Classification of Alzheimer's Electroencephalograms using Artificial Neural Networks and Logistic Regression

Pedro Rodrigues[*1)], João Teixeira [1)], Roberto Hornero[2)], Jesús Poza[2)] and Alicia Carreres[3)]

1) Polytechnic Institute of Bragança, Portugal.
2) Biomedical Engineering Group, E.T.S. Ingenieros de Telecomunicación, University of Valladolid, Spain.
3) Hospital Universitario "Pío del Rio Hortega",Valladolid, Spain.
E-mail: p_mlrodrigues@hotmail.com

Abstract

The Artificial Neural Networks have been used over the years to solve complex problems and their development has strongly grown in recent years. In particular, this work, focused on the development and a comparison between Artificial Neural Networks (ANN) and a traditional statistical technic known as Logistic Regression (LR) in Electroencephalogram (EEG) classification. The Wavelet Transform was seen as the main technique of signal processing, in order to analyze the EEG signals of this study. Some features were extracted by the EEG signals like relative power (RP) in conventional frequency bands and two spectral ratios. The best feature combination was selected by Principal Components Analysis method to increase the accuracy of the ANN and LR to discriminate their entries between Alzheimer Disease and Controls.

Keywords: EEG, Classification, Wavelet Transform, Artificial Neural Networks, Logistic Regression;

1. Introduction

The increase in life expectancy associated with lower rates of birth that have been evident especially in recent years in developed countries, led to an aging population and to an increase in the incidence of diseases related old age, such as Alzheimer's disease (AD). AD is a progressive brain degeneration that initially affects memory for recent events, advancing to the overall deterioration of mental faculties. AD is associated with an increase of power in low frequencies (delta and theta band) and a decrease of power in higher frequencies (alpha and beta) [1].

The EEG is a noninvasive technique that records the electromagnetic fields produced by brain activity with good temporal resolution [1].

The main objective of this study is doing an Artificial Neural Network (ANN) and using a Logistic Regression (LR) for the identification of AD, and making a comparison between the both methods.

2. Methodology

This work was conducted mainly in four phases: getting the EEG signals, processing of EEG signals (using Wavelet Transform as the main processing), selecting the best features combination for the input of the ANN and LR by Principal Components Analysis (PCA) and finally the classification of the signals by ANN and LR.

The EEG signals were obtained in Hospital Universitario "Pío del Rio Hortega",Valladolid, Spain. They provided 20 EEG's from patients with possibility of AD and 14 EEG's from a group of control. EEG signals were recorded at the sampling frequency of 200Hz. Each EEG consists of 19 signals recorded by the 19 electrodes. EEGs were organized in 5 seconds artifact-free epochs (1000 points). All the recordings were digitally filtered with a band-pass filter with cut-off frequencies at 1 Hz and at 40 Hz.

In order to extract some distinct information between AD and Control subjects in EEG signals we decomposed the signal segments of 5 seconds by WT at the decomposition 5. The Wavelet Biorthogonal 3.5 was used for that. We calculated the percentage of energy (*PE*) corresponding to the Detail coefficients in level 2, 3, 4, 5 and at the Approximation coefficient in level 5. These levels of decomposition corresponded to gamma (25-50Hz), beta (12,5-25Hz), alpha (6,25-12,5Hz), theta (3,125-6,25Hz) and delta (0-3,125Hz) waves respectively. We also computed two spectral ratios to summarize the deceleration of the EEG spectrum of AD patients [2].

We averaged the features extracted by the segments of 5 seconds per channel and at the last because we had a few data for classification, we assumed that each channel corresponds to one subject.

After that we selected the best feature combination by PCA technique. PCA summarized the information and detect correlations among our variables [3]. The method generates a new set of variables, called principal components. Each principal component is a linear combination of the original variables. The values of each principal component can be analyzed using statistical techniques like analysis of variance and regression analysis, among others, in order to remove the similar components that can lead to errors during the ANN and LR classification [3]. The best feature combination was done by r_1 and r_2 (Variance analysis of the all principal components obtained by PCA), given by eq. 1 and eq. 2, and so we had an ANN with 2 input nodes (Table 1). We divided the 646 cases of this

study in Training set and Test set. The Training set was constituted by 494 cases and the Test set by 152 cases (Table 1). Although we assumed that an electrode of a real subject corresponds to an individual subject, the electrodes belonging to a real subject were forced to belong to the same set.

$$r_1 = \frac{PE\ at\ alpha}{PE\ at\ theta} \quad (1)$$

$$r_2 = \frac{PE\ at\ alpha + PE\ at\ beta + PE\ at\ gamma}{PE\ at\ delta + PE\ at\ theta} \quad (2)$$

Table 1. Join training and test sets for the entry of the ANN and LR

Dimension		***Input nodes of ANN***
Training set	Test set	
2x494	2x152	2

The ANN was a pattern recognition ANN with a Logsig as activation function, SCG as the training algorithm and cross-entropy as the error function. To prevent the overfitting of ANN to the training set we calculated the optimum Weight Decay (WD) parameter. WD prevented the weights to participate fully in the modeling process of ANN to the training set [4]. In order to select the best WD and the number of Hidden Layer Nodes for a better classify of the test set, we used the graphic in Fig.1. There can be seen the distribution of AUC parameter, resulting by the ROC analysis of the ANN leave one out cross-validation process results for the training set, along the nodes with a different WD. This type of WD analysis per nodes allowed us to avoid the use of the validation set.

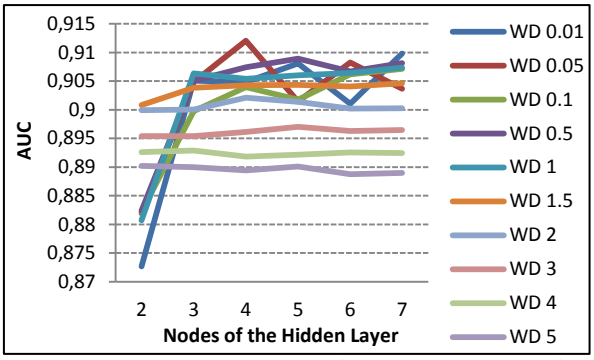

Fig. 1. Graphic that shows the values of AUC along the Nodes of the Hidden Layer for different kinds of WD values in the process of ANN cross-validation of the training data set.

After a short analysis of the graphic (Fig.1), we choose 0.05 of WD and 4 nods of the Hidden Layer, to classify the test set by ANN because this parameters provide more AUC for the training data set.

On the other hand the LR was utilized for prediction of the probability of an event occurrence by fitting data to a logistic curve [5]. The same training set and test set used in ANN methodology was employed in LR methodology (Table 1). Finally the results were evaluated by some parameters extracted by the ROC curve like: Area Under Curve (AUC) that summarized all the performance of the process, Sensibility represented the percentage of patients correctly classified, Specificity were the proportion of controls properly identified and Accuracy were the percentage of subjects correctly recognized [6].

3. Results

The methodologies ROC results for the classification of the test set can be observed in the following Table.

Table 2. Classification ROC results of test set by ANN and LR

Parameters	**Methods**	
	ANN	LR
AUC	0,9	0,8
Sensitivity	77,6%	77,6%
Specificity	90,8%	73,7%
Accuracy	84,2%	75,7%

4. Discussion and Conclusion

Observing the Table 2 we could see that the classification ROC results significantly increased in ANN when compared with the results of the LR. The sample size of the sets could be a good reason to explain the differences between the results of two classifications. LR had more efficiency with small dimensional sets [5]. The classification results of the ANN were optimums because we obtained 0.9 of AUC and 84.2% of Accuracy. The results of AUC were in line with the AUC obtained in the process of ANN cross-validation of the training data set. This means that the test set was a representative group of all the population involved in this study.

We could demonstrate that the ANN may be a promising tool for the detection of AD but some limitations for this type of study arise because we loss some spatial information when we retaining only the average measures over the channels and we assumed that one electrode were a subject because of the small database. Furthermore the detected increase of EEG regularity is not specific to AD [1].

References

[1]Jeong J. EEG dynamics in patients with Alzheimer's disease. Clin. Neurophysiol. 115, 1490-1505, 2004.

[2]Poza J, Hornero R, Abásolo D, Fernández A and Mayo A. Evaluation of spectral ratio measures from spontaneous MEG recordings in patients with Alzheimer's disease. Computer Methods And Programs In Biomedicine 90, 37-47, 2008.

[3]Jolliffe I.T. Principal Component Analysis,2nd ed. Springer, 2002

[4]Bishop M. Neural Networks for Pattern Recognition Christopher. Oxford University Press, 1995.

[5]Hosmer, D W. Stanley Lemeshow. Applied Logistic Regression, 2nd ed. New York; Wiley, 2000.

[6]Williams C, Lee S, Fisher R and Dickerman L. A comparison of statistical methods for prenatal screening for Down syndrome. Applied Stochastic Models in Business and Ind. 15, 89-101, 1999.