Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 126 (2018) 156–165

Procedia Computer Science

www.elsevier.com/locate/procedia

International Conference on Knowledge Based and Intelligent Information and Engineering Systems, KES2018, 3-5 September 2018, Belgrade, Serbia

A Study of Combined Lossy Compression and Seizure Detection on Epileptic EEG Signals

Binh Nguyen[a], Wanli Ma[a], Dat Tran[a,*]

[a] *Faculty of Science and Technology, University of Canberra, ACT 2601, Australia*

Abstract

Electroencephalogram (EEG) has been widely used in diagnosing and detecting epileptic seizure. Large epileptic EEG databases have been built, the use of EEG compression is therefore becoming necessary. Epilepsy causes a change on EEG characteristics, especially on frequency, hence exploiting these features may improve the performance of EEG lossy compression techniques that are mostly working on frequency domain. In this paper, we propose a lossy compression method for epileptic EEG data, by exploiting the characteristics of EEG under epilepsy. Moreover, the recovered EEG signals processed by the proposed method are used by an EEG-based seizure detection system to evaluate the possibility of applying in real world as well as the impact of lossy compression on seizure detection. The results show that the proposed method gives a higher result, and applying the proposed method to EEG-based seizure detection system is feasible and has the advantage compared to using lossless ones.

Keywords: EEG; Epileptic EEG signal; lossy compression; DWT-AAC; EEG-based seizure detection;

1. Introduction

Epilepsy is one of the most common neurological diseases affecting people of all ages, race and social class, and 50 million people with epilepsy in the world are estimated [1,2]. However, diagnosis, detection and treatment of epilepsy are not easy tasks. Electroencephalogram (EEG) has been widely used in detecting epileptic seizures. Kannathal *et al*. pointed out that one of the most effective way in diagnosis and detection of epileptic seizures by neurologists and neurophysiologists is the usage of EEG signals and it has a long history [3]. According to Garry *et al*., albeit EEG signals can be used in many applications, the most popular application using EEG is in the diagnosis and detection of epilepsy [4]. Nevertheless, one of the major challenges in using EEG signals for seizure detection is very large epileptic EEG databases have been built. Therefore, an effective EEG compression technique for epileptic EEG signals is necessary to reduce data for transmitting, processing and storing.

Lossless and lossy compressions are two techniques that have been proposed for EEG data. Lossless compression reduces the size of the compressed data without any loss of information in the reconstructed data, whilst lossy compression does not ensure for perfect reconstruction of the data. However, the compression ratio (CR) of lossy

* Corresponding author. Tel.: +61-02-6201-2394.
E-mail address: Dat.Tran@canberra.edu.au

1877-0509 © 2018 The Authors. Published by Elsevier Ltd.

10.1016/j.procs.2018.07.219

compression technique is much higher than that of the lossless one. Consequently, data compression studies mostly focus on the lossy technique. Currently, lossy compression techniques can be categorised into four groups, namely Wavelet-based, Filter-based, Predictor and Other. In addition, Wavelet-based indicated is the most common in lossy techniques [5].

Although EEG signals are complex and non-stationary, some studies indicated that they become less complex and chaotic in epilepsy condition. In particular, Kannathal *et al*. and Sanei *et al*. pointed out that during epilepsy condition, the cortex becomes inactive and EEG signals become less random [3,6]. Additionally, the power spectrum is concentrated in specific frequency bands during epilepsy condition [7]. Moreover, Arroyo *et al*. and Jirsch *et al*. also stated that epileptic seizures cause the change of EEG frequencies [8,9]. It is said that epileptic EEG signals are different from normal EEG ones in some aspects such as frequency. However, current lossy techniques have been proposed for general EEG data rather than for specific ones such as epileptic EEG data. Hence, exploiting the difference in epileptic EEG signals may improve the performance of compression techniques.

In this paper, we propose an EEG lossy compression technique that extended from the one developed by Nguyen *et al*. in [10], by exploiting the characteristics of EEG signals under epilepsy to improve the CR while unchanging the loss of information in recovered signals. Moreover, a seizure detection system is used to evaluate not only the proposed lossy technique performance, but also the impact of this lossy technique on seizure detection.

The rest of the paper is organized as follows. Sections 2 and 3 present the background information and the methods and materials, respectively. Section 4 outlines the results, followed by a conclusion in Section 5.

2. Background information

2.1. Epileptic EEG signals

Epilepsy is a clinical condition characterised by recurrent two or more epileptic seizures, unprovoked by any direct identified cause. According to the presumption of the clinical manifestation, epileptic seizures are the result from an abnormal and excessive discharge of a set of neurons in the brain. The sudden and transitory abnormal phenomena including alterations of consciousness, motor, sensory are indications of the clinical manifestation [11]. EEG signals that are recorded during a seizure activity are called ictal, whilst inter-ictal EEG signals are captured in the period between seizures. Moreover, pre-ictal and post-ictal refer to the state immediately before and after the actual seizure, respectively [12].

According to Sanei and Chambers, after transforming into frequency domain, the signal energy spreads over different frequencies due to the nature, complex and diversity of EEG signals [6]. Nevertheless, epileptic EEG signals become less chaotic and complex than normal ones [3,6], which may result in the energy of the signal becomes more concentrated on some frequency bands. This assumption is compatible with findings reported by Mormann *et al*. that the power spectrum will be concentrated in specific frequency bands during epileptic seizures [7]. These characteristics of epileptic EEG signals will be exploited to improve CR of the proposed compression scheme.

2.2. Lossy compression technique based on Discrete Wavelet Transform and Adaptive Arithmetic Coder (DWT-AAC)

Figure 1 illustrates the lossy compression technique based on Discrete Wavelet Transform (DWT) and Adaptive Arithmetic Coder (AAC) proposed by Nguyen *et al*. in [10]. Firstly, EEG signals are transformed from one-dimensional into two-dimensional matrix. In this scheme, matrix size chosen is 128x128. Each matrix is passed through a DWT, which decomposes signals into sub-bands. Afterwards, the decomposed coefficients are quantised and then thresholded such that those having the absolute values below the threshold are set to zero. However, for recovered signals in decompression stage, positions of zeroed coefficients must be kept to create a binary significance map where bits 1 and 0 indicate the position of significant and zeroed coefficients, respectively. Finally, the binary significance map and indices of significant coefficients are encoded using an AAC to generate the compressed signals. To decompress the signals, an inverse process is employed.

2.3. Compression performance measures

Two widely-used metrics to measure the performance of compression algorithms are CR and PRD.

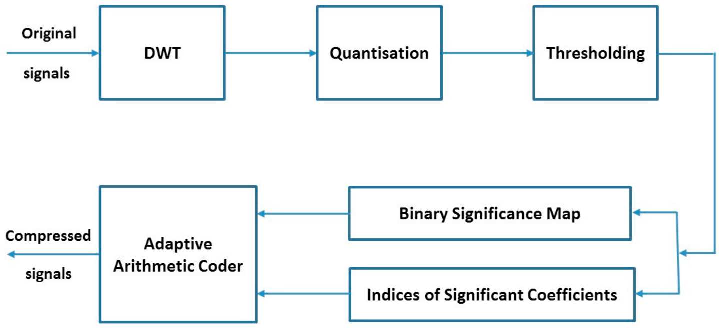

Fig. 1. The architecture of DWT-AAC

CR is defined as the ratio between the uncompressed size and compressed size:

$$CR = \frac{L_0}{L_c} \tag{1}$$

where L_0 and L_c denote the number of bits of original EEG signal and the number of bits of compressed EEG signal, respectively.

PRD is a standard metric to evaluate the distortion between the original and reconstructed signals. It is defined as:

$$PRD = \sqrt{\frac{\sum_{i=1}^{N}(x[i] - x'[i])^2}{\sum_{i=1}^{N}(x[i])^2}} \times 100\% \tag{2}$$

where $x[i]$ represents the original EEG signal, $x'[i]$ represents the compressed signal, and N is the number of samples.

Beside CR and PRD, $Increase\ of\ CR$ is used to compare CRs of DWT-AAC and proposed method.

$$Increase\ of\ CR = \frac{CR_{Proposed\ method} - CR_{DWT-AAC}}{CR_{DWT-AAC}} \times 100\% \tag{3}$$

When both CRs are equal, the value is zero. The value of $Increase\ of\ CR$ is positive if $CR_{Proposed\ method}$ is greater than $CR_{DWT-AAC}$, otherwise that value is negative.

2.4. EEG-based seizure detection system

The architecture of EEG-based seizure detection system proposed in [13] is illustrated in Figure 2. Initially, EEG signals are transformed into the time-frequency representation (TFR) by using Hilbert-Huang Transform. The TFR can be considered as time-frequency image (TFI), so the image processing methods can be used to solve the problem of seizure detection. Afterwards, TFI is segmented into five sub-images corresponding to the frequency bands of the rhythms. In present work, TFI is divided into five bands, namely delta (0.5-4 Hz), theta (4-8 Hz), alpha (8-13 Hz), beta (13-30 Hz) and gamma (30-50 Hz)[6]. The next step is feature extraction in which three features including mean (μ), variance (σ^2) and skewness (α) of pixel intensity are calculated from the histogram of the grayscale sub-images. Finally, these features are fed into the Support Vector Machine (SVM) to classify seizures.

3. Methods and Materials

3.1. Proposed compression scheme

Studies in [14,15] pointed out that ECG signal energy concentrates most on low frequencies after decomposition, so thresholding will generate the longer 0s consecutive streams, which helps algorithms based on Run Length Coding

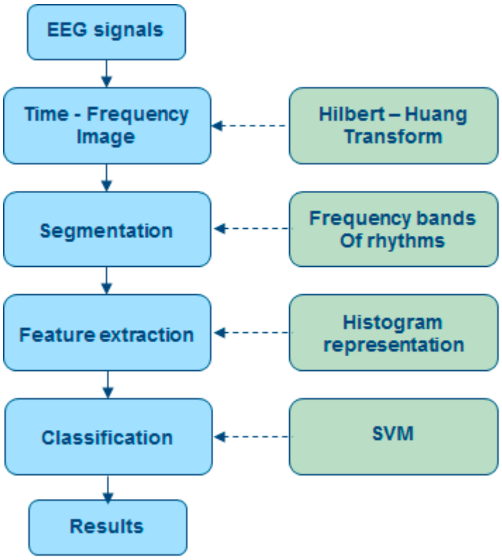

Fig. 2. The architecture of EEG-based Seizure Detection System

(RLC) achieve a high compression. Similarly, Bazán-Prieto also stated that when signal energy spreads over different frequencies, thresholding will not produce a larger number of zero-valued coefficients [16]. Although normal EEG signal characteristics are random, complex and non-stationary [6], epileptic EEG signals are less chaotic, complex and signal energy concentrates on some frequency bands [3,6,7]. Hence, it is said that epileptic EEG signals have some features that are quite similar to ECG signals, so it is expected to create the longer 0's consecutive streams in binary significance map and a smaller number of significant coefficients after thresholding. However, initial experiments on epileptic EEG data from Born University [17] conducted using both AAC and RLC on the binary significance map, the results indicated that using RLC was not as efficient as using ACC. This could be because epileptic EEG signals are still more random, complex and the signal energy spreads out more than ECG ones, so applying RLC is not effective. Therefore, the binary significance map will be coded by using AAC in the proposed method like in DWT-AAC.

Table 1. Statistics on normal EEG signals

PRD(%)	Number of coefficients	Last position of significant coefficient	Number of last consecutive insignificant coefficients
11.4672	16,384	16,384	0
18.9782	16,384	16,378	6
28.146	16,384	16,000	384
37.64	16,384	12,736	3,648
46.7581	16,384	12,672	3,712

Table 2. Statistics on epileptic EEG signals

PRD(%)	Number of coefficients	Last position of significant coefficient	Number of last consecutive insignificant coefficients
12.0986	16,384	16,384	0
18.4871	16,384	15,872	512
28.0209	16,384	15,616	768
32.7586	16,384	8,178	8,206
47.9265	16,384	8,060	8,324

Although the EEG signal under epilepsy is not completely similar to ECG one, its characteristics helps to produce the longer 0's consecutive streams in binary significance map compared to the normal EEG one, which is exploited to improve CR of the proposed method. Table 1 and 2 demonstrate the statistics of binary significance map performed on two block (128x128) of normal and epileptic EEG signals respectively. At the approximately same PRD, after thresholding, epileptic EEG data give much longer number of last consecutive insignificant coefficients than normal ones. At PRD of roughly 19%, for example, the number of last consecutive insignificant coefficients of normal EEG is only 6 whilst the figure of epileptic one is 512. Similarly, at PRD of roughly 47%, the number of the former is 3,712, compared to 8,324 for the latter.

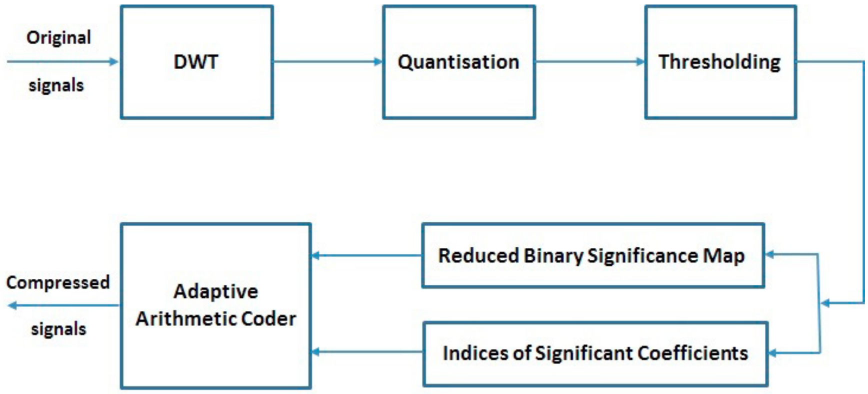

Fig. 3. Typical architecture of proposed method

In the scheme of DWT-AAC, the whole binary significance map of runs of ones and zeros is coded by using AAC. However, the proposed scheme will use a reduced binary significance map that its length equals the position of the last significant coefficient instead of the whole one. The number of total coefficients can be calculated from block size and the mode of wavelet, and the remaining insignificant coefficients will be added in the decoder to make the integrity of the binary significance map. For instance, from Table 2, at PRD of 32.7586%, the proposed method will use a reduced binary significance map of 8,178 bits instead of using the whole one of 16,384 bits, which improves significantly CR of the proposed method while unchanging the value of PRD.

Figure 3 illustrates the typical architecture of the proposed method in which a reduced binary significance map is used to replace the whole one. All configurations from DWT-AAC is kept the same because these settings gave high compression performance [10]. Particularly, the proposed method is set up to run with 5 levels of biorthogonal 4.4 2-D DWT. Moreover, 6 bits uniform quantisation and hard threshold are used. The performance of proposed method is not affected when compressing normal EEG signals because the reduced binary significance map will become the whole one in the worst case in which the proposed method becomes DWT-AAC. However, the proposed method is expected to be suitable with epileptic EEG signals thanks to its characteristics.

3.2. EEG database

The public EEG dataset used is obtained from Born University [17]. There are five sets denoted A-E in this dataset. Each set containing 100 single-channel EEG segments of 23.6-sec duration was sampled at the frequency of 173.61 Hz. Sets A and B were EEG recordings of five healthy volunteers recorded with eyes open and eyes close respectively. Signals in set D were recorded from within the epileptogenic zone while those in set C from the hippocampal formation of the opposite hemisphere of the brain. Moreover, signals in sets C and D were measured during seizure free intervals. Conversely, set E only contained seizure activities. In this paper, the epileptic EEG signals from datasets D (inter-ictal) and E (ictal) are used to evaluate the performance of proposed technique whilst datasets A (normal) and E, and datasets D and E are used to examine the impact of lossy compression on seizure detection.

3.3. Test conditions

The signals from datasets D and E were compressed at a range of different CRs by using both DWT-AAC and the proposed method. Increase of CRs was then calculated from CRs of both techniques to compare the performance.

The reconstructed EEG signals from datasets A and D, and datasets D and E processed by DWT-AAC and the proposed method at different CRs were used by the EEG-based seizure detection system to determine: 1) if seizure activities can still be detected from EEG signals that processed by those lossy compression techniques and 2) the impact of increasing compression and then the loss of signal fidelity on seizure detection performance.

The signal was processed and features were extracted individually from each EEG channel, and then features from all channels are merged together. We used 2/3 of the dataset for cross validation training and 1/3 for testing. Linear SVM classifiers [18] were trained in 3-folds cross validation scheme with parameter C ranging from 1 to 1000 in 5 steps.

Features for training and testing were randomly chosen ten times, and the final experimental result is the average of these ten results.

4. Results

4.1. Compression performance

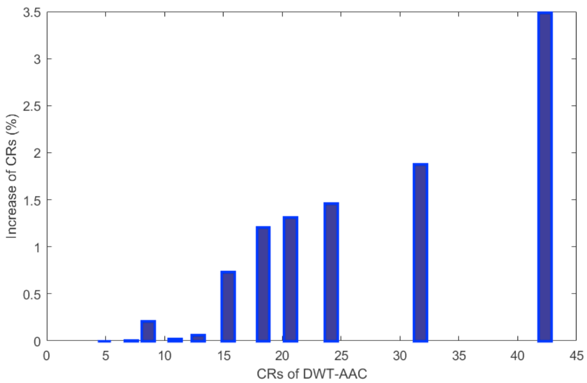

Fig. 4. Increase of CRs between proposed method and DWT-AAC on dataset E

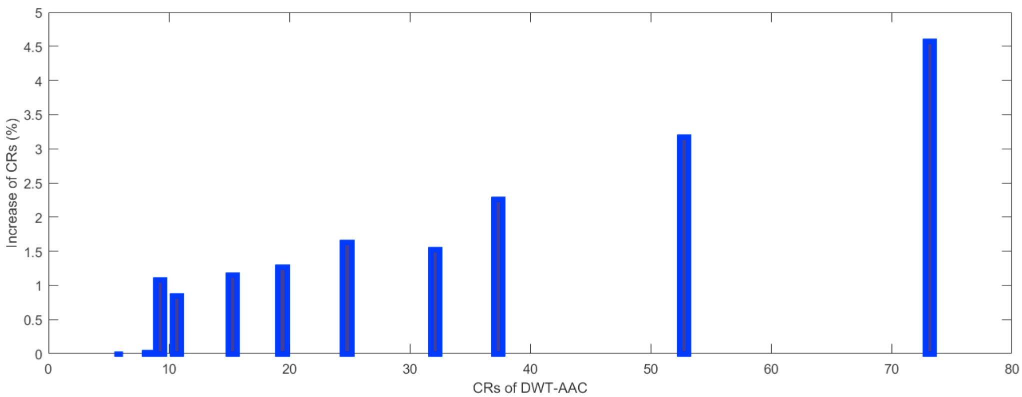

Fig. 5. Increase of CRs between proposed method and DWT-AAC on datataset D

Figure 4 and 5 illustrate the Increase of CRs between the proposed method and DWT-AAC on datasets E and D, respectively. Generally, no negative values of Increase of CR has been reported and there are increase trends of Increase of CRs on both datasets. In other words, the proposed method obtains better compression results than DWT-AAC, especially at high CRs. In particular, on dataset E (ictal EEG data), CR of the proposed method increases 0.734% when CR of DWT-AAC is 15.4. This figure augments to 3.5% when CR of DWT-AAC reaches 42.3. Similarly, on dataset D (inter-ictal EEG data), the Increase of CR of proposed method is 0.85% when CR of DWT-AAC is 10.6, while this number is 3.17% when DWT-AAC obtains CR of 52.76.

4.2. Seizure detection performance with increasing compression

Table 3 demonstrates the results of seizure detection using original (uncompressed) EEG signals from datasets D and E. It can be seen that Theta band gives the best performance while Delta has the worst one. These results are suitable with findings reported in[13]. Hence, the Theta waveform was used to classify epileptic seizure in this paper.

Table 3. Seizure detection performance using original EEG signals from datasets D and E

Bands	Accuracy rate(%)
Delta EEG	84.84
Theta	**96.20**
Alpha EEG	95.45
Beta EEG	93.93
Gamma EEG	95.45

Figures 6 and 7 show average accuracy rates versus average CRs for DWT-AAC and the proposed method using Theta band on datasets A and E, and datasets D and E, respectively. It can be seen that when CRs augment, the classification accuracies degrade gradually. This is because PRDs will rise as CRs increase, generating more loss of diagnostically information in recovered signals. Additionally, the results on datasets A and E are better than those on datasets D and E. For example, at the seizure classification accuracy of 91%, CR of both techniques is around 21 on the former, whilst this figure is only 13 on the latter. The most probable reason is that the differences between normal and ictal EEG signals are greater than those between inter-ictal and ictal ones, making higher classification accuracies. Moreover, although the seizure detection performances for both DWT-AAC and the proposed method are nearly similar, those for the latter are slightly better than those for the former, especially at high CRs. On datasets A and E, for instance, at accuracy of 88%, CR of proposed method is 32.33, compared to 31.7 for DWT-AAC. Similarly, on datasets D and E, CR of proposed method is 24.5 at accuracy of 86.3% while that of DWT-AAC is 24.1 at the same accuracy.

4.3. Maximising compression and Information lost

According to Faul and Higgins, a classification accuracy greater than 90% is considered very good performing classifier for seizure detection[19,20]. Hence, 90% accuracy rate is used as a threshold limit for compression in this work.

Referring back to Figure 6, at accuracy rate of 90% on datasets A and E, CRs of DWT-AAC and the proposed method are 24 and 24.48, respectively. From Figure 8, the equivalent PRD is 34% at 90% accuracy. Similarly, referring back to Figure 7 and from Figure 9, CRs of DWT-AAC and the proposed method are 15.4 and 15.5, correspondingly at the same PRD of 18.48% on datasets D and E. For these results, it is said that the proposed method obtains slight higher CRs than DWT-AAC at the same accuracy rate of 90%.

Studies in[21,22] stated that CR of lossless techniques of EEG signals achieves only between 2 and 3. Therefore, applying lossy compression to EEG-based seizure detection system achieves much higher CRs while having little effect on seizure detection performance, compared to using lossless ones.

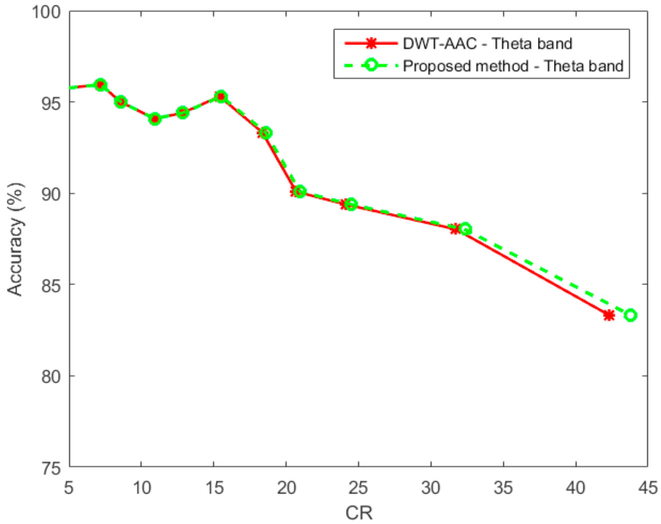

Fig. 6. Average accuracy rates versus average CRs for both compression techniques using the theta band on datasets A and E.

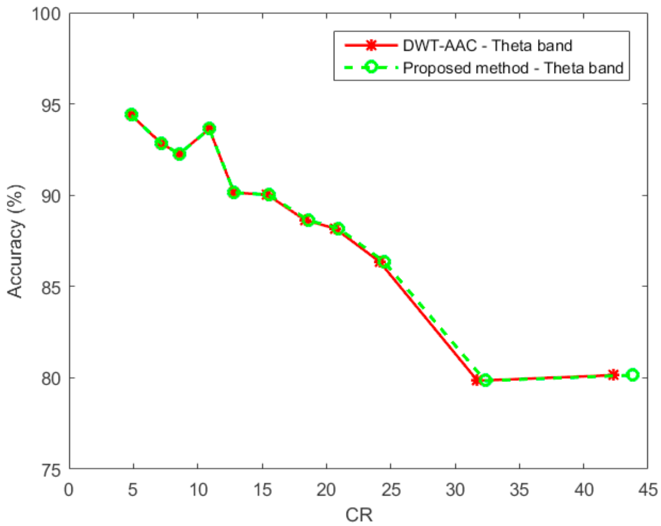

Fig. 7. Average accuracy rates versus average CRs for both compression techniques using the theta band on datasets D and E.

5. Conclusion

We have presented a lossy compression technique for epileptic EEG signals based on DWT-AAC proposed in[10], by exploiting the epileptic EEG signals' characteristics. In this scheme, the reduced binary significance map is used instead of using the whole one, improving the CR while keeping PRD unchanging.

Inter-ictal and ictal EEG data were used to evaluate the compression performance of proposed method. Moreover, the reconstructed EEG signals from normal and ictal EEG, and inter-ictal and ictal ones were fed into an EEG-based seizure detection system to validate the proposed method performance as well as investigate the impact of lossy technique on seizure detection.

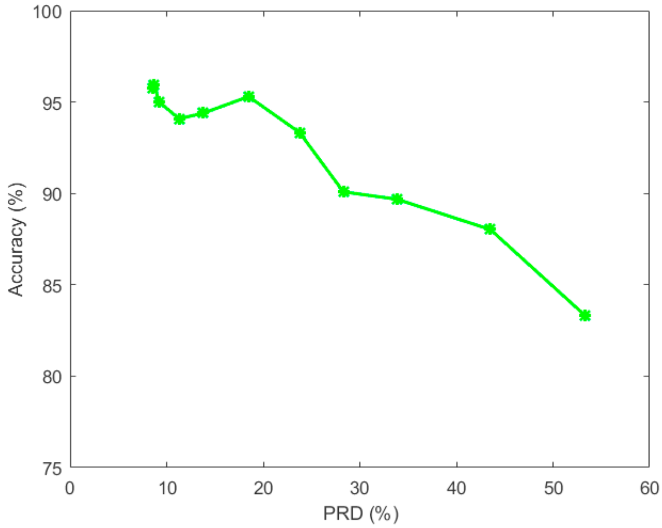

Fig. 8. Average accuracy rates versus average PRDs for both compression techniques using the theta band on datasets A and E.

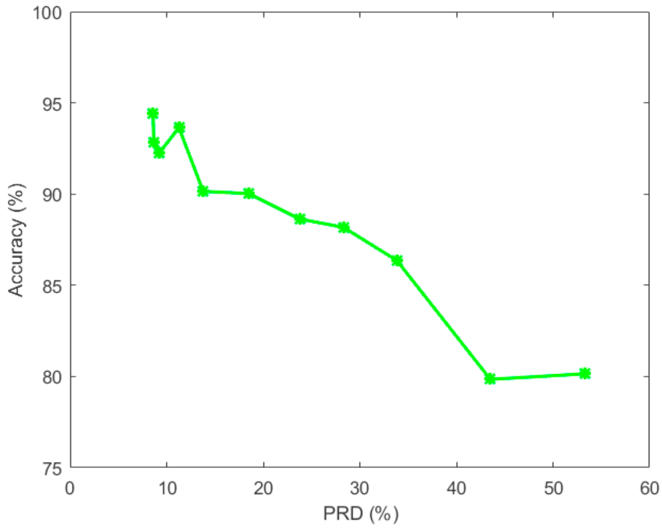

Fig. 9. Average accuracy rates versus average PRDs for both compression techniques using the theta band on datasets D and E.

Our results show that the proposed technique gives better results than DWT-AAC, especially at high CRs. Furthermore, although the seizure detection performance will reduce when increasing compression, applying lossy compression is feasible and still has more benefits compared to using lossless ones.

References

1. Neligan, A., Sander, J.. The incidence and prevalence of epilepsy. *London: UCL Institute of Neurology* 2009;.
2. WHO, . Epilepsy 2017;URL: `http://www.who.int/mediacentre/factsheets/fs999/en/`.
3. Kannathal, N., Acharya, U.R., Lim, C.M., Sadasivan, P.. Characterization of eega comparative study. *Computer methods and Programs in Biomedicine* 2005;**80**(1):17–23.
4. Garry, H., McGinley, B., Jones, E., Glavin, M.. An evaluation of the effects of wavelet coefficient quantisation in transform based eeg compression. *Computers in biology and medicine* 2013;**43**(6):661–669.

5. Dao, P.T., Li, X.J., Do, H.N.. Lossy compression techniques for eeg signals. In: *Advanced Technologies for Communications (ATC), 2015 International Conference on*. IEEE; 2015, p. 154–159.
6. Sanei, S., Chambers, J.A.. *EEG signal processing*. Wiley-Interscience; 2007.
7. Mormann, F., Lehnertz, K., David, P., Elger, C.E.. Mean phase coherence as a measure for phase synchronization and its application to the eeg of epilepsy patients. *Physica D: Nonlinear Phenomena* 2000;**144**(3):358–369.
8. Arroyo, S., Uematsu, S.. High-frequency eeg activity at the start of seizures. *Journal of Clinical Neurophysiology* 1992;**9**(3):441–448.
9. Jirsch, J., Urrestarazu, E., LeVan, P., Olivier, A., Dubeau, F., Gotman, J.. High-frequency oscillations during human focal seizures. *Brain* 2006;**129**(6):1593–1608.
10. Nguyen, B., Nguyen, D., Ma, W., Tran, D.. Wavelet transform and adaptive arithmetic coding techniques for eeg lossy compression. In: *Neural Networks (IJCNN), 2017 International Joint Conference on*. IEEE; 2017, p. 3153–3160.
11. International League Against Epilepsy, C.o.E., Prognosis, . Guidelines for epidemiologic studies on epilepsy. *Epilepsia* 1993;**34**(4).
12. Mula, M., Monaco, F.. Ictal and peri-ictal psychopathology. *Behavioural neurology* 2011;**24**(1):21–25.
13. Fu, K., Qu, J., Chai, Y., Dong, Y.. Classification of seizure based on the time-frequency image of eeg signals using hht and svm. *Biomedical Signal Processing and Control* 2014;**13**:15–22.
14. Blanco-Velasco, M., Cruz-Roldán, F., López-Ferreras, F., Bravo-Santos, Á., Martínez-Muñoz, D.. A low computational complexity algorithm for ecg signal compression. *Medical engineering & physics* 2004;**26**(7):553–568.
15. Abo-Zahhad, M., Rajoub, B.A.. An effective coding technique for the compression of one-dimensional signals using wavelet transforms. *Medical engineering & physics* 2002;**24**(3):185–199.
16. Bazán-Prieto, C., Cárdenas-Barrera, J., Blanco-Velasco, M., Cruz-Roldán, F.. Electroencephalographic compression based on modulated filter banks and wavelet transform. In: *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE; 2011, p. 7067–7070.
17. Andrzejak, R.G., Lehnertz, K., Mormann, F., Rieke, C., David, P., Elger, C.E.. Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state. *Physical Review E* 2001;**64**(6):061907.
18. Chang, C.C., Lin, C.J.. Libsvm: a library for support vector machines. *ACM transactions on intelligent systems and technology (TIST)* 2011;**2**(3):27.
19. Faul, S., Temko, A., Marnane, W.. Age-independent seizure detection. In: *Engineering in Medicine and Biology Society, 2009. EMBC 2009. Annual International Conference of the IEEE*. IEEE; 2009, p. 6612–6615.
20. Higgins, G., McGinley, B., Faul, S., McEvoy, R.P., Glavin, M., Marnane, W.P., et al. The effects of lossy compression on diagnostically relevant seizure information in eeg signals. *IEEE journal of biomedical and health informatics* 2013;**17**(1):121–127.
21. Antoniol, G., Tonella, P.. Eeg data compression techniques. *IEEE Transactions on Biomedical engineering* 1997;**44**(2):105–114.
22. Srinivasan, K., Reddy, M.R.. Efficient preprocessing technique for real-time lossless eeg compression. *Electronics Letters* 2010;**46**(1):26–27.