

# Non-linear Analysis of the Electroencephalogram in Alzheimer's Disease by Means of Symbolic Sequence Decomposition Method

Pinar Deniz Tosun<sup>\*1</sup>, Samantha Simons<sup>1</sup>, Raphaele Winsky-Sommerer<sup>2</sup>, Daniel Abasolo<sup>1</sup>

Accepted 15<sup>th</sup> August 2014

**Abstract:** In this pilot study, a symbolic sequence decomposition method was used in conjunction with Shannon's entropy to investigate the changes in electroencephalogram signals of 11 patients with Alzheimer's disease and 11 age-matched control subjects. Results were statistically analysed by student t-test and later classified with receiver operating curves. Statistically significant differences between both groups were found at electrodes Fp1, O2, P3, T4 and T5. Sensitivity (defined as percentages of correctly classified patients) and specificity (defined as correctly classified controls) were evaluated using the receiver operating curves method. Accuracy of the methods was calculated according to sensitivity and specificity measures of electrodes showing statistically significant differences between the control group and Alzheimer's disease patients and ranged between 72.73-77.27%. These accuracy values were in agreement with previously published entropy studies on this data set. Although combining these methods did not provide any greater accuracy over previous findings, using a symbolic sequence decomposition method enhanced the data processing.

**Keywords:** Alzheimer's Disease, Electroencephalogram, Non-linear Analysis, Symbolic Sequence Decomposition, Shannon's Entropy.

## 1. Introduction

Alzheimer's disease (AD) is the most frequent cause of dementia in the western world and is caused by excessive amyloid deposition and accumulation of abnormal tau protein in the brain which affect the cognitive ability of the sufferer [1].

The clinical diagnosis of AD is made primarily on the basis of medical history, psychiatric evaluation and different memory and mental health tests [2]. However, an indisputable diagnosis is only possible post-mortem [3]. Symptoms of the disease vary from patient to patient along with the severity of the disease. Early diagnosis is crucial in terms of lessening the effects of the disease with available drug treatments or making necessary life change adjustments for the patient and care takers to ensure optimum quality of life [4].

The electroencephalogram (EEG) has been used as a tool in the investigation of dementia for several decades. EEG signals reflect brain electrical activity and can be non-invasively recorded with surface electrodes. AD is a cortical dementia and as such electrical abnormalities in the brain signals caused by AD can be captured with cranial surface electrodes [5]. Generally, EEG signals of AD patients show a shift to lower frequencies in spectral analysis which suggests a decreased cohesion of cognitive networks [6]. Moreover, AD patients' EEGs display less complexity and contain more regular patterns compared to

Due to the intrinsic irregular and aperiodic nature of the EEG signal, spectral analyses techniques might not be sufficient to characterise the dynamics of the events underlying the EEG signals. Thus, additional techniques, such as non-linear time series analysis techniques were applied to provide a better understanding of EEG signals [9].

Correlation dimension (D2) is used to identify the complexity of a system [10]. AD patients' D2 measures show decreased values indicating less complex dynamics of neural networks in the brain, possibly due to the loss of neurons and synapses [11]. The Lyapunov exponents (L1) have been used to characterise non-linear behaviour of a system and can be seen as a measure of unpredictability. AD patients' EEGs have lower L1 values describing a more reliable or regular signal, which could suggest information processing, is less flexible in diseased brain [12].

Nevertheless, the amount of data required for meaningful results with these methods is very high. Also, a detailed signal conditioning is necessary in order to apply algorithms for both analysis methods [9]. Therefore, in this pilot study a new method has been applied on AD EEGs which is combining two non-linear analysis methods, i.e., symbolic sequence decomposition and Shannon's entropy measures.

Symbolic dynamical analysis is a family of non-linear signal processing techniques which investigate a signal in small, discrete time dynamics which relate to portions of the original signal. Symbolic sequence decomposition takes a finite number of samples and reforms a symbol series out of the original sample series depending on the value of the original sample relative to a threshold value of the whole sample series [13]. The overall process ensures an approximate analysis of a complex biological system [14, 15]. Entropy studies on the other hand, analyse the randomness or predictability of systems. First used as a thermodynamics term, in biomedical engineering, Shannon's entropy identifies the amount of information within the biological signal where greater entropy indicates more information than lower entropy within a signal. This feature can be interpreted as

<sup>1</sup> Centre for Biomedical Engineering, Department of Mechanical Engineering Sciences, Faculty of Engineering and Physical Sciences, University of Surrey, Guildford, GU2 7XH, United Kingdom

<sup>2</sup> Surrey Sleep Research Centre, Faculty of Health and Medical Sciences, University of Surrey, Guildford, United Kingdom

\* Corresponding Author: Email: p.tosun@surrey.ac.uk

# This paper has been presented at the International Conference on Advanced Technology and Sciences (ICAT'14) held in Antalya (Turkey), August 12-15, 2014.

those of control subjects [5, 7, 8].

regularity when the entropy is low and as complexity in parallel with a higher entropy value [16].

The current paper is organised as follows. Section 2 describes selection of AD patients and controls, as well as the process for data collection. The symbolic sequence decomposition and statistical analyses are also defined. The results are presented in Section 3. Finally, Section 4 contains a discussion of the results and conclusions.

## 2. Research Method

### 2.1. EEG Signal Database

The database consisted of 22 subjects, 11 AD patients (5 men and 6 women,  $72.5 \pm 8.3$  years (mean  $\pm$  standard deviation (SD)) and 11 age-matched controls (7 men and 4 women  $72.8 \pm 6.1$  years mean  $\pm$  SD). All patients were diagnosed with AD after a detailed medical test and a Mini-Mental State Examination (MMSE) which evaluates cognitive impairment [4, 17]. The MMSE scores, which identifies cognitive impairment, were  $13.1 \pm 5.9$  and  $30.0 \pm 0$  (mean  $\pm$  SD) respectively. The data collection and evaluation of the signals received ethical approval along with the permission of care takers of patients.

Signals were recorded with a sampling frequency of 256 Hz in a resting but awake state with eyes closed using the international 10-20 electrode placement system (electrodes Fp1, Fp2, F3, F4, C3, C4, P3, P4, O1, O2, F7, F8, T3, T4, T5, T6, Fz, Cz and Pz).

Recordings were carried over five minutes in order to reduce artefacts on the EEG recordings. Five minute recordings were sufficient to collect EEGs without occurrence of sleep. Signals were then checked by a specialist to eliminate signals contaminated with muscle movement and cardiac signal artefacts. For each subject, artefact-free 5-sec epochs (1280 data points) were selected and contributed to further analyses. An average of  $30.0 \pm 12.5$  (mean  $\pm$  SD) epochs was selected for each electrode of each subject.

### 2.2. Symbolic Sequence Decomposition

Symbolic sequence decomposition is a method which can be used to convert a raw series into a symbolic series. This approach provides the advantage to create more numerical computations since these discrete symbolic time series are represented in binary codes [15]. Each region is associated with a unique symbol and these symbols are involved in creating another series of data, called the symbol series, out of the original series depending on the region which the original value falls into [14].

First, the threshold value for partitioning is calculated. In the median technique, the threshold is the statistically calculated median value. By using this median value, data series were separated into two parts, i.e., partitioned into two different regions. Time series' value was defined in binary codes, "0" being in region 1, values which are less than the threshold value and "1" being in region 2 for values higher than the threshold. After the symbolisation was performed, a two-step template window was slid over the symbols in order to create code series which allows further extraction of additional information. Template size and window sliding steps are two main components to consider when extracting information from the data series. To take into account the general trend of the EEG signal, slide step and window size can be selected equally to avoid overlapping [14]. On the other hand, local changes are captured in small step sizes. However, small sliding steps affect reliability and smoothness of the probability density function

(PDF) of the time series [16]. In the current study, a two-step template with a single step windowing was performed on the time series to obtain code series in order to identify local events without affecting the PDF of data series.

### 2.3. Shannon's Entropy

Shannon's entropy (SE) was first introduced by Claude Shannon [18] as a communication technology application. It quantifies the amount of information that is carried within a signal, i.e., the significance of the signal is extracted from a group of data with relevance to its statistical mechanics [19].

Let us assume  $X$  being an arbitrary variable taking values on a set  $(a_1, a_2, \dots, a_m)$  with probability;

$$P(X = a_i) = p_i, i = 1, \dots, m \quad (1)$$

Then, SE of  $X$  can be defined as follows;

$$H(X) \equiv H(p_1, \dots, p_m) = -\sum_{i=1}^m p_i \log p_i \quad (2)$$

$H(X)$  is the entropy value of a random variable  $X$  or the entropy of the probability distribution  $(p_1, \dots, p_m)$ . The sum of all these probability distributions is equal to 1.

$$H_s = -\frac{1}{(\log N_{obs})} \sum_{i=1}^m p_i \log p_i \quad (3)$$

$H_s$  stands for the normalised Shannon's entropy value and is ranged between 0-1 where values which are approximately 1 represent an irregularity whereas 0 represent regularities within the data set [20]. When the randomness of the spectral distribution is relatively flat, lower entropy values were evaluated.

$N_{obs}$  (equation 3) is the number of possible sequences in the code series. Because a two-step window was used when creating code-series, this number is equal to 4 in our calculations.  $p_i$  values are thus  $p_1, p_2, p_3, p_4$ . There are four possible sequences for a two size window template where,  $p_1$  is the probability of '00' sequence within the code-series;  $p_2$  is the probability of '01';  $p_3$  is the probability of '10' and  $p_4$  is the probability of '11'.

The total number of samples is determined by the length of the epoch and sampling rate of the data set (i.e., 1280 bits for each subject and electrode).

### 2.4. Receiver Operating Curves

The receiver operating curves (ROC) gives a graphical representation of the accuracy of a statistical test. In this study, ROC was used to validate student's t-test analysis performed on our two sets of data (AD patients vs. age-matched controls). Sensitivity and specificity of a given data set is sketched to show number of true positives and true negatives. These numbers are usually expressed by percentages and highly depend on the Type I and Type II errors [21]. Thus, percentages of correctly identified patients, called as sensitivity, and the percentage of correctly identified controls, also known as specificity, are two important components of this method. Incorrect identification of patients and controls lead to Type I and II errors mentioned earlier.

$$fp\ rate = FP/N \quad (4)$$

$$tp\ rate = TP/P \quad (5)$$

True positives are abbreviated as  $TP$  and false positives as  $FP$ . Values,  $FP$  are equal to incorrectly classified negatives and  $N$  is

number of total negatives, which includes *FP* and correctly classified controls. *TP* is correctly classified positives and *P* is the total number of positives, which includes *TP* and incorrectly identified controls. Both *N* and *P* values are equal to 11 in this study with 11 AD patients and 11 age-matched controls. Sensitivity is calculated as the percentage of true positives among true positives and false negatives and equals to *fp rate* (equation 4). Specificity is evaluated as the percentage of true negatives among true negatives and false positives and equals to *tp rate* (equation 5). Once these percentages are calculated, accuracy can be established as the ratio of the number of correctly identified subjects (both AD and controls) and the total number of subjects, which is 22 (11 AD patients and 11 age-matched controls).

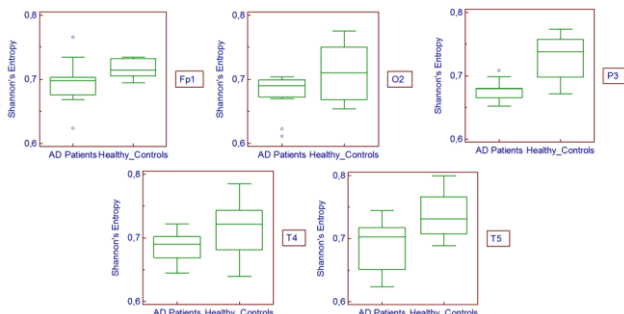
### 3. Results

SE values of each electrode are listed in Table 1. Statistically significant electrodes ( $p < 0.01$ ) marked with asterisk.

**Table 1.** Shannon's Entropy values for AD patients and controls for each electrode. Statistically significant electrodes are marked with an asterisk.

Electrode	AD Patients (mean±SD)	Controls (mean±SD)	p-value
C3	0.6916±0.0342	0.7137±0.0256	0.0962
C4	0.7007±0.0311	0.7140±0.0283	0.0716
F3	0.6749±0.0315	0.6907±0.0327	0.0717
F4	0.6884±0.0314	0.6878±0.0298	0.4063
F7	0.6817±0.0395	0.6874±0.0311	0.0103
F8	0.6732±0.0319	0.6977±0.0251	0.0299
Fp1	0.6956±0.0244	0.7165±0.0236	0.0092*
Fp2	0.7252±0.0322	0.7062±0.0259	0.2359
O1	0.6803±0.0309	0.6972±0.0275	0.2304
O2	0.6781±0.0334	0.7120±0.0294	7E-05*
P3	0.6769±0.0259	0.7289±0.0188	5E-05*
P4	0.7062±0.0256	0.7378±0.0207	0.0245
T3	0.6921±0.0254	0.7365±0.0231	0.2173
T4	0.6872±0.0304	0.7161±0.0311	0.0087*
T5	0.6912±0.0256	0.7358±0.0242	0.0035*
T6	0.7198±0.0278	0.7415±0.0305	0.1594

Box plots of the mean SE values for statistically significantly electrodes can be seen in Figure 1. Sensitivity values (true positives, correctly identified AD patients) against specificity (true negatives, correctly identified healthy controls) were calculated using an online software programme called MedCalc.



**Figure 1.** Box plots for the mean SE values for statistically significant electrodes of Fp1, O2, P3, T4 and T5.

Accuracies of statistically significant electrodes were calculated according to these sensitivity and specificity values. Table 2 summarises the accuracy values calculated for statistically significant electrodes. The threshold value is the mean SE value selected for electrodes Fp1, O2, P3, T4 and T5 respectively.

Sensitivity and specificity percentages of this particular threshold value provide the accuracy of the electrode.

**Table 2.** Sensitivity, specificity, and accuracy for the electrodes showing statistically significant differences between AD Patients and controls.

Electrode	Sensitivity (%)	Specificity (%)	Threshold	Accuracy (%)
Fp1	100	50.0	0.6686	77.27
O2	100	50.0	0.6222	77.27
P3	85.7	75.0	0.6797	77.27
T4	71.4	75.0	0.6899	72.73
T5	85.7	75.0	0.6987	77.27

### 4. Discussions and Conclusions

In this study, symbolic sequence decomposition method was used together with Shannon's entropy in order to investigate EEG changes caused by Alzheimer's disease compared to age-matched controls.

O2, P3, T4 and T5 have been frequently reported as significant differences showing electrodes between AD patients and control subjects [8, 20]. In particular, P3 electrode was the only statistically significant electrode which is common for all entropy studies performed using this data set.

Except for electrode Fp1, all other electrodes which were statistically different, were located on the posterior side of the skull.

Sensitivity of Fp1 is observed as 100% which is higher than the only value for the same electrode of 90.91% sensitivity of a previous work by Abásolo [7, 8]. O2 accuracy was at 100% which is higher than any other entropy measure previously mentioned by Abásolo and collaborators [8]. P3 has a greater sensitivity together with the other two electrodes, compared to the previous database studies. On the other hand, specificity values of these five electrodes were lower than any other value evaluated before. Our results showed similar features to previous studies. Statistically significant electrodes were situated at the posterior part of the skull for Sample entropy, Spectral entropy, Approximate entropy and Lempel-Ziv complexity measures [6, 7, 8]. Statistically significant electrodes of SE were observed for the same brain regions.

To conclude, the symbolic sequence decomposition supports current knowledge about the effects of Alzheimer's disease on the EEG. While this method did not provide any improvement over other entropy techniques in terms of the possible use of it as a diagnostics tool, it proved to be relatively faster, hence can be used as a part of an analysis method to characterize EEG in other cerebral disorders. Also a future work on the statistically significant electrodes' locations can be conducted since it may be a beneficial indicator of a cerebral disorder.

### Acknowledgements

We would like to thank Pedro Espino (Hospital San Carlos, Madrid, Spain), for the EEG epoch selection. Also, we would like to thank Turkish Republic Ministry of Education for their Higher Education Bursary Scheme which provided us with the possibility to work together in this project.

### References

- [1] M. Graeber, S. Kösel, R. Egensperger, R. Banati, U. Müller, K. Bise, P. Hoff, H. Möller, K. Fujisawa and P. Mehraein. Rediscovery of the case described by alois

- alzheimer in 1911: Historical, histological and molecular genetic analysis. *Neurogenetics* 1(1), pp. 73-80. 1997.
- [2] G. M. McKhann, D. S. Knopman, H. Chertkow, B. T. Hyman, C. R. Jack, C. H. Kawas, W. E. Klunk, W. J. Koroshetz, J. J. Manly, R. Mayeux, R. C. Mohs, J. C. Morris, M. N. Rossor, P. Scheltens, M. C. Carrillo, B. Thies, S. Weintraub and C. H. Phelps. The diagnosis of dementia due to Alzheimer's disease: Recommendations from the national institute on aging-Alzheimer's association workgroups on diagnostic guidelines for alzheimer's disease *Alzheimer's & Dementia* 7(3), pp. 263-269. 2011.
- [3] D. S. Knopman, S. T. DeKosky, J. L. Cummings, H. Chui, J. Corey-Bloom, N. Relkin, G. W. Small, B. Miller and J. C. Stevens. Practice parameter: Diagnosis of dementia (an evidence-based review). report of the quality standards subcommittee of the american academy of neurology. *Neurology* 56(9), pp. 1143-1153. 2001.
- [4] B. Reisberg, R. Doody, A. Stöffler, F. Schmitt, S. Ferris and H. J. Möbius. Memantine in moderate-to-severe alzheimer's disease. *N. Engl. J. Med.* 348(14), pp. 1333-1341. 2003.
- [5] J. Jeong. EEG dynamics in patients with alzheimer's disease. *Clinical Neurophysiology* 115(7), pp. 1490-1505. 2004.
- [6] D. Abásolo, R. Hornero, P. Espino, D. Alvarez and J. Poza. Entropy analysis of the EEG background activity in alzheimer's disease patients. *Physiol. Meas.* 27(3), pp. 241-244. 2006.
- [7] D. Abásolo, R. Hornero, P. Espino, J. Poza, C. I. Sánchez and R. de la Rosa. Analysis of regularity in the EEG background activity of Alzheimer's disease patients with approximate entropy. *Clinical Neurophysiology* 116(8), pp. 1826-1834. 2005.
- [8] D. Abásolo, R. Hornero, C. Gómez, M. García and M. López. Analysis of EEG background activity in Alzheimer's disease patients with Lempel-Ziv complexity and central tendency measure. *Med. Eng. Phys.* 28(4), pp. 315-322. 2006.
- [9] C. J. Stam. Nonlinear dynamical analysis of EEG and MEG: Review of an emerging field. *Clinical Neurophysiology* 116(10), pp. 2266-2301. 2005.
- [10] B. Jelles, J. H. van Birgelen, J. P. J. Slaets, R. E. M. Hekster, E. J. Jonkman and C. J. Stam. Decrease of non-linear structure in the EEG of Alzheimer patients compared to healthy controls. *Clin Neurophysiology* (110), pp.1159-1167. 1999.
- [11] J. Jeong, D. J. Kim, J. H. Chae, S. Y. Kim, H. J. Ko and I. H Paik. Nonlinear analysis of the EEG of schizophrenics with optimal embedding dimension. *Med Eng Phys* (20), pp. 669-676. 1998.
- [12] K. Keller and H. Lauffer. Symbolic analysis of high-dimensional time series. *International Journal of Bifurcation and Chaos* 13(09), pp. 2657-2668. 2003.
- [13] A. Abbott. Sequence analysis: New methods for old ideas. *Annual Review of Sociology* 21(1), pp. 93-113. 1995.
- [14] C. S. Daw, C. E. A. Finney and E. R. Tracy. A review of symbolic analysis of experimental data. *Rev. Sci. Instr.* 74(2), pp. 915-930. 2003.
- [15] S. Tong and N. V. Thakor. Advances in quantitative electroencephalogram analysis methods. *Annual Review of Biomedical Engineering* 6(01), pp. 453-456. 2004.
- [16] P. E. Rapp. A guide to dynamical analysis. *Integrative Physiological and Behavioural Science* 29(3), pp. 311-317. 1994.
- [17] L. Kurlowicz and M. Wallace. The Mini Mental State Examination (MMSE). *The Hartford Institute of Geriatric Nursing Issue* 3. 1999.
- [18] C. E. Shannon. A mathematical theory of communication. *The Bell Technical Journal* 27, pp. 379-423. 1948.
- [19] J. Bruhn, L. E. Röpcke, H. Bouillon and A. Hoeff. Shannon entropy applied to the measurement of the electroencephalographic effects of desflurane. *Anesthesiology* 95(1), pp. 30-35. 2001.
- [20] J. W. Sleigh, D. A. Steyn-Ross, M. L. Steyn-Ross, C. Grant and G. Ludbook. Cortical entropy changes with general anaesthesia: theory and experiment. *Physiological Measurement* 25(4), pp. 921-927. 2004
- [21] T. Fawcett. An introduction to ROC analysis. *Pattern Recognition Letters* 27, pp. 861-874. 2006.