Classification of EEG signals for epileptic seizure prediction using ANN

JasminKevric, AbdulhamitSubasi

International Burch University, Faculty of Engineering and Information Technologies,

FrancuskeRevolucije bb, Ilidža, Sarajevo, 71210, Bosnia and Herzegovina.

E-mails:*jkevric@ibu.edu.ba*, *asubasi@ibu.edu.ba*

Abstract

In this paper, we developed a model for classification of EEG signals. The aim of the study is to determine whether this model can be used for epileptic seizure prediction if “pre-ictal” stages were successfully detected. We analyzed long-term Freiburg EEG data. Each of 21 patients contains datasets called “ictal” (seizure) and “inter-ictal” (seizure-free). We extracted 4096-samples (or 16 seconds) long segments from both datasets of each patient. These segments were decomposed into time-frequency representations using Discrete Wavelet Transform (DWT). The statistical features from the DWT sub-bands of EEG segments were calculated and fed as inputs to Multilayer Perceptron (MLP) and Radial Basis Function (RBF) network classifiers using 10-fold cross validation. We also applied multiscale PCA (MSPCA) de-noising method to determine if it can further enhance the classifiers’ performance. MLP-based approach outperformed RBF classifier with or without MSPCA, which significantly improved the classification accuracy of both classifiers. The proposed MLP-approach with MSPCAachieved a classification accuracy of 95.09%. We showed that a high classification accuracy of EEG signals can be accomplished in cases when additional “pre-ictal” class is introduced. Therefore, the proposed approach may become an efficient tool to predict epileptic seizures from EEG recordings.

Keywords: Electroencephalogram (EEG); Epileptic seizure; Discrete Wavelet Transform (DWT); Multilayer Perceptron (MLP); Radial Basis Function (RBF) network; Multiscale PCA (MSPCA); Machine learning.

1.INTRODUCTION

Noninvasive electrodes on the scalp can record the brain's electrical activity called as electroencephalogram (EEG), produced by billions of neurons firing within the nervous system. The EEG signal is characterized by a nonstationarity in the waveforms and semistationary time-dependent states, and detection of these characteristics is a difficult task (Bigan, 1998). Over 50 million people in the world are affected by the epilepsy, the second mostcommon neurological disorder after stroke (D'Alessandro et al., 2003). Abnormal movements and seizures, resulting from the brain cells' excessive electrical discharge, are the signs of epilepsy.

One of the most important causes of stress, morbidness and anxiety in epileptic patients is the inability of predicting seizure onset (Murray, 1993; Buck et al., 1997). Thereliable predictability of seizure onset would dramatically improve the safety and quality of life of these patients who cannot be treated successfully by common therapeutic options (Schachter, 1994). For example, patients would be able to prevent dangerous situations when being warned of upcoming seizure. Various automated intervention systems and measures could be implemented like applying electrical brain stimulations or delivering short-acting anticonvulsant drugs by using implanted devices (Stein et al., 2000;Elger, 2001). Additionally, the investigation of the pathophysiological mechanisms causing seizures could be improved by the accurate detection of states preceding seizures.

Mormann et al., (2007) stated that seizure prediction is the long and winding road in their review article. D'Alessandro et al., (2003)used intelligent genetic search technique to classify preseizure and non-preseizure classes from four patients by a probabilistic neural network, reporting a sensitivity of 62.5% with 90.5% specificity. Costa et al., (2008)compared 6 types of neural network architectures which used 14 features extracted from EEG of two patients to classify brain states into four classes: inter-ictal, pre-ictal, ictal and pos-ictal. The accuracies of up to 99% were achieved. Mirowski et al., (2009)achieved 71% sensitivity and 0 false positives using convolutional networks combined with wavelet coherence. Chisci et al., (2010) used Autoregressive (AR) models to classify pre-ictal and inter-ictal classes from nine patients, reporting 100% sensitivities and average false positive rates of 0.174/h (on the inter-ictal dataset).

This paper is organized as follows. Section 2 describes the EEG data, signal processing and feature extraction methods, and the artificial neural networks with a brief description of each one. In section 3, the performance of the proposed system is presented and discussed. Finally, section 4 presents concluding remarks and perspectives for future work.

2.Materials and methods

2.1 Subjects and data recording

We analyzed long-term EEG data recorded during invasive pre-surgical epilepsy monitoring at the Epilepsy Center of the University Hospital of Freiburg, Germany. The Neurofile NT digital video EEG system with 128 channels, 256 Hz sampling rate, and a 16 bit analogue-to-digital converter was used to acquire the EEG data. Each of 21 patients, suffering from medically intractable focal epilepsy, contains datasets called “ictal” and “inter-ictal”. The “ictal“ dataset consists of files containing epileptic seizures, each having a seizure-free "pre-ictal" period of at least 50 minutes. The “inter-ictal“ dataset consist of approximately one day of seizure-free EEG recordings for each patient. Each patient had between two and five seizures, with an average of 4.2 seizures per patient or a total number of 87 seizures(Maiwald et al., 2004).The onset and end times of each seizure were determined by visual examination of skilled epileptologists.

2.2 Multiscale Principal Component Analysis

Multiscale Principal Component Analysis (MSPCA) combines the wavelet analysis with PCA. The MSPCA method incorporates the decomposition of each variable on a selected family of wavelets during which the wavelet coefficients are thresholded. After that, the PCA model is separately built for the coefficients at each scale. In order to yield one model for all scales together, the models at important scales, which show process disturbances or abnormal operation, are merged in an effective scale-recursive way(Bakshi 1998; Ganesan, Das, & Venkataraman, 2004).Because of its multiscale type, it is suitable to use MSPCA for modeling of data consisting of contributions from events which behavior changes over time and frequency. MSPCA is powerful tool for monitoring autocorrelated measurements without time-series modeling or matrix augmentation due to approximate decorrelation of wavelet coefficients. The MSPCA method not only selects and monitors the significant signal features but also conforms to the nature of the signal (Bakshi 1998).

2.3 Discrete Wavelet Transform

Signals like EEG may contain transitory or non-stationary characteristics. That is why Fourier Transform, which can be applied to the stationary signals, is not an ideal method to be directly applied to signals like EEG. Therefore, time-frequency methods like Wavelet Transform should be used.

The analysis based on Discrete Wavelet Transform is best explained in terms of filter banks. Multi-resolution decomposition of a signal is the procedure of using a group of filters to separate that signal into various spectral components. Every stage of this procedure consists of two digital filters and two down-samplers by 2. The first filter is the discrete mother wavelet, being high-pass in nature. The second filter is its mirror version, being low pass in nature. Outputs of the first high-pass and low-pass filters, once being down-sampled, provide the detail D1 and the approximation A1, respectively (Adeli, Zhou, & Dadmehr, 2003;Marchant, 2003; Semmlow, 2004).

In DWT analysis it is very important to choose the appropriate number of decomposition levels and appropriate wavelet selection. The components of the dominant frequency of the signal are the main base for choosing the number of decomposition levels.

Distribution of energy of the EEG signal in frequency and time is shown by a compact representation of the extracted wavelet coefficients. Using statistics over the wavelet coefficients sets helped in decreasing the dimensionality of the extracted feature vectors (Kandaswamy et al., 2004).Subasi (2007) and Subasi&Gursoy (2010) achieved high accuracies in classifying EEG signals using statistical feature vectors extracted from wavelet coefficients.

2.4 Multilayer Perceptron

Multilayer feedforward networks is composed of a set of source nodes which serve as sensory units that form the input layer, one or more hidden layers and an output layer. Hidden layers and an output layer consist of computational nodes. The input signal is transmitted through the network in a forward direction, layer by layer. This type of neural networks, which represents a generalization of the single-layer perceptron, is generally known as multilayer perceptron (MLP). When trained in a supervised manner using highly popular and computationally efficienterror back-propagation algorithm, multilayer perceptrons can successfully solve complex and different problems, but certainly do not provide an optimal solution for all solvable problems. Essentially, error back-propagation learning consists of a forward pass and a backward pass. In the forward pass, the effect of an input vector, when being applied to the sensory nodes, propagates through the network. At the end, a set of outputs, as the real response of the network, is formed. The synaptic weights are all fixed during this stage. However, these synaptic weights are being tuned according to error-correction rule during the backward pass. Namely, an error signal is produced as the real response of the network is subtracted from a desired (target) response. This error signal is then propagated backward through the network during which the synaptic weights are tuned so that the difference between the real and the desired response of the network decreases.One or more layers of hidden neurons enhance network's learning of difficult problems by extracting more significant features from the input vectors(Haykin 1999).

2.5 Radial Basis Function Network

The design of a neural network can also be perceived as a curve-fitting (approximation) problem in a high-dimensional space,where learning is viewed as finding a surface which represents a best fit to the training data in a multidimensional space. This multidimensional surface is then used to interpolate the test data. The method of radial-basis functions is motivated by such a viewpoint. The early work on radial-basis functions is reviewed in Powell (1985).A radial-basis function (RBF) network basically consists of three layers having completely different tasks. The input layer connects the network to the environment via source nodes that serve as sensory units. A nonlinear transformation from the input space to the hidden space of high dimensionality is applied in the second layer as the only hidden layer in the network. The output layer, producing the response of the network to the input vector, is linear. The effect of applying nonlinear transformation prior to a linear transformation is explained by Cover (1965). As stated by him, there is a higher change of a pattern recognition problem to be linearly separable in a high-dimensional space. Therefore, the dimension of the hidden space in an RBF network is often made high. Moreover, the higher the dimension of the hidden space, the more accurate the approximation of smooth mapping is(Mhaskar, 1996; Niyogi and Girosi, 1996).

3. Experimental results and discussion

3.1 Experiment

Classification of EEG signals consists of data acquisition and preparation, signal processing, feature extraction and classification. We propose a method based on MSPCA for denoising, DWT for feature extraction and ANNs for classification.We extracted 4096-samples-long segments from both datasets of each patient. Approximately two segments per hour were extracted from “inter-ictal” dataset, producing 1050 inter-ictal segments. We also extracted two types of segments from “ictal” dataset: ictal and pre-ictal. We used minimum number of 4096-samples-long segments to cover all 87 seizure activities, producing 652 ictal segments. We extracted five segments within a seizure-free "pre-ictal" period of 50-60 minutes, producing 435 pre-ictal segments. Only one out of six channels was used for extraction of EEG segments, although results from the different authors presented a poor performance of univariate measures (Mormann et al., 2005).

We selected the number of decomposition levels for DWT to be 5 since EEG signals contain no useful frequency components above 30 Hz, and because of 256Hz sampling rate of Neurofile NT used to acquire the EEG data. Daubechies 4 (DB4) wavelet filter was used to reconstruct the detail and approximation records.All 2137 EEG segments, which belong to three different classes, were divided into sub-band frequencies A5 (0-4 Hz), D5 (4-8 Hz), D4 (8-16 Hz), D3 (16-32 Hz), D2 (32-64 Hz) and D1 (64-128 Hz). Sub-band frequencies A5 and D3-D5 almost perfectly correspond to δ (0-4 Hz), θ (4-8 Hz), α (8-12 Hz) and β (12-26 Hz) frequencies of EEG signals (Bylsma et al., 1994).

A set of fifteen statistical features was then extracted from the wavelet coefficients representing these sub-band frequencies and fed as inputs to classifiers. A Multiscale PCA (MSPCA) de-noising method was also applied to determine if it can further enhance the classifiers' performance. We implemented a classification system based on MLP and RBF network using wavelet statistical features as inputs and 10-fold cross validation method, to guarantee validity of the results.

3.2 Results

We performed two types of experiment: with and without MSPCA de-noising method applied.In Table 1, we have seen that MSPCA drastically improved the classification accuracy of both classifiers, while MLP network achieved higher total classification accuracy than RBF network. The accuracies for each class are also presented in Table 1.

Classifier	Accuracy (Pre-ictal)	Accuracy (Inter-ictal)	Accuracy (Ictal)	Total Accuracy
MLP +DWT	2.76 %	89.43 %	60.58 %	62.99 %
RBFN +DWT	7.13 %	90.57 %	54.45 %	62.56 %
MLP + MSPCA+DWT	87.82 %	97.43 %	96.17 %	95.09 %
RBFN + MSPCA+DWT	71.49 %	97.14 %	94.02 %	90.97 %

Table 1. Accuracies of MLP and RBF network classifiers with and without MSPCA.

MSPCA significantly improved the classification accuracy for ictal and pre-ictal samples, while accuracy performance for inter-ictal class was only slightly improved. Classifiers are totally useless for seizure prediction if MSPCA is not applied.

3.3 Discussion

The experiment results show that MSPCA is an effective denoising method for improving the classification performance. Without MSPCA, our method classified many pre-ictal/ictal data samples as being inter-ictal.Aminghafari, Cheze, & Poggi, (2006) showed that de-noised signals by MSPCA magnify the spikes more clearly. Therefore, MSPCA enhanced our classifier's performance for about 50%.

Our approach outperformed the one explained in D'Alessandro et al., (2003). Theyalso used data of only four patients to developfour different classifiers for each patient. Although Costa et al., (2008)introduced one more class (pos-ictal) and achieved accuracies of 99%, using data of only two patients from Freiburg database is insufficient to successfully train and develop a model. Mirowski et al., (2009) predicted all seizures without false positives for 15 patients, without mentioning how classifier performed on data belonging to six remaining patients from Freiburg database. Thus, sensitivity of 71% is reported, which is lower than classification accuracies for pre-ictal class of both of our classifiers. Chisci et al., (2010)used nine patients for which additional electro-corticographic recordings (grid–strip electrodes) were available and achieved 100% sensitivity with low false positive rates. However, they developed patient-specific system by training nine classifiers, where each classifier used train and test data of only one patient. Our proposed system is more general because only one classifier is developed for all patients and it is not bound to specific group of epileptic patients.

4. CONCLUSION

We showed that a high classification accuracy of EEG signals can be accomplished in cases when additional "pre-ictal" class is introduced. Many research papers showed that DWT coefficients well represent the EEG signals and ensure a good differentiation between classes. However, we managed to achieve high accuracies only when MSPCA de-noising method was applied to Freiburg dataset. The accuracy may be further improved by applying dimension reduction or feature selection methods like ICA or LDA on the feature vectors. Measures that characterize the relations between two or more channels can be used to further enhance the performance. Using only inter-ictal and pre-ictal samples to train the classifier could be investigated since our aim is not seizure detection. Freiburg dataset can serve as a challenge for trying other feature extraction methods rather than DWT. The proposed approach may become an efficient tool to predict epileptic seizures from EEG recordings.

REFERENCES

Adeli, H., Zhou, Z., & Dadmehr, N. (2003). Analysis of EEG records in an epileptic patient using wavelet transform. Journal of Neuroscience Methods 123, 69-87.

Aminghafari, M., Cheze, N., & Poggi, J.-M. (2006). Multivariate denoising using wavelets and principal component analysis. Computational Statistics & Data Analysis, 50, 2381-2398.

Bakshi, B. R. (1998). Multiscale PCA with Application to Multivariate Statistical Process Monitoring. AlChE, 44(7), 1596-1610.

Bigan, C. (1998). A recursive time-frequency processing method for neural networks recognition of EEG seizures. In E. C. Ifeachor, A. Sperduti, & A. Starita (Eds.), Neural Networks and Expert Systems in Medicine and Healthcare. Singapore: World Scientific.

Buck, D., Baker, G. A., Jacoby, A., Smith, D. F., & Chadwick, D. W. (1997). Patients' experiences of injury as a result of epilepsy. Epilepsia, 38, 439-44.

Bylsma, F., Peyser, C., Folstein, S., Ross, C., & Brandt, J. (1994). EEG power spectra in Huntington's disease: clinical and neuropsychological correlates. Neuropsychologia 32(2), 137-150.

Chisci, L., Mavino, A., Perferi, G., Sciandrone, M., Anile, C., Colicchio, G., et al. (2010). Real-Time Epileptic Seizure Prediction Using AR Models and Support Vector Machines. IEEE Transactions on Biomedical Engineering, 57(5).

Costa, R. P., Oliveira, P., Rodrigues, G., Leitao, B., & Dourado, A. (2008). Epileptic Seizure Classification Using Neural Networks with 14 Features. In I. Lovrek, R. J. Howlett, & L. C. Jain (Eds.), KES 2008, Part II, LNAI 5178 (pp. 281-288). Springer-Verlag Berlin Heidelberg.

Cover, T. M. (1965). Geometrical and Statistical properties of systems of linear inequalities with applications in pattern recognition. IEEE Transactions on Electronic Computers, 326-334.

D'Alessandro, M., Esteller, R., Vachtsevanos, G., Hinson, A., Echauz, J., & Litt, B. (2003). Epileptic seizure prediction using hybrid feature selection over multiple intracranial EEG electrode contacts: a report of four patients. IEEE Transactions on Biomedical Engineering 50 (5), 603-615.

Elger, C. E. (2001). Future trends in epileptology. Current Opinion in Neurology, 14, 185-186.

Ganesan, R., Das, T. K., & Venkataraman, V. (2004). Wavelet-based multiscale statistical process monitoring: A literature review. IIE Transactions, 36(9), 787-806.

Haykin, S. (1999). Neural Networks: A Comprehensive Foundation (Second ed.). Prentice Hall.

Kandaswamy, A., Kumar, C., Ramanathan, R., Jayaraman, S., & Malmurugan, N. (2004). Neural classification of lung sounds using wavelet coefficients. Computers in Biology and Medicine, 34(6), 523-537.

Maiwald, T., Winterhalder, M., Aschenbrenner-Scheibe, R., Voss, H. U., Schulze-Bonhage, A., & Timmer, J. (2004). Comparison of three nonlinear seizure prediction methods by means of the seizure prediction characteristic. Physica D 194, 357-368.

Marchant, B. P. (2003). Time–frequency analysis for biosystem engineering. Biosystems Engineering, 85(3), 261-281.

Mhaskar, H. N. (1996). Neural networks for optimal approximation of smooth and analytic functions. Neural Computation, 8, 164-177.

Mirowski, P., Madhavan, D., LeCun, Y., & Kuzniecky, R. (2009). Classification of patterns of EEG synchronization for seizure prediction. Clinical Neurophysiology, 120(11), 1927-1940.

Mormann, F., Andrzejak, R. G., Elger, C. E., & Lehnertz, K. (2007). Seizure prediction: the long and winding road. Brain, 130, 314-333.

Murray, J. (1993). Coping with the uncertainty of uncontrolled epilepsy. 2, 167-178.

Niyogi, P., & Girosi, F. (1996). On the relationship between generalization error, hypothesis complexity, and sample complexity for Radial Basis Functions. Neural Computation, 8, 819-842.

Powell, M. J. (1985). Radial basis functions for multivariable interpolation: a review. IMA Conference on Algorithms for the Approximation of Functions and Data. Shrivenham: RMCS.

Schachter, S. C. (1994). The brainstorms companion: epilepsy in our view. New York: Raven Press.

Semmlow, J. L. (2004). Biosignal and biomedical image processing: MATLAB-based applications. New York: Marcel Dekker, Inc.

Stein, A. G., Eder, H. G., Blum, D. E., Drachev, A., & Fischer, R. S. (2000). An automated drug delivery system for focal epilepsy. Epilepsy Res, 39, 103-114.

Subasi, A. (2007). EEG signal classification using wavelet feature extraction and a mixture of expert model. Expert Systems with Applications 32, 1084-1093.

Subasi, A., & Gürsoy, M. I. (2010). Comparison of PCA, ICA and LDA in EEG signal classification using DWT and SVM. Expert Systems with Applications 37, 8659-8666.