Proceedings of the 29th Annual International
Conference of the IEEE EMBS
Cité Internationale, Lyon, France
August 23-26, 2007.

Characterization of EEGs in Alzheimer's Disease using Information Theoretic Methods

Peng Zhao[1], Peter Van-Eetvelt [1], Cindy Goh[1], Nigel Hudson[2], Sunil Wimalaratna[3], and Emmanuel Ifeachor[1]

[1]University of Plymouth, UK, [2]Derriford Hospital Plymouth, UK, [3]Radcliffe Infirmary, Oxford UK,

Abstract—**The number of people that now go on to develop Alzheimer's disease (AD) and other types of dementia is rapidly rising. For maximum benefits from new treatments, the disease should be diagnosed as early as possible, but this is difficult with current clinical criteria. Potentially, the EEG can serve as an objective, first line of decision support tool to improve diagnosis. It is non-invasive, widely available, low-cost and could be carried out rapidly in the high-risk age group that will develop AD. Changes in the EEG due to the dementing process could be quantified as an index or marker. In this paper, we investigate two information theoretic methods (Tsallis entropy and universal compression algorithm) as a way to generate potentially robust markers from the EEG. The hypothesis is that the information theoretic makers for AD are significantly different to those of normal subjects. An attraction of the information theoretic approach is that, unlike most existing methods, there may be a natural link between the underlying ideas of information theoretic methods, the physiology of AD and its impact on brain functions. Data compression has not been investigated as a means of generating EEG markers before and is attractive because it does not require a priori knowledge of the source model. In this paper, we focus on the LZW algorithm because of its sound theoretical foundation. We used the LZW algorithm and Tsallis model to compute the markers (compression ratios and normalized entropies, respectively) from two EEG datasets. The results show that the information theoretic methods can be used to compute EEG markers for AD.**

Key words - **EEG, dementia, Alzheimer's disease, markers, information theory, Tsallis entropy, LZW algorithm, universal compression.**

I. INTRODUCTION

As more and more people live longer, the number of people that develop Alzheimer's disease (AD) and other types of dementia is rapidly rising. Treatments to slow the progress of the disease are being developed, but to get maximum benefits the disease should be diagnosed as early as possible. However, the delay between actual onset and clinical diagnosis of AD using the current clinical criteria may be up to 5 years so that a significant amount of irreversible cell damage will have occurred by the time a diagnosis is made. Potentially, the electroencephalogram (EEG) may be used as an effective first line of decision support tool to improve diagnosis of AD. It is non-invasiveness, widely available, low-cost and could be carried out rapidly in the high-risk age group that will develop AD.

EEG changes could be detected fairly early in the dementing process and the changes could be quantified as an index or marker. Nonlinear methods provide a potentially powerful method for deriving suitable markers, and recent studies suggest that they are better suited to EEG analysis than conventional approaches [1][2]. However, there are several limitations of existing nonlinear measures which impact on their potential usefulness as a marker [1][2].

In this paper, we investigate an important class of nonlinear methods, information theoretic approaches, which offers a potentially powerful method for deriving robust markers. An attraction of these approaches is that, unlike most existing methods, there may be a natural link between the underlying ideas of information theoretic methods, the physiology of AD and its impact on brain functions [3]. Conceptually, information processing activities in the brain are thought to be reflected in the information content of the EEG. Thus, the effects of damage or disruption to nerve cells/pathways in the brain due to AD may be reflected in information theory-based markers. By implication, changes in the features of the EEG due to AD may be related to changes in the information content of the EEG.

There have been recent attempts to develop nonlinear AD measures based on information theory (e.g. sample entropy), but these tend to approach the problem from nonlinear dynamical concepts [2], and the performance is sensitive to values of the parameters of the algorithm used, such as embedding parameters. In this paper, we present two different information theory-based approaches and their use to generate markers from the EEG. In particular, we seek to exploit knowledge from universal data compression methods and entropy to quantify the EEG. The hypothesis is that the information theoretic measures for AD are significantly different to those of normal subjects.

The rest of the paper is organized as follows. In Section II, a brief description of the two information theoretic methods is given, and in Section III we present details of the data used in the study. The results from the use of the two information theory methods are presented and discussed in Section IV, and finally Section V concludes the paper.

II. INFORMATION THEORETIC METHODS

A key information theoretic measure that can be computed from the EEG record is the entropy. Entropy is a fundamental parameter and provides a measure of the information content of the data [4]. In terms of the EEG it can be viewed as a

This work is supported by the European Union under the BIOPATTERN project, through the Network of Excellence Project FP6-2002-IST-1 No 508803.

measure of the uncertainty or order/disorder in the time-series [4], and it is low for an ordered source and maximum for a disordered source.

Various methods exist for estimating the entropy of a time-series such as the EEG [4][5], but most are based on the pioneering work of Shannon [6]. An entropy measure (Shannon's entropy) can be calculated directly from the EEG data samples by examining the probability distribution of the amplitudes of the data values. An alternative measure due to Tsallis generalizes the Shannon entropy and assumes that the EEG is a non-extensive source [7]. For computing the EEG markers, normalized entropies are used to ensure that the entropy lies between 0 and 100%, making it easier to compare markers for AD and for normal subjects. The normalized Tsallis entropy is show in (1):

$$NTSE(N,q) = \frac{\sum_{i=1}^{N}(p_i - p_i^q)}{1 - N^{1-q}} \quad (1)$$

where N is the number of states that the amplitudes of the EEG are quantized into, p_i is the probability associated with the i^{th} state, and q is the Tsallis parameter.

In practice, changes in the EEG activities (e.g. due to changes in the condition of the subject or experimental conditions, such as during cognitive and event-related potential (ERP) tasks) mean that the source model is not fixed and so most methods of computing estimates of entropy may not be appropriate in all cases. Ideally, what is needed is a method that can be used to estimate information theoretic EEG measures regardless of the underlying source model. Potentially, universal compression algorithms provide a way to satisfy such a requirement because they adapt to the source statistics and do not require a priori knowledge of the source model [8]. Given that compression and entropy [9] are related, this provides a means of estimating entropy and other information theoretic measures. From information theoretic perspective, compression involves reduction in the redundancy in the data. In lossless compression, the lower bound is the entropy. Universal method provides an alternative approach to estimate the entropy such that the estimations converge to the 'true' value of entropy for almost every source. Potentially, these may be more suited to the analysis of EEG in a wide range of situations where the features change than model-based approaches which look for probability of occurrence.

Data compression has not been investigated as a means of generating EEG markers before. In this paper, we focus on the LZW algorithm because of its sound theoretical basis [8][10] [11].

The LZW (Lempel-Ziv-Welch) algorithm is a variant of the basic Lemple-Ziv (LZ) algorithm [8][10][11]. Like the LZ algorithm, it adapts to the source statistics and so requires no prior information about the input data and exploits different types of redundancies in the data. Although the compression effectiveness is about the same, LZW is simpler.

The LZW examines the characters in the input data stream serially using a parsing algorithm and from this dynamically builds a string table. For the EEG, the characters are derived from the data samples. As the algorithm examines the characters in the input data stream, character strings that have not been encountered before are added to the string table and each given a fixed, unique code (typically, 12 bits long). By the end of the compression process, the string table essentially represents the input data in compressed form. The effectiveness of compression may be expressed as the ratio of the number of bits needed to represent the data before and after compression. The more times character strings are repeated in the input data, the longer the strings in the table and the better the compression achieved. The LZW compression is reversible and so the original data can be recovered without loss by decompression.

III. SUBJECTS AND EEG RECORDINGS

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

A. Subjects

Dataset A includes 3 Alzheimer's patients, and 8 age-matched controls (over 65 years old) of which all have normal EEGs confirmed by a consultant clinical neurophysiologist. Within age-matched controls, one of the volunteers (Vol1) is of particular interest because it subsequently developed to AD, therefore potentially in transition from normal to AD.

Dataset B includes 24 normal subjects and 17 probable AD, which are not perfectly age matched. In normal groups, mean age is 69.4±11.5, minimum is 40, maximum is 84, 42% are male; in AD group, mean age is 77.6±10.0, minimum is 50, maximum is 93, 53% are male. These probable AD subjects were not previously diagnosed, and 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.

B. 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 by averaging two consecutive samples for storage reasons, which was confirmed by analysis to have no significant effect on the data because the band of the interest

is low and less than 30Hz [1]. A predetermined protocol was applied to avoid the possibility of inadvertently or unconsciously selecting only data segments that 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 give a standard 4 minutes data to analyze.

IV. RESULTS AND DISCUSSIONS

The LZW algorithm and the Tsallis model were applied to datasets A and B to generate compression ratios and normalized entropies, respectively, for each subject. The channel numbered from 1 to 21 correspond respectively to montage FP1, FP2, F7, F3, FZ, F4, F8, A1, T3, C3, CZ, C4, T4, A2, T5, P3, PZ, P4, T6, O1, O2. Fig. 1 shows plots of the compression ratios for each of the 21 channels and for each subject in dataset A using the LZW.

A. LZW Ratio of Dataset A

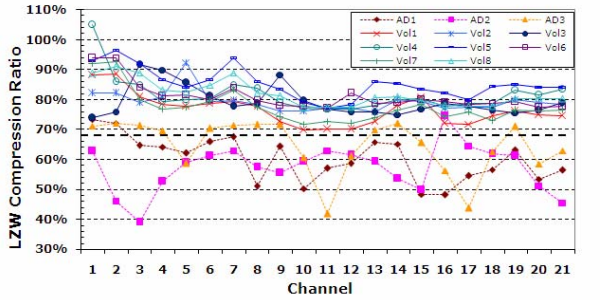

Fig. 1. LZW compression ratio for all cases in channels (Dataset A).

The sensitivity and specificity measures provide an indication of performance, but acceptable values for these depend on the application. 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 can 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 the Dataset A, this gives the sensitivity of 77.78%.

Fig. 2, Fig. 3 and Fig. 4 show, respectively, the average compression values across channels for each case, the average compression values for each channel and across subjects in each of the two subject groups, and the average compression values in key regions of the scalp across subjects in each of the two groups. These clearly show that the AD group has a lower LZW compression ratio than normal group, and that the two groups are well separated by examining the channels, as well as the regions (for example, the *p*-values in TABLE I). The sensitivity for all channels is 95.24% and 83.33% for all regions.

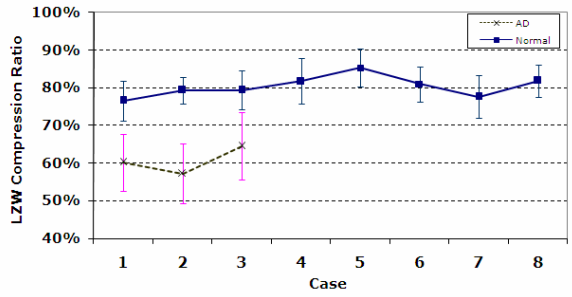

Fig. 2. The average and standard deviation of all channels in each case (Dataset A).

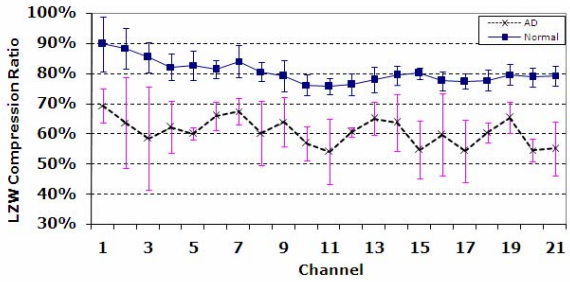

Fig. 3. The average and standard deviation of all cases in each channel (Dataset A).

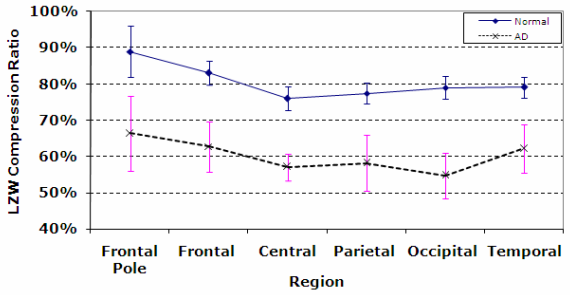

Fig. 4. The average and standard deviation of all cases for each region (Dataset A).

TABLE I
P-VALUE FOR REGIONS USING LZW (DATASET A)

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

B. Tsallis Entropy of Dataset A

We also carried out a similar study using normalized Tsallis Entropy (see Fig. 6, Fig. 7 & Fig. 8). In this case, a threshold of 0.22 is used to achieve 100% specificity, with a corresponding sensitivity of 80.95% (see Fig. 5). The sensitivity across channels is 85.71 % and 77.78% across regions.

These also clearly show that the AD group has a lower LZW compression ratio than normal group, and that the two groups are well separated by examining the channels, as well as the

regions (for example, the *p*-values in TABLE II).

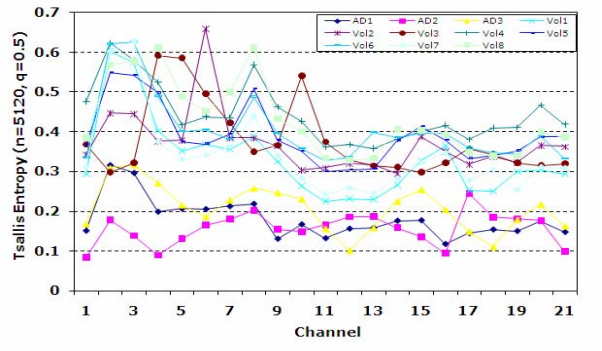

Fig. 5. Tsallis entropy for all cases in channels (Dataset A).

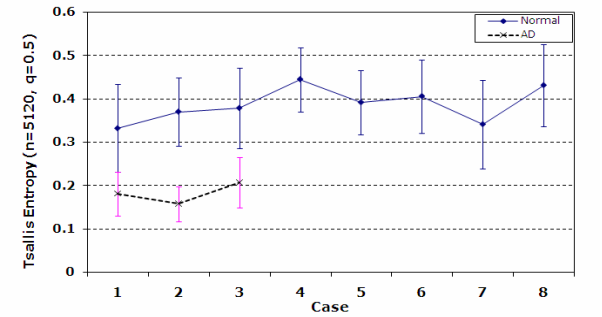

Fig. 6. The average and standard deviation of all channels in each case (Dataset A).

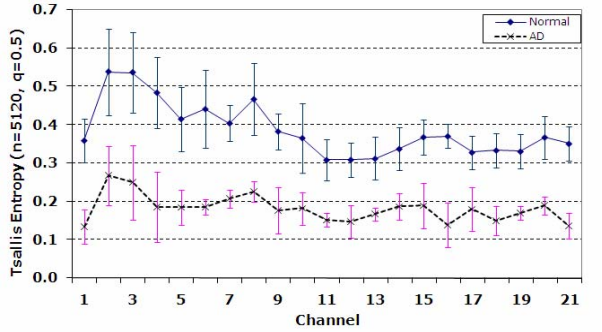

Fig.7. The average and standard deviation of all cases in each channel (Dataset A).

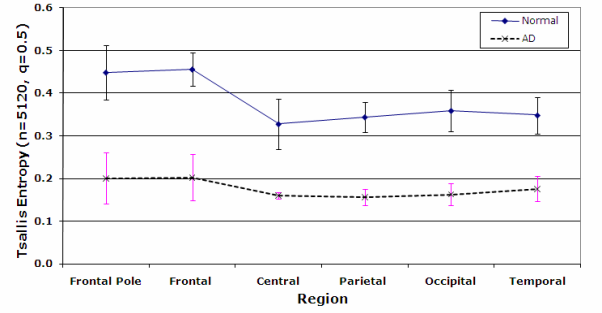

Fig.8. The average and standard deviation of all cases for each region (Dataset A).

TABLE II
P-VALUES FOR REGIONS USING TSALLIS ENTROPY (DATASET A)

Region	Control subjects (Mean ± SD)	AD patients (Mean ± SD)	*p*-value
Central	0.3279 ± 0.0592	0.1604 ± 0.0080	1.0760E-03
Frontal Pole	0.4488 ± 0.0637	0.2009 ± 0.0606	2.5544E-04
Temporal	0.3484 ± 0.0428	0.1759 ± 0.0291	1.3272E-04
Occipital	0.3591 ± 0.0489	0.1627 ± 0.0258	1.1562E-04
Frontal	0.4561 ± 0.0398	0.2026 ± 0.0541	1.2013E-05
Parietal	0.3440 ± 0.0351	0.1560 ± 0.0184	1.2007E-05

C. LZW Ratio and Tsallis Entropy of Dataset B

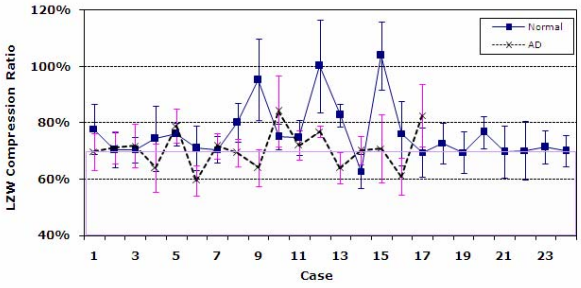

Fig. 9. The average and standard deviation of all channels in each case (Dataset B).

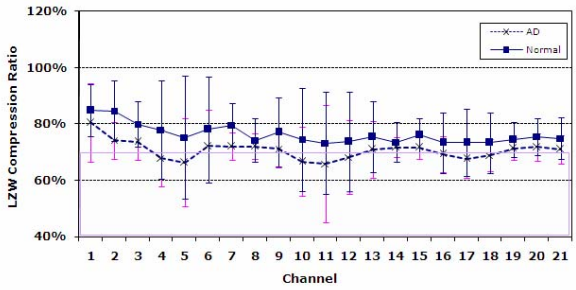

Fig. 10. The average and standard deviation of all cases in each channel (Dataset B).

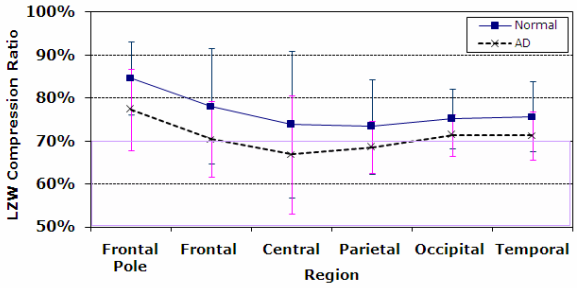

Fig. 11. The average and standard deviation of all cases for each region (Dataset B).

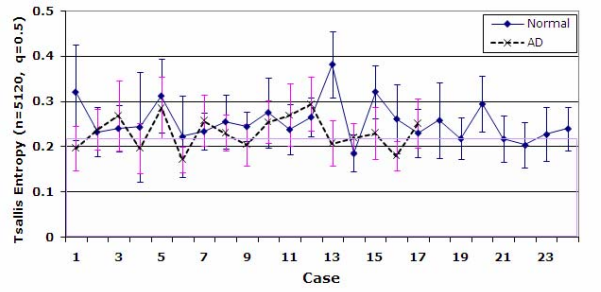

Fig. 12. The average and standard deviation of all channels in each case (Dataset B).

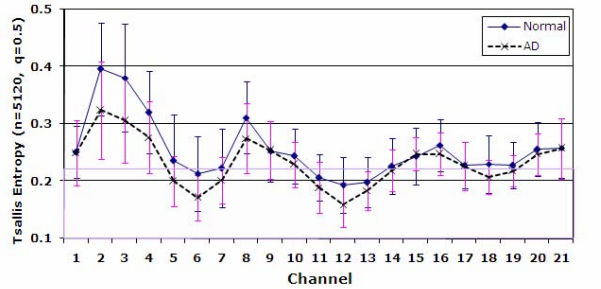

Fig. 13. The average and standard deviation of all cases in each channel (Dataset B).

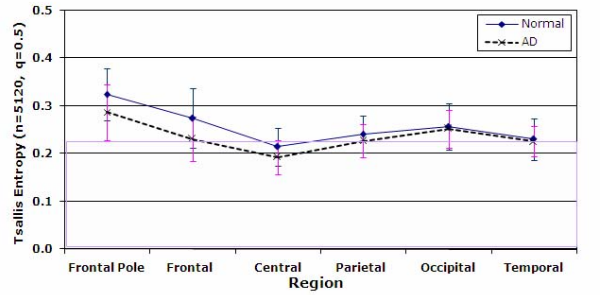

Fig. 14. The average and standard deviation of all cases for each region (Dataset B).

V. CONCLUSIONS

The investigation reported in this paper suggests that information theoretic methods provide a potentially useful way for generating EEG markers for AD. The results show that both measures (compression ratio and normalized Tsallis entropy) are lower in AD than for the normal group. However, for noisy data there may be a need to pre-process the data before the methods are applied.

In future, we will investigate the impacts of the condition of the subject on the information theoretical markers (e.g. sleep, ERP experiments).

REFERENCES

[1] G. T. Henderson *et al.*, "Development and assessment of methods for detecting dementia using the human electroencephalogram," *IEEE Trans. Biomed. Eng.*, vol. 53, no. 8, pp.1557-1568, Aug. 2006.

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

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

[4] N. V. Thakor and S. Tong, "Advances in quantitative electroencephalogram analysis methods," *Annu. Rev. Biomed. Eng.*, vol. 6, pp. 453-496, 2004.

[5] C. Gomez, R. Hornero, D. Abasolo, M. Lopez, and A. Fernandez, "Decreased Lempel-Ziv complexity in Alzheimer's disease patients' magnetoencephalograms," in *Proc. Ann. Int. Conf. IEEE Engineering in Medicine and Biology Soc.*, 2005, pp. 4514- 4517.

[6] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, pp. 379–423 and vol. 28, pp. 623–656, Jul. and Oct. 1948.

[7] C. Tsallis, "Possible generalization of Boltzmann-Gibbs statistics," *J. Stat. Phys.*, vol. 52, pp. 479-487, 1988.

[8] J. Ziv and A. Lempel, "A universal algorithm for sequential data compression," *IEEE Trans. Inf. Theory*, vol. 23, no. 3, pp. 337-343, May 1978.

[9] D. S. Ornstein and B. Weiss, "Entropy and data compression schemes," *IEEE Trans. Inf. Theory*, vol. 39, no.1, pp. 78-83, Jan. 1993.

[10] T. A. Welch, "A technique for high-performance data compression," *IEEE Computer*, vol. 17, no. 6, pp. 8-19, Jun. 1984.

[11] J. Ziv and A. Lempel, "Compression of individual sequences via variable-rate coding," *IEEE Trans. Inf. Theory*, vol. 24, no. 5, pp. 530-536, Sep. 1978.