EEG ASSESSMENT OF ALZHEIMER'S DISEASE USING UNIVERSAL COMPRESSION ALGORITHM

P. Zhao*, **E.C. Ifeachor†**

*PhD student, peng.zhao@plymouth.ac.uk, †Professor, e.ifeachor@plymouth.ac.uk
†Smeaton207, SPMC Group, University of Plymouth, Drake Circus, PL4 8AA,
Tel: 0044-(0)1752-232574, Fax: 0044-(0)1752-232583

Keywords: EEG, dementia, Alzheimer's diseases, LZW algorithm, universal compression algorithm.

Abstract

The prevalence of Alzheimer's disease (AD) and other types of dementia is rapidly rising with the aging population. Early diagnosis is important for effective treatment and care for AD, e.g. to slow its progression and to extend the quality of life for sufferers. The Electroencephalogram (EEG), potentially, could be used as a first line of decision support tool adjunct to current clinical assessments to improve diagnosis as the dementing process could be reflected in the changes in the EEG signals at the early stages. The EEG is attractive because it is non-invasive, widely available, relatively low-cost, and could be quickly carried out by non-specialist clinician for regular check-ups within the fast growing number of people at-risk. The work reported in this paper is part of a programme of study to investigate a new class of nonlinear methods based on information theory, which potentially offers a powerful way of objective assessment of AD. The assumption is that changes in the information content of the EEG may also reflect the pathological changes in the brain due to AD. In this paper, we seek to exploit knowledge from universal data compression method to quantify changes in the EEG. The basic hypothesis is that the information content of the EEGs for AD and normal groups, and hence the extent to which the EEGs can be compressed, are significantly different. Data compression has not been investigated as a means of characterising EEG signals in dementia research before and is very attractive because it does not require a priori knowledge of the source model. In our study, we focus on the LZW algorithm because of its sound theoretical foundation. We used it to compute the compression ratios for two EEG datasets and the results suggest that this approach may be used in objective assessment of Alzheimer's disease.

1 Introduction

The prevalence of Alzheimer's disease (AD) and other types of dementia is rapidly rising with the aging population. Early diagnosis is important for effective treatment and care for AD, e.g. to slow its progression and to extend the quality of life for sufferers. However, there may be a delay of up to 5 years between actual onset and the patient meeting the current clinical diagnostic criteria for AD. Thus, by the time a diagnosis is made, significant amount of irreversible cell damage might have already occurred. Although a definite diagnosis can only be made at post-mortem, the combined use of clinical assessment together with neuropsychological tests and objective methods (e.g. electrophysiological, neuroimaging and genomic diagnostic tests) can provide a reasonably accurate and early diagnosis.

The Electroencephalogram (EEG), potentially, could be used as a first line of decision support tool adjunct to current clinical assessments to improve the diagnosis. The EEG is attractive because it is non-invasive, widely available, relatively low-cost, and could be quickly carried out by a non-specialist clinician for regular check-ups within the fast growing number of people at-risk.

Based on the assumption that the dementing process could be reflected on the changes in the EEG signals at very early stages, the associated abnormities have been investigated using various linear and nonlinear methods [1,2]. Recent studies suggest that nonlinear measures are better suited to EEG analysis than conventional approaches, but with several limitations [1,2].

The work reported in this paper is part of a programme of study to investigate a new class of nonlinear methods based on information theory, which potentially offers a powerful way of objective assessment of AD. The approach is based on the assumption that the information content in the electrical activity recorded by the EEG is related to the information processing in the brain, and that a natural link may exist between the underlying ideas of information theoretic methods, the physiology of AD and its impact on brain functions [3]. Thus, changes in the features of the EEG due to AD may be related to changes in the information content of the EEG, which may also reflect the pathological changes in the brain due to AD.

Recent studies based on information theoretic methods (e.g. sample entropy) attempt to tackle the problem from nonlinear dynamical analysis perceptive [2]. In this paper, we seek to exploit knowledge from universal data compression to quantify the EEG for the assessment. The basic hypothesis is that the information content of the EEGs for AD and normal groups, and hence the extent to which the EEGs can be compressed, are significantly different. Data compression has not been investigated as a means of characterising EEG signals in dementia research before and is very attractive because it does not require a priori knowledge of the source model. In our study, we focus on the LZW algorithm because of its sound theoretical foundation.

The specific objectives of this paper are: (1) to study changes in the EEGs of AD and normal subjects using the universal lossless data compression algorithm, LZW (Lempel-Ziv-Welch); (2) to investigate if compression could be a useful tool for objective assessment of AD.

The rest of the paper is organized in four main parts: (1) a brief description of the universal compression algorithm, LZW, (2) details of the two datasets used in the study, (3) the results from the use of LZW algorithm, and (4) conclusion of the paper.

2 The Universal Compression Algorithm

A way to quantify the information content of the EEG is to obtain estimates of its entropy using, for example, Shannon's formula. In practice, this approach may not be appropriate because changes in the EEG activities for a variety of reasons mean that the source model is not fixed. Thus, in order to obtain good estimates of the information content, a method that works regardless of the underlying source model may be more appropriate. Potentially, universal compression algorithms satisfies such a requirement because of their ability to adapt to the source statistics and there is no need for a priori knowledge of the source model [4].

In the study, we focus on the LZW algorithm because of its sound theoretical basis and simplicity [4,5,6]. It belongs to the family of LZ algorithms, which are adaptive, dictionary-based and lossless. It exploits different types of redundancies apparent in the data [4,6] and is widely used to compress text and other types of data .

The LZW algorithm is based around a dictionary and code table. In its basic form, the LZW examines the data stream, byte by byte, using a parsing algorithm. From this, it creates a dictionary of sub-sequences that occur in the data and then assigns unique codes to them. As the process continues, the LZW algorithm identifies repeated sub-sequences in the data. Thus, the more times they are repeated the better compression will be achieved. At the end of the process, the code table in effect represents the data in compressed form. The extent of compression can be computed as the ratio of the file size before and after compression.

3 Subjects and EEG Recordings

In our experiment, we applied the LZW algorithm to two datasets: Dataset A and Dataset B, which were obtained from the Derriford Hospital and had been collected using normal hospital practices in conjunction with a strict protocol. The classification between normal and Alzheimer's disease was taken from hospital diagnosis notes.

3.1 Subjects

Dataset A includes 3 Alzheimer's patients and 8 age-matched controls (over 65 years old) all of which have normal EEGs confirmed by a consultant clinical neurophysiologist. Within the age-matched controls, one of the subjects subsequently developed AD and therefore, potentially, was in transition from normal to AD.

Dataset B includes 24 normal subjects and 17 probable AD, which are not perfectly age matched. In the normal group, the mean age is 69.4±11.5 years, the minimum is 40, and the maximum is 84; 42% are male; in the AD group, the mean age is 77.6±10.0 years, the minimum is 50, and the maximum is 93; 53% are male. These probable AD subjects were not previously diagnosed, and were still in the early stages of exhibiting symptoms; some of them were not referred for dementia diagnosis but came in for investigation of other disease, such as seizures.

3.2 EEG recordings

Dataset A was recorded using the traditional 10-20 system in a Common Reference Montage (using the average of all channels as the reference), and later converted to Common Average and Bipolar Montages in software.

Dataset B was recorded using the modified Maudsley system which is similar to the traditional 10-20 system.

In both datasets, the EEG recordings include various states: awake, hyperventilation, drowsy and alert with periods of eyes closed and open.

The sampling rate was reduced from 256Hz to 128Hz for analysis by averaging two consecutive samples for storage reasons. This was confirmed by analysis to have no significant effect on the data [1] which is to be expected because the frequency band of interest does not exceed 30Hz. A predetermined protocol was applied to avoid the possibility of inadvertently or unconsciously selecting only data segments that are suitable for analysis. Thus, whole recordings including artefacts were used without à priori selection of elements 'suitable for analysis'. This was to get an idea about the robustness and usefulness of the methods in practice [1]. Data from a fixed interval, 60s to 300s, was used to avoid electrical artefacts which commonly occur at the beginning of a record, therefore gives 4 minutes of data to analyse.

4 Results and Discussions

We employ Dataset A for development and Dataset B for evaluation of performance. Fig. 1 shows plots of the compression ratio of each of the 21 channels for each case in the Dataset A using LZW.

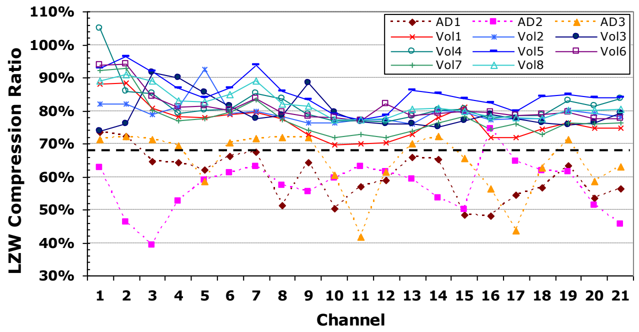

Fig. 1 LZW compression ratio for all cases and all channels (Dataset A).

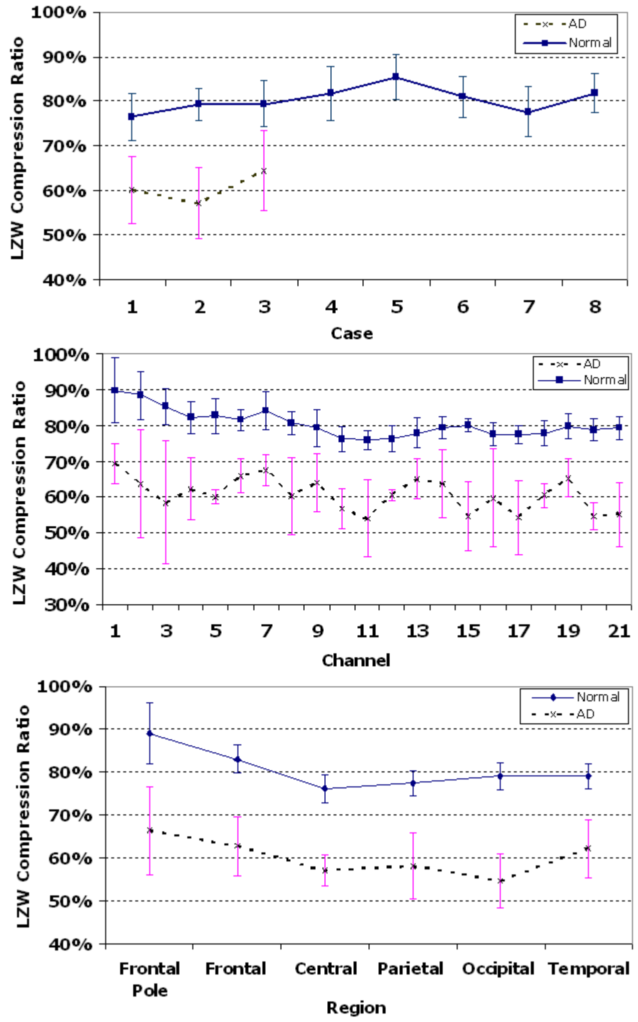

Fig. 2a/b/c The average and standard deviation in case/channel/ region (Dataset A).

As a screening tool, a very high specificity is desirable as this would reduce the possibility of many patients being sent for unnecessary and resource consuming follow-up tests. To achieve 100% specificity in Dataset A, we select a compression ratio of 69% as a threshold. In this case, compression ratios in the range 0~69% is indicative of AD and ratios of >69% is normal. For Dataset A, this gives a sensitivity of 77.78%.

Fig. 2a, b and c shows, respectively, the average compression ratios across channels for each case, the average compression ratios across cases for each channel, and the average compression ratios in key regions of the scalp across cases in AD and normal groups. These clearly show that the AD group has lower LZW compression ratios than the normal group, and that the two groups are well separated (according to the *p*-values listed in Table 1 and Table 2). With a 100% specificity, the sensitivity for all channels is 95.24% and for all regions it is 83.33%.

Region	Control subjects mean ± SD	AD patients mean ± SD	*p*-value
Frontal Pole	88.99% ± 0.0714	66.34% ± 0.1027	2.2597E-03
Temporal	79.09% ± 0.0286	62.16% ± 0.0673	1.6519E-04
Parietal	77.46% ± 0.0285	58.11% ± 0.0772	1.1664E-04
Frontal	83.07% ± 0.0328	62.73% ± 0.0694	7.2100E-05
Central	76.03% ± 0.0317	57.08% ± 0.0367	1.3363E-05
Occipital	78.99% ± 0.0314	54.70% ± 0.0620	9.2036E-06

Table 1: *p*-Values for regions using LZW (Dataset A).

Channel	Electrode	Control subjects (mean±SD)	AD patients (mean±SD)	*p*-value
1	FP1	69.16% ± 0.0559	89.65% ± 0.0909	5.8761E-03
16	P3	59.66% ± 0.1354	77.49% ± 0.0316	4.3481E-03
9	T3	63.87% ± 0.0816	79.18% ± 0.0511	4.1225E-03
2	FP2	63.52% ± 0.1503	88.33% ± 0.0677	3.3283E-03
13	T4	64.98% ± 0.0538	77.74% ± 0.0423	2.3773E-03
3	F7	58.38% ± 0.1704	85.26% ± 0.0515	1.9882E-03
14	A2	63.61% ± 0.0934	79.37% ± 0.0314	1.5429E-03
7	F8	67.33% ± 0.0427	83.98% ± 0.0540	1.0347E-03
4	F3	62.12% ± 0.0862	82.07% ± 0.0441	5.3588E-04
8	A1	60.13% ± 0.1058	80.43% ± 0.0314	5.1991E-04
19	T6	65.21% ± 0.0521	79.51% ± 0.0338	3.9685E-04
11	CZ	53.96% ± 0.1087	75.70% ± 0.0273	3.0564E-04
17	PZ	54.29% ± 0.1039	77.28% ± 0.0248	1.3651E-04
6	F4	65.88% ± 0.0466	81.35% ± 0.0290	8.1116E-05
12	C4	60.52% ± 0.0155	76.28% ± 0.0372	6.9310E-05
10	C3	56.77% ± 0.0562	76.12% ± 0.0347	5.9001E-05
21	O2	54.96% ± 0.0883	79.26% ± 0.0329	5.8370E-05
18	P4	60.37% ± 0.0331	77.62% ± 0.0339	3.4758E-05
5	FZ	59.95% ± 0.0197	82.71% ± 0.0485	3.0688E-05
15	T5	54.59% ± 0.0954	79.95% ± 0.0194	2.7445E-05
20	O1	54.43% ± 0.0375	78.73% ± 0.0310	1.5802E-06

Table 2: *p*-Values for channels using LZW (Dataset A).

Fig. 3 shows plots of the compression ratios for each of the 21 channels and for each case in Dataset B. Using the threshold of 69% selected for Dataset A, gives a specificity of 72.22% and a sensitivity of 35.29%.

Fig. 4a, b and c, respectively, show the average compression ratios across channels for each case, the average compression ratios for each channel across cases for the AD and normal groups, and the average compression ratios in key regions of the scalp across cases for the AD and normal groups. The last two figures clearly show that the AD group has a lower LZW compression ratio than the normal group, but according to the *p*-values in Table 3 and 4, only channels 2 and 7, the frontal regions give the best separation.

Using the threshold of 69% gives a specificity of 90.48% and sensitivity of 38.10% across all channels; across all regions, this gives a specificity of 77.08% and a sensitivity of 41.18%.

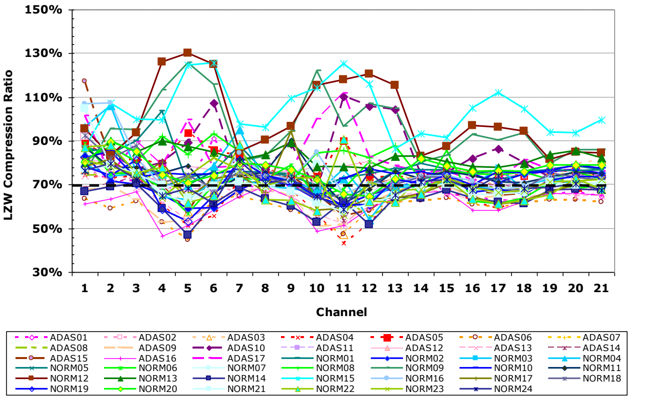

Fig. 3 LZW compression ratio for all cases and all channels (Dataset B).

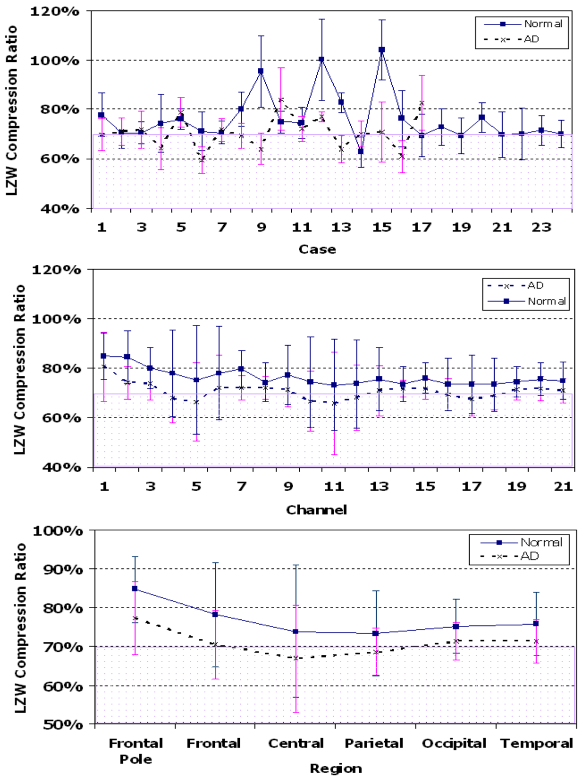

Fig. 4a/b/c The average and standard deviation in case/ channel/region (Dataset B).

Region	Control subjects Mean ± SD	AD patients Mean ± SD	*p* -value
Central	73.82% ± 0.1705	66.91% ± 0.1381	0.1753
Parietal	73.35% ± 0.1096	68.60% ± 0.0613	0.1147
Temporal	75.75% ± 0.0817	71.29% ± 0.0567	0.0598
Occipital	75.22% ± 0.0688	71.39% ± 0.0487	0.0562
Frontal	78.16% ± 0.1351	70.42% ± 0.0885	0.0455
Frontal Pole	84.74% ± 0.0848	77.38% ± 0.0950	0.0129

Table 3: *p*-Values for regions using LZW (Dataset B).

Channel	Electrode	Control subjects Mean ± SD	AD patients Mean ± SD	*p* -value
14	A2	71.72% ± 0.0341	73.55% ± 0.0706	0.3274
8	A1	72.07% ± 0.0446	74.27% ± 0.0776	0.3002
12	C4	68.17% ± 0.1318	73.68% ± 0.1765	0.2830
6	F4	72.15% ± 0.1291	78.02% ± 0.1879	0.2719
1	FP1	80.52% ± 0.1397	84.87% ± 0.0923	0.2370
11	CZ	65.88% ± 0.2079	73.26% ± 0.1829	0.2357
13	T4	70.98% ± 0.1012	75.46% ± 0.1271	0.2354
16	P3	69.35% ± 0.0643	73.36% ± 0.1068	0.1759
5	FZ	66.34% ± 0.1572	75.18% ± 0.2189	0.1626
10	C3	66.69% ± 0.1230	74.53% ± 0.1831	0.1330
18	P4	68.83% ± 0.0575	73.29% ± 0.1074	0.1277
19	T6	71.29% ± 0.0411	74.40% ± 0.0627	0.0814
17	PZ	67.63% ± 0.0692	73.42% ± 0.1185	0.0791
9	T3	71.25% ± 0.0667	77.16% ± 0.1214	0.0766
21	O2	70.98% ± 0.0501	74.90% ± 0.0736	0.0646
20	O1	71.79% ± 0.0483	75.53% ± 0.0654	0.0523
4	F3	67.82% ± 0.1005	77.97% ± 0.1752	0.0381
15	T5	71.65% ± 0.0414	75.98% ± 0.0597	0.0138
3	F7	73.72% ± 0.0649	80.00% ± 0.0822	0.0125
2	FP2	74.23% ± 0.0659	84.61% ± 0.1071	0.0011
7	F8	72.07% ± 0.0497	79.63% ± 0.0761	0.0009

Table 4: *p*-Values for channels using LZW (Dataset B).

5 Conclusions

The investigation reported in this paper suggests that universal compression methods provide a potentially useful way for objective assessment of AD. The results show that compression ratios are lower in AD group than in the normal group. In future, we will investigate the impact of artefacts on the performance of the method and its robustness in practice.

Acknowledgements

We acknowledge the financial support of the European Commission (The BIOPATTERN Project, Contract No. 508803).

References

[1] Henderson G. T., Ifeachor E. C., Hudson N. R., Goh C., Outram N., Wimalaratna S. K., Percio C. De., and Vecchio F. (2006): "Development and assessment of methods for detecting dementia using the human electroencephalogram", *IEEE Trans. Biomed. Eng.*, **53**, pp.1557-1568.

[2] Abásolo D., Gómez C., Poza J., García M., Sánchez C. I. and López M. (2005): "EEG background activity analysis in Alzheimer's disease patients with sample entropy", in *Int. Conf. Computational Bioengineering*, Lisbon, Portugal.

[3] National Institute on Aging [online] (2006): "Progress report on Alzheimer's Disease 2004-2005", available: http://www.nia.nih.gov/Alzheimers/Publications/ADProgress2004_2005/

[4] Ziv J. and Lempel A. (1978): "A universal algorithm for sequential data compression", *IEEE Trans. Inf. Theory*, **23**, pp. 337-343.

[5] Welch T. A. (1984): "A technique for high-performance data compression", *IEEE Computer*, **17**, pp. 8-19.
[6] Ziv J. and Lempel A.(1978):"Compression of individual sequences via variable-rate coding", *IEEE Trans. Inf. Theory*, **24**, pp. 530-536.