iMEDiTEC 2017

ORIGINAL PAPER

Complexity Analysis on EEG Signal via Lempel-Ziv and Approximate Entropy: Effect of Multiresolution Analysis

Mohd Syakir b Fathillah [a,*], Rosmina Jaafar [a], Kalaivani Chellappan [a], Rabani Remli [b], Wan Asyraf Wan Zaidi [b]

[a] Jabatan Kejuruteraan Elektrik, Elektronik & Sistem, Fakulti Kejuruteraan dan Alam Bina, Universiti Kebangsaan Malaysia
[b] Neurology Unit, Department of Medicine, Pusat Perubatan Universiti Kebangsaan Malaysia

* Corresponding author: syakirfathillah92@gmail.com

ABSTRACT

Lempel Ziv has been applied in biomedical signal interpretation for measuring depth of anaesthesia, characterization of brain function and study of emotional and cognitive processing. Despite Lempel Ziv (LZ) popularity as a technique to measure complexity, the understanding of the LZ application on biomedical signal has not been fully addressed. This paper is focusing on comparison study between LZ and Approximate Entropy (ApEn) and how multiresolution analysis (MRA) affects both techniques. We assess the performance of both techniques in time domain first using ANOVA to investigate the ability of both techniques in distinguishing between normal, interictal and ictal seizure dataset. MRA then is implemented and a similar test is conducted. Our findings show that ApEn is more sensitive to the presence of epileptic discharge compare to LZ. MRA has positive effect on ApEn where it provides more detail analysis based on sub-bands. The effect of MRA provides better enhancement for ApEn compare to LZ in terms of differentiating between normal, interictal and ictal.

INTRODUCTION

Biomedical signal is a continuous, time-varying record that conveys information regarding an internal functioning of a biomedical system. A biomedical signal usually acquired by two ways, whether by a transducer and is converted to a voltage or current for further analysis or by recording directly using electrode (Robert B . Northrop 2010)

Most of the biomedical signals are non-stationary including electroencephalogram (EEG) signal. The condition producing these signal vary over time, making EEG signal to exhibit nonstationary, nonlinear, stochastic, dynamic and also complex behaviour (Klonowski 2009). These properties have been used by the previous researcher to investigate in EEG study related to characterization of brain function (Grassberger et al. 1991), study of emotional (Li et al. 2016), cognitive processing (Natarajan et al. 2004), measurement of depth of anaesthesia (Zhang et al. 2001), and seizure detection (Ocak 2009).

To evaluate the complexity of a time series, Lempel and Ziv have proposed a method related to the number patterns and the rate of their occurrence in a sequence (Lempel & Ziv 1976). From data compression algorithm, LZ has been adopted in many biomedical data such as schizophrenia (Fernández et al. 2013), mechanomyography (Sarlabous et al. 2013), Alzheimer (Abásolo et al. 2006), monitoring the brain state in anaesthesia (Bai, Liang, Li, et al. 2015) and seizure detection (Bai, Liang & Li 2015). Currently, there has been extensive use of LZ in EEG analysis because it has several advantages (Ibáñez-Molina et al. 2015): (a) can be applied to any time series signal (b) can be applied to short time signal (c) lower computational cost.

Despite of its popularity as a technique to measure complexity, the understanding of the LZ application on biomedical signal has not been fully addressed. Jing Hu et al. (Hu et al. 2006) studies the effect of finite size data on LZ complexity and drawn out conclusion that LZ is almost independent with sequence length. The normalized LZ complexity is able to outperform correlation entropy in detecting epileptic seizure. Aboy et al. (Aboy et al. 2006) investigated interpretability of LZ complexity in concepts such as frequency, frequency variability of the harmonics and signal bandwidth.

In this paper, we present a comparison study on LZ with another commonly used complexity measurement technique, the Approximate Entropy (ApEn) through the detection of seizure in EEG. We also test the effect of multiresolution analysis (MRA) on both techniques.

MATERIALS AND METHOD

EEG Data

This study utilizes four sets of online public EEG data (A, B, and C) which were acquired from the Department of Epileptology, University of Bonn database (Andrzejak et al. 2012). Each set contains 100 single channels that were recorded using 128-channel amplifier system. The duration of each data is 23.6 seconds, sampled at 173.61Hz and was band-pass filtered from 0.53 Hz to 40 Hz. Set A consists of 5 healthy subjects where the subjects were awake and relaxed with eye opened. For set B, it contains interictal epileptic discharge (interictal seizure) and was recorded within epileptogenic region respectively. Set C contained ictal epileptic discharge (ictal seizure) activities. The interictal epileptic discharge has the same characteristic as ictal, only in shorter duration and no symptom shown by the patient (Staley & Dudek 2006). The electrodes placement is according to the international 10-20 system and the summarization of data sets detail are as in Table 1.

Table 1 Summary of clinical data.

	Set 1 (A)	Set 2 (B)	Set 3 (C)

eISBN 978-967-0194-93-6 FBME

Subject Condition	Healthy subject with eyes open	Interictal epileptic discharge	Ictal epileptic discharge
Electrode placement	International 10-20 systems	Within epileptogenic zone	Within epileptogenic zone

Seizure data is utilized in this study to demonstrate the change of regularity during normal, interictal and seizure period. During the release of epileptic discharge, EEG signal appear regular and repetitive (Ocak 2009) allowing LZ and ApEn to measure it complexity.

Discrete Wavelet Transform (MultiResolution Analysis)

Discrete wavelet transform is adopted to utilize multiresolution analysis (MRA). The wavelet transform is mathematical techniques where it can convert the signal into a scaled and shifted version of the mother wavelet and express it in terms of frequency and time. The main benefit of the wavelet transformation is to decompose a signal into a sub-band frequency which is MRA and can be described by the equation (1):

$$f(t) = \sum_{j\in Z} 2^{j/2} c_j(k)\varphi\left(2^j t - k\right) + \sum_{j=0}^{j-1}\sum_{k=0}^{\infty} 2^{j/2} d_j(k)\omega\left(2^j t - l\right) \quad (1)$$

where ϕ(t) is a scaling function, ψ (t) is a basis function and j is the scale index. The signal undergoes high pass filter and low pass filter that will produce an approximation of f(t) and detail of f(t) respectively which will be presented in finer scale. The types of wavelets play important role in the wavelet transform. In this study, we adopted Daubechies 4 (db4) as mother wavelet because its smoothing feature is suitable in detecting EEG changes (Omerhodzic et al. 2010). Decomposition level is set to six to correlate it with a classification of wave frequencies as shown in Table II.

Table 2 Classification of sub-band according to type of wave.

Coefficient	Frequency Band (Hz)	Type of wave	Level of Decomposition
D1	43.40-86.81	Noise	1
D2	21.70-43.40	Beta-Gamma	2
D3	10.85-21.70	Alpha-Beta	3
D4	5.43-10.85	Theta-Alpha	4
D5	2.71-5.43	Delta-Theta	5
A5	0.5-2.71	Delta	5

Lempel-Ziv Complexity

To compute LZ complexity, the signal S(n) must be transformed into a symbolic sequence first. Typically, converting the signal into a binary sequence is adequate for biomedical signal analysis (Aboy et al. 2006). This can be done by comparing threshold T_d with the signal data, S as follows:

$$S = s(1), s(2), \dots, s(n) \quad (2)$$

where

$$s(i) = \begin{cases} 1, & if\ s(i) < T_d \\ 0, & otherwise \end{cases} \quad (3)$$

Commonly the median is used as the threshold due to its robustness to outliers (Aboy et al. 2006). The signal S is scanned from left to right and the complexity counter c(n) will increase each time new subsequence of consecutive character is come up. The complexity measure can be obtained by following algorithm (Abásolo et al. 2007):

1) Let P be a signal which contain two subsequences, S and Q. Let SQ be the sequence of S and Q, while SQπ is derived from SQ after its last character deleted. Let v(SQπ) indicate the vocabulary of all different subsequences of SQπ. For started, value c(n)=1, S=s(1), Q=s(2) and SQπ=s(1).
2) For generalization, S=s(1), s(2), …, s(r), Q=s(r+1) and SQπ=s(1), s(2), …, s(r). If Q fits in v(SQπ), then Q is a subsequence of SQπ which is not a new sequence.
3) Renew Q to be s(r+1), s(r+1) and check if Q belongs to v(SQπ) or not.
4) Step 3 is repeat until Q does not belongs to v(SQπ), which means Q is not a subsequence of v(SQπ).
5) SQπ is a new sequence, so value of c(n) increase by one.
6) S is renewed to be S=s(1), s(2), …, s(r+i) and Q=s(r+i+1).
7) Repeat procedure until Q is the last character.

The complexity measure depends on number of different subsequences in P. To acquire complexity independent to sequence length, c(n) need to be normalized as follow (4):

$$C(n) = \frac{c(n)\times\log_3 n}{n} \quad (4)$$

Approximate Entropy

Approximate Entropy first was developed by Pincus to measure system complexity (Pincus 1991). In signal analysis, ApEn helped to measure the regularity and predictability of a signal. The value of ApEn can be determined by the following procedure.

1) Let a data sequence containing n data points be S_n = {u(1), u(2), u(3),….,u(n)]
2) Choose value of m and r where m = pattern length and r = criterion of similarity $r = k \times SD$ for k=0,0.1,0.2,0.3,…,0.9. SD will be present as standard deviation of data S_n
3) Let X be sequence of x(i) such that x(i)=[u(i),u(i+1),u(i+2),…,u(i+m-1)] where i=1,2,3,…,(n-m+1)
4) Find the distance between vector x(i) and x(j) by using formula $d[x,x^*] = \max_a |u(a) - u^*(a)|$, if $d[x,x^*] < r$ the pattern are likely similar
5) Calculate $C_i^m = \frac{number\ of\ d[x,x^*] less\ than\ r}{(n-m+1)}$ and $C_i^{m+1} = \frac{number\ of\ d[x,x^*] less\ than\ r}{(n-m+1)}$
6) Define $\Phi^m(r) = \frac{\sum_{i=1}^{n-m+1}\ln(C_i^m(r))}{n-m+1}$ and $\Phi^{m+1}(r) = \frac{\sum_{i=1}^{n-m+1}\ln(C_i^m(r))}{n-m+1}$
7) ApEn(m,r,n) is determined as follow: $ApEn(m,r,n) = \Phi^m(r) - \Phi^{m+1}(r)$

Large ApEn value indicates the signal is unpredictable and irregular while a small ApEn value indicate higher regularity and repetitive pattern. To determine the ApEn, the (m) and (r) are set to 2 and 0.2xSD respectively based on Srinivasan et. al (Srinivasan et al. 2007) to obtain the highest percentage of efficiency.

RESULTS AND DISCUSSION

Signal Complexity without MRA

Initially, we applied LZ and ApEn to three different set of EEG data, Set A, B and C in time domain. We analysed the significant level of complexity mean of difference group and the result is tabulated in Table 3 and Table 4 for LZ and ApEn, respectively.

Table 3 Mean complexity of LZ with significant value between dataset.

Dataset	Mean±Std.Deviation	Dataset	sig.
Set A	0.0695±0.0235	Set B	0.389
		Set C	0.000*
Set B	0.0656±0.0255	Set A	0.389
		Set C	0.000*
Set C	0.0233±0.0109	Set A	0.000*

eISBN 978-967-0194-93-6 FBME

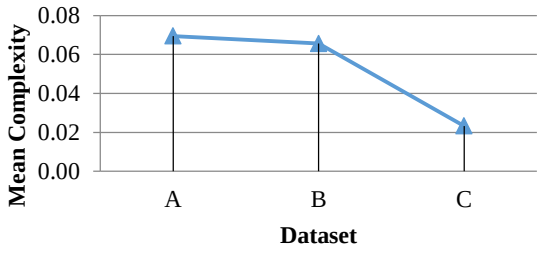
		Set B	0.000*

Table 4 Mean complexity of ApEn with significant value between dataset.

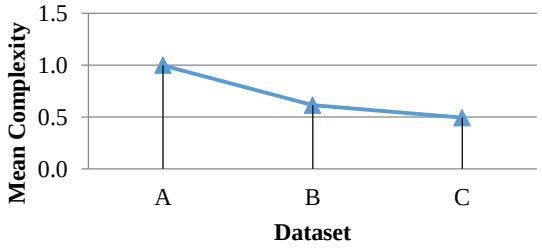
Dataset	Mean±Std.Deviation	Dataset	sig.
Set A	0.9971±0.1558	Set B	0.000*
		Set C	0.000*
Set B	0.6137±0.1961	Set A	0.000*
		Set C	0.000*
Set C	0.4926±0.1438	Set A	0.000*
		Set B	0.000*

Based on Table 3, Set A yields the highest mean complexity followed by Set B and Set C. The complexity decrease in EEG that contain epileptic discharge. The outcome of ANOVA test reveal that there is significant difference between Set C with set A and B but there is no significant difference between Set A and Set B (p-value>0.01).

On the other hand ApEn outcome (Table 4) is consistent with LZ where it exhibit decreasing trend from Set A followed by Set B and Set C. However, the ANOVA show positive outcome where there is significant difference among Set A, B and C (*p*-value<0.01). This test shows that ApEn is more sensitive in detecting the presence of epileptic discharge compare to LZ. Graph of mean complexity for LZ and ApEn can be observed in Fig 1(a) and Fig1(b), respectively.



(a)



(b)

Fig.1 Mean complexity vs dataset for (a) LZ (b) ApEn

Signal Complexity with MRA in Time-Frequency Domain

Next, we apply DWT on signal to do multiresolution analysis. The signal is decomposed into sub-band. Complexity can be assessed in each sub-band. The average complexity of each sub-band is tabulated in Table 4 and Table 5 for LZ and ApEn, respectively.

Table 4 Mean complexity of LZ using MRA with significant value between dataset.

Dataset	Sub-band	Mean±Std.Deviation	Dataset	sig.
A	cA5	0.2387±0.036058	Set B	0.000*
			Set C	0.000*
	cD5	0.6643±0.0412	Set B	0.051
			Set C	0.342
	cD4	1.1686±0.0613	Set B	0.000*
			Set C	0.000*
	cD3	1.7630±0.0928	Set B	0.000*
			Set C	0.000*
	cD2	2.1157±0.1511	Set B	0.000*
			Set C	0.000*
	cD1	3.2088±0.1870	Set B	0.006*
			Set C	0.882
B	cA5	0.2761±0.0477	Set A	0.000*
			Set C	0.000*
	cD5	0.6470±0.0445	Set A	0.051
			Set C	0.609
	cD4	1.1022±0.0587	Set A	0.000*
			Set C	0.020
	cD3	1.6841±0.1032	Set A	0.000*
			Set C	0.007*
	cD2	2.2672±0.0845	Set A	0.000*
			Set C	0.000*
	cD1	3.2795±0.1416	Set A	0.006*
			Set C	0.001*
C	cA5	0.3198±0.0490	Set A	0.000*
			Set B	0.000*
	cD5	0.6540±0.0667	Set A	0.342
			Set B	0.609
	cD4	1.0710±0.1130	Set A	0.000*
			Set B	0.020
	cD3	1.626±0.1861	Set A	0.000*
			Set B	0.007*
	cD2	1.9956±0.1875	Set A	0.000*
			Set B	0.000*
	cD1	3.1979±0.1525	Set A	0.882
			Set B	0.001*

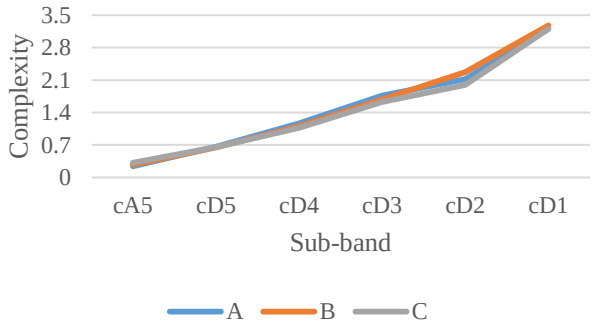
Table 5 Mean complexity of ApEn using MRA with significant value between dataset.

Dataset	Sub-band	Mean±Std.Deviation	Dataset	sig.
A	cA5	0.1619±0.022268	Set B	0.000*
			Set C	0.000*
	cD5	0.4347±0.023383	Set B	0.001*
			Set C	0.000*
	cD4	0.5939±0.01886	Set B	0.000*
			Set C	0.005*
	cD3	0.8004±0.026509	Set B	0.000*
			Set C	0.000*
	cD2	1.0029±0.050259	Set B	0.003*
			Set C	0.000*
	cD1	1.5230±0.065739	Set B	0.004*
			Set C	0.000*
B	cA5	0.1909±0.043464	Set A	0.000*
			Set C	0.000*
	cD5	0.4079±0.073705	Set A	0.001*
			Set C	0.000*
	cD4	0.5438±0.084005	Set A	0.000*
			Set C	0.069
	cD3	0.7022±0.140353	Set A	0.000*
			Set C	0.000*
	cD2	0.9121±0.203186	Set A	0.003*
			Set C	0.000*
	cD1	1.4161±0.249503	Set A	0.004*
			Set C	0.000*
C	cA5	0.2309±0.041916	Set A	0.000*
			Set B	0.000*
	cD5	0.4719±0.045282	Set A	0.000*
			Set B	0.000*
	cD4	0.5646±0.075406	Set A	0.005*
			Set B	0.069
	cD3	0.6258±0.17937	Set A	0.000*
			Set B	0.000*
	cD2	0.5978±0.263945	Set A	0.000*
			Set B	0.000*
	cD1	0.7676±0.318454	Set A	0.000*
			Set B	0.000*

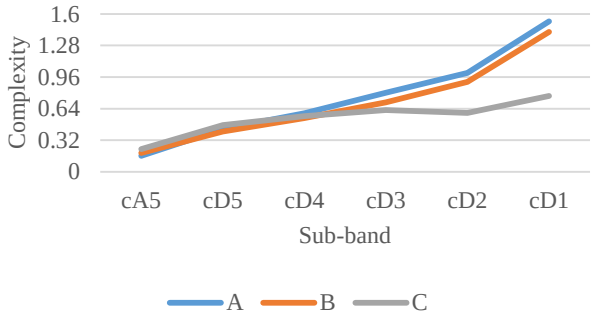
eISBN 978-967-0194-93-6 FBME

Graph of mean complexity for LZ and ApEn can be observed in Fig 2(a) and Fig 2(b), respectively. The complexity of Set C is consistent between LZ and ApEn. This result supports the fact that epileptic discharge during seizure make EEG signal be more deterministic. However, we found out that LZ is less sensitive to the presence of epileptic discharge. This is due to its ability to discriminate the normal (Set A) and interictal seizure (Set B). This trait can be seen in analysis without MRA where there is no significant difference between Set A and Set B (p-value>0.01). In MRA analysis, particularly in sub-band cD2 and cD1, complexity of Set B is higher than Set A, which is contradicting to the fact that presence of epileptic discharge exhibit regular pattern. This result also contradicts with ApEn result. The sub-band cD5 in LZ show no significant difference among three datasets (p-value>0.01). However, ApEn show good performance in distinguishing all sub-bands with significant difference (p-value<0.01) except between Set B and Set C in sub-band cD4 (p-value<0.0619). This indicates that effect of MRA is able to provide better enhancement to differentiate between dataset for ApEn compare to LZ.

A study was conducted to investigate the relationship of LZ complexity with frequency (Aboy et al. 2006). It is stated that LZ complexity increases as the frequency of a sinusoid increase. This concluded that the LZ complexity is independent with frequency content. However, an objection on this matter was expressed in (Balasubramanian et al. 2013). The author claims the fact that LZ complexity value increases only at low frequency (0.1 to 1 Hz) and hovers around a constant value at frequency 5 to 50 Hz. In our case, the effect of MRA proves that the LZ complexity is dependent with frequency. This can be seen by the increase of LZ value as sub-band increases even in the frequency higher than 5 Hz. Our finding correlates with (Aboy et al. 2006).



(a)



(b)

Fig. 2 Mean complexity vs sub-band for (a) LZ (b) ApEn.

CONCLUSION

This study discussed the complexity analysis using LZ and ApEn. It is shown that the ApEn is more sensitive in detecting the presence of epileptic discharge compare to LZ. The MRA has a positive effect on ApEn, enabling more detail analysis and exhibit a clear trend to distinguish subject normal with interictal and ictal seizure. LZ on the other hand shows a random trend in complexity. We conclude MRA enhance ApEn better compare to LZ in terms of distinguishing between dataset. Further analysis is needed to demonstrate the performance of MRA and LZ combination especially as feature for seizure detection algorithm.

ACKNOWLEDGEMENT

This work was financially supported by the grant from Universiti Kebangsaan Malaysia (LRGS/2013/UMK-UKM/SS/04).

REFERENCES

Abásolo, D., Hornero, R., Gómez, C., García, M. & López, M. 2006. Analysis of EEG background activity in Alzheimer's disease patients with Lempel-Ziv complexity and central tendency measure. *Medical Engineering and Physics* 28(4): 315–322.

Abásolo, D., James, C. J. & Hornero, R. 2007. Non-linear analysis of intracranial electroencephalogram recordings with approximate entropy and Lempel-Ziv complexity for epileptic seizure detection. *Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Conference* 1953–1956.

Aboy, M., Hornero, R., Abásolo, D. & Álvarez, D. 2006. Interpretation of the Lempel-Ziv complexity measure in the context of biomedical signal analysis. *IEEE Transactions on Biomedical Engineering* 53(11): 2282–2288.

Andrzejak, R. G., Schindler, K. & Rummel, C. 2012. Nonrandomness, nonlinear dependence, and nonstationarity of electroencephalographic recordings from epilepsy patients. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics* 86(4).

Bai, Y., Liang, Z. & Li, X. 2015. A permutation Lempel-Ziv complexity measure for EEG analysis. *Biomedical Signal Processing and Control* 19: 102–114.

Bai, Y., Liang, Z., Li, X., Voss, L. J. & Sleigh, J. W. 2015. Permutation Lempel–Ziv complexity measure of electroencephalogram in GABAergic anaesthetics. *Physiological Measurement* 36(12): 2483–2501.

Balasubramanian, K., Prabhu, G. R. & Nagaraj, N. 2013. Comment on'Interpretation of the Lempel-Ziv Complexity Measure in the context of Biomedical Signal Analysis'. *arXiv preprint arXiv:1308.0130* 10–12.

Fernández, A., Gómez, C., Hornero, R. & López-Ibor, J. J. 2013. Complexity and schizophrenia. *Progress in Neuro-Psychopharmacology and Biological Psychiatry* 45: 267–276.

Grassberger, P., Schreiber, T. & Schaffrath, C. 1991. Nonlinear Time Sequence Analysis. *International Journal of Bifurcation and Chaos* 1(3): 521–547.

Hu, J., Gao, J. & Principe, J. C. 2006. Analysis of biomedical signals by the Lempel-Ziv complexity: The effect of finite data size. *IEEE Transactions on Biomedical Engineering* 53(12): 2606–2609.

Ibáñez-Molina, A. J., Iglesias-Parro, S., Soriano, M. F. & Aznarte, J. I. 2015. Multiscale Lempel-Ziv complexity for EEG measures. *Clinical Neurophysiology* 126(3): 541–548.

Klonowski, W. 2009. Everything you wanted to ask about EEG but were afraid to get the right answer. *Nonlinear biomedical physics* 3: 2.

Lempel, A. & Ziv, J. 1976. On the Complexity of Finite Sequences. *IEEE Transactions on Information Theory* 22(1): 75–81.

Li, X., Hu, B., Sun, S. & Cai, H. 2016. EEG-based mild depressive detection using feature selection methods and classifiers. *Computer Methods and Programs in Biomedicine* 136: 151–161.

Natarajan, K., Acharya U, R., Alias, F., Tiboleng, T. & Puthusserypady, S. K. 2004. Nonlinear analysis of EEG signals at different mental states. *Biomedical engineering online* 3(1): 7.

Ocak, H. 2009. Automatic detection of epileptic seizures in EEG using discrete wavelet transform and approximate entropy. *Expert Systems with Applications* 36(2 PART 1): 2027–2036.

Omerhodzic, I., Avdakovic, S., Nuhanovic, A. & Dizdarevic, K. 2010. Energy Distribution of EEG Signals : EEG Signal Wavelet-Neural Network Classifier. *Neurocomputing* 2(1): 210–215.

eISBN 978-967-0194-93-6 FBME

Pincus, S. M. 1991. Approximate entropy as a measure of system complexity. *Proceedings of the National Academy of Sciences* 88(6): 2297–2301.
Robert B . Northrop. 2010. *Signals and Systems Analysis In Biomedical Engineering*. CRC Press. doi:doi.org/10.1201/b15856
Sarlabous, L., Torres, A., Fiz, J. A., Morera, J. & Jané, R. 2013. Index for estimation of muscle force from mechanomyography based on the Lempel-Ziv algorithm. *Journal of Electromyography and Kinesiology* 23(3): 548–557.
Srinivasan, V., Eswaran, C. & Sriraam, N. 2007. Approximate entropy-based epileptic EEG detection using artificial neural networks. *IEEE Transactions on Information Technology in Biomedicine* 11(3): 288–295.
Staley, K. J. & Dudek, F. E. 2006. Interictal Spikes and Epileptogenesis. *Epilepsy Currents* 6(6): 199–202.
Zhang, X. S., Roy, R. J. & Jensen, E. W. 2001. EEG complexity as a measure of depth of anesthesia for patients. *IEEE Transactions on Biomedical Engineering* 48(12): 1424–1433.

eISBN 978-967-0194-93-6 FBME