025	-.151	-.115
D1	-.945	.002	.070	-.008	-.172	.032	.109	.112
D1	.164	.896	-.054	-.028	-.137	-.373	-.013	.010
D1	.948	.002	-.018	.044	.240	-.047	.116	.111
D2	.953	-.019	-.043	-.019	.225	-.003	-.097	-.019
D2	-.959	.002	.079	-.015	-.183	.030	.081	.015
D2	-.116	-.965	-.113	.104	.000	.141	.010	-.004
D2	.956	.007	-.032	.047	.209	-.036	.143	.115
D3	.970	.000	-.031	-.008	.141	.005	-.006	-.013
D3	-.973	.006	.063	-.016	-.088	.016	-.001	-.013
D3	-.083	.733	.084	.096	.136	.648	-.010	.013
D3	.972	.006	-.021	.033	.085	-.004	.194	.067
D4	.943	.001	.023	-.079	-.236	.060	.037	-.141
D4	-.937	.009	-.010	.016	.273	-.076	-.047	.103
D4	.021	-.114	.767	-.611	.151	-.008	.023	-.005
D4	.923	.016	.002	-.041	-.246	.054	.214	-.127
A	.885	-.021	.106	.005	-.369	.065	-.125	.076
A	-.920	.113	-.070	.071	.206	-.036	.151	-.152
A	.153	-.032	.685	.700	.025	-.113	-.008	-.040
A	.927	-.058	.075	.015	-.281	.063	-.044	.118

Table 4. Principal Component Analysis, displaying the first 8 components that were extracted (in the form of a component matrix). The symbols in the left most column refer to the level of the detail (D) or approximation (A) coefficients. The order is the following: Max, Min, Mean, and Standard Deviation.

LVQ Applied to	Class A	Class E
All A & E Attributes	1600	1558
Attribute MaxD4	1600	1596
8 attributes using PCA	1568	1558
Rough sets	1600	1600

Table 5. Summary of the correctly classified objects in the testing set for each of the classification algorithms employed in this study (note the maximum number was 1,600)

8. Conclusions

The results from this study indicate that the hybrid approach to the classification of a complex dataset such as an EEG time series can be achieved with a high degree of accuracy. This dataset contains both a spatial and a temporal component – the electrodes are placed on spatially distinct regions of the calvarium. There are several diseases that yield a characteristic signature that can be detected reproducibly using standard EEG equipment. For instance epilepsy yields a characteristic change in the power spectrum within the

temporal lobe region. This would indicate that there would be a spatial signal that requires proper spatial localisation within the appropriate brain region. In addition, symptoms may change over time – and thus the temporal resolution of the recording must be such that it is samples at the correct frequency – without yielding Nyquist or other sampling errors. In the present work, we employed a discrete wavelet transform to the dataset in order to extract temporal information in the form of changes in the frequency domain over time – that is they are able to extract non-stationary signals embedded in the noisy background of the human brain. In this study, we examined the difference(s) between normal and epileptic EEG signals – over a reasonable duration of approximately 24 seconds. We extracted statistical information from the wavelet coefficients, which we used as inputs to a set of supervised learning algorithms – LVQ 2.1 based neural networks. The attributes (inputs) used were measures of dispersion – which captured the statistical variations found within the particular time series. The results from this preliminary study will be expanded to include a more complete range of pathologies. In this work, we focused on the extremes that are found within the EEG spectrum – normal and epileptic time series. These two series were chosen as they would more than likely lead to the maximal dispersion between the 2 signals and is amenable for training of the classifiers. In the next stage of this research, we have datasets that are intermediate in the signal changes they present. This will provide a more challenging set of data to work with – and will allow us to refine our learning algorithms and/or approaches to the problem of EEG analysis. In a future work, we will also investigate additional pre-processing steps such as clustering techniques.

9. Acknowledgements

The authors wish to thank Andrzejak et al., 2001 for making the data publicly available at (http://www.meb.uni-bonn.de/epile-ptologie/science/physik/eegdata.html

10. References

Adeli H, Zhou Z, Dadmehr N.(2003) Analysis of EEG records in an epileptic patient using wavelet transform. J Neurosci Methods;123(1):69–87.

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

Daubechies I (1992), Ten Lectures on Wavelets, CBMS-NSF Regional , Conference Series in Applied Mathematics, Vol. 61, Capital City Press, Montpelier, Vermont

Daubechies I, (1990), The wavelet transform, time-frequency localization and signal analysis. IEEE T Inform Theory ; 36(5); pp 961-1005.

Folkers, A., Mosch, F., Malina, T., & Hofmann, U. G.(2003) Real-time bioelectrical data acquisition and processing from 128 channels utilizing the wavelet-transformation. Neurocomputing, 52–54, 247–254.

Guler I, Ubeyli ED.(2004) Application of adaptive neuro-fuzzy inference system for detection of electrocardiographic changes in patients with partial epilepsy using feature extraction. Expert Syst Appl;27(3):323–30.

Guler, I., Kiymik, M. K., Akin, M., & Alkan, A. (2001) AR spectral analysis of EEG signals by using maximum likelihood estimation. Computers in Biology and Medicine, 31, 441–450.

Jahankhani[1], K. Revett[2], V. Kodogiannis[1] , (2005 a), Automatic Detection Of EEG Abnormalities Using Wavelet Transforms, [1]Mechatronics Group, [2]Artificial Intelligent and Multimedia School of Computer ScienceUniversity of Westminster,

London HA1 3TP UNITED KINGDOM WSEAS Transactions on Signal Processing Issue 1, Volume 1, ISSN 1790-5022, pp 55-61

Jahankhani [1], Ken Revett[2], Lynne Spackman3, Vassilis Kodogiannis [1]and Stewart Boyd[3,] (2005 b) An Automated Anomaly EEG Detection Algorithm Using Discrete Wavelet Transforms Proceeding of the 2nd Indian International Conferrence on Aritficial Intelligence, ISBN 0-9727412-1-6 1.Mechatronics Group, School of Computer Science, University of Westminster, London HA1 3TP, UK 2. Artificial Intelligent and Multimedia, School of Computer Science, University of Westminster,London HA1 3TP, UK 3. Great Ormond Street Hospital for Children NHS Trust, Great Ormond Street, London WC1N 3JH, UK

Jahankhani [1], Kenneth Revett[2], Vassilis Kodogiannis [1], (2005 c) Detecting Clinically Relevant EEG Anomalies Using Discrete Wavelet Transforms [1]Mechatronics Group, [2]Artificial Intelligent and Multimedia School of Computer Science, University of Westminster, London HA1 3TP, UK Proceedings of the WSEAS International Conference 5TH WSEAS Int.Conf. on WAVELET ANALYSIS AND MULTIRATE SYSTEMS (WAMUS'05), Sofia, Bulgaria, October 27-29, ISBN: 960-8457-36-X

Jolliffe, T.(2002). Principal Component Analysis, (2nd ed.), New York: Springer-Verlag. Kandaswamy, A., Kumar, C. S., Ramanathan, R. P., Jayaraman, S., & Malmurugan, (2004) Neural classification of lung sounds using wavelet coefficients. Computers in Biology and Medicine, 34(6), 523–537.

Kohonen ,T,(2001), self-Organising Maps, Berlin Heidelberg: Springer-Verlag, third ed.

Kohonen. T,(1988), An Introduction to Neural Computing, Neural Networks pp. 3-16

Crammer,K., R. Gilad-Bachrach, and A.Tishby, (2002) Marging Analysis of the LVQ Algorithm, Proc. 15th Ann.Conf. Neural Information Processing systems.

Bottou,L., (2004), Stochastic Learning, Lecture Notes in Artificial Intelligence, Vol. 3176, pp 146-168

Lay, David, (2000), Linear Algebra and its Applications. Addison-Wesley, New York

Boulougoura,M., E. Wadge, V.S. Kodogiannis, H.S. Chowdrey,(2004), Intelligent systems for computer-assisted clinical endoscopic image analysis, 2nd IASTED Int. Conf . on BIOMEDICAL ENGINEERING, Innsbruck, Austria, pp. 405-408.

Pawlak, Z. (1991): Rough sets – Theoretical aspects of reasoning about data.

Pawlak, Z. (1982), Rough Sets, International Journal of Computer and Information Sciences, 11, pp. 341-356.

Petrosian, A., Prokhorov, D., Homan, R., Dashei, R., & Wunsch, D.(2000) Recurrent neural Network based prediction of epileptic seizures in intraand extracranial EEG. Neurocomputing, 30, 201–218.

Pradhan, N., Sadasivan, P. K., & Arunodaya, G. R.(1996) Detection of seizure activity in EEG by an artificial neural network: A preliminary study. Computers and Biomedical Research, 29, 303–313.

Soltani S.(2002), On the use of the wavelet decomposition for time series prediction. Neurocomputing ; 48:267-77

Subasi, A.(2005), Automatic recognition of alertness level from EEG by using neural network and wavelet coefficients. Expert Systems with Applications, 28, 701–711.

Weng, W., & Khorasani, K. (1996) An adaptive structure neural network with application to EEG automatic seizure detection. Neural Networks, 9, 1223–1240.

Zoubir, M., & Boashash, B., (1998), Seizure detection of newborn EEG using a model approach. IEEE Transactions on Biomedical Engineering, 45, 673–685.

Data Mining in Medical and Biological Research

Edited by Eugenia G. Giannopoulou

ISBN 978-953-7619-30-5
Hard cover, 320 pages
Publisher InTech
Published online 01, November, 2008
Published in print edition November, 2008

This book intends to bring together the most recent advances and applications of data mining research in the promising areas of medicine and biology from around the world. It consists of seventeen chapters, twelve related to medical research and five focused on the biological domain, which describe interesting applications, motivating progress and worthwhile results. We hope that the readers will benefit from this book and consider it as an excellent way to keep pace with the vast and diverse advances of new research efforts.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Pari Jahankhani and Vassilis Kodogiannis (2008). Intelligent Decision Support System for Classification of Eeg Signals Using Wavelet Coefficients, Data Mining in Medical and Biological Research, Eugenia G. Giannopoulou (Ed.), ISBN: 978-953-7619-30-5, InTech, Available from: http://www.intechopen.com/books/data_mining_in_medical_and_biological_research/intelligent_decision_support_system_for_classification_of_eeg_signals_using_wavelet_coefficients

INTECH
open science | open minds

InTech Europe
University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China
Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821

© 2008 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike-3.0 License, which permits use, distribution and reproduction for non-commercial purposes, provided the original is properly cited and derivative works building on this content are distributed under the same license.

IntechOpen

IntechOpen