Linear and Nonlinear Feature Extraction for Neural Seizure Detection

Mohamed A. Elgammal*, Omar A. Elkhouly†, Heba Elhosary§, Mohamed Elsayed¶, Ahmed Nader Mohieldin*, Khaled N. Salama‖ and Hassan Mostafa*‡

*Department of Electronics and Communications Engineering, Cairo University, Egypt
†Department of Electronics and Communications Engineering, The American University in Cairo (AUC), Egypt
‡Nanotechnology Department, Zewail City of Science and Technology, Egypt
¶Communication and Information Engineering Department, Zewail City of Science and Technology, Egypt
§Department of Electronics Engineering, The German University in Cairo(GUC), Egypt
‖King Abdullah University of Science and Technology (KAUST), Saudi Arabia

***Abstract*—In this paper, both linear and nonlinear features have been reviewed with linear support vector machine (SVM) classifier for neural seizure detection. The work introduced in the paper includes performance measurement through different metrics: accuracy, sensitivity, and specificity of multiple linear and nonlinear features with linear support vector machine (SVM). A comparison is performed between the performance of different combinations between 11 linear features and 9 nonlinear features to conclude the best set of features. It is found that some features enhance the detection performance greatly. Using a combination of 3 features of them, a linear SVM classifier detects seizures with sensitivity of 96.78%, specificity of 97.9%, and accuracy of 97.9%.**

***Index Terms*—Seizure, EEG, SVM, Linear features, Nonlinear features.**

I. INTRODUCTION

Epilepsy is a neural disease that affects approximately 1% of the world population (700 million people). Epilepsy is defined by seizures, which is sudden discharges in the brain neurons that can affect the contraction and relaxation of any muscle in the body. Seizures may be fatal and hence some precautions must be taken for an epileptic patient to cope with his life, like having someone around him all the time which may be troublesome for lots of people and may be impossible in some cases. Hence, Automatic seizure detection algorithms have been evolved.

Many research is being done for the automatic detection and prediction of epileptic seizures with different approaches like analysis of the muscles movement, studying the electrocardiogram (ECG) signal of the heart and studying EEG signals [1], [2]. After the signal is obtained, artifacts such as blinking of the eye effects are removed from EEG. Then, some features are extracted from EEG signal. Finally, a classifier is used to detect whether the patient is in a state of seizure or not. Figure 1 shows the four main blocks of automatic seizure detection system.

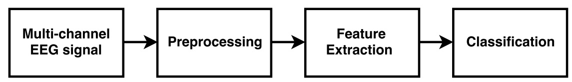


Figure 1: Automatic seizure detection system

Different work is found on the analysis of EEG signals for seizure detection in the literature [3], [4]. Features extracted from EEG along with different machine learning algorithms are used to detect seizure. Yuan Q. et al. used nonlinear feature extraction strategies such as approximate entropy and Hurst exponent and got 93.75% and 79.75% sensitivity respectively [5]. Support vector machine (SVM) is also used in many papers with Radial Basis function (RBF) kernel for classification. Generally, the results obtained through SVM with RBF kernel are usually more accurate, however a hardware implementation for an RBF kernel consumes much more power than linear kernel. Other approaches send EEG signal to be processed offline [6], [7].

Section II provides explanation of the different features extracted from EEG. Section III describes how these features are combined, tested and compared. Section IV describes the simulation results, and provides analysis and discussion of the obtained results. Finally, some conclusions are drawn in Section V.

II. FEATURE EXTRACTION

EEG signal is divided into time epochs each of 4 seconds. In each epoch, different linear and nonlinear features are extracted as shown in the following subsections.

A. Linear Features

Different linear features are implemented, extracted and tested. The 11 linear features are as follows:

- Mean Absolute Value (MAV)

$$MAV = \frac{1}{N}\sum_{i=1}^{N}|x_i|$$

- Root Mean Square (RMS)

RMS was used combined with other features for seizure prediction in [8]. RMS is calculated as follows:

$$RMS = \sqrt{\frac{1}{N}\sum_{i=1}^{N}x_i^2}$$

978-1-5386-7392-8/18/$31.00 ©2018 IEEE

- Standard Deviation (STD)

Standard Deviation is a measure of the average deviation from the mean. It was used in [9] and achieved high performance.

$$SD = \sqrt{\frac{\sum_{i=1}^{N}(x_i - mean(x))}{N-1}}$$

- Variance

Variance is the standard deviation raised to the power of two. It is easier to calculate the variance rather than calculate SD. Hence, both SD and variance are tested to check if easier calculation would reflect on the performance or not.

- Maximum Absolute Value

Calculating the maximum absolute value for every epoch of time. It was used in [9] with other features achieving performance more than 98%.

- Minimum Absolute Value

Calculating the minimum absolute value for every epoch of time.

- Average Energy

In epileptic seizures, the amplitude and frequency of the EEG signal increases. This was a motivation to include the Average Energy of the epoch as a feature. It is defined as follows:

$$E = \sum_{i=1}^{N} x_i^2$$

- Fluctuation Index

Fluctuation Index (FI) measures the fluctuation in the signal. During seizure periods, it is found that EEG exhibits high fluctuations relative to non-seizure periods. FI is defined as follows:

$$FI = \sum_{i=1}^{N} |x_{i+1} - x_i|$$

- Hjorth parameters: Mobility

Mobility is the square root of the variance of the first derivative divided over the variance of the signal.

- Hjorth paramteres: Complexity

Complexity represents the change in frequency with respect to a pure sine wave

- Skew

Skew measures how non symmetric the data is. It was used with other features for classification by Zhang [9]. It is calculated as follows:

$$Skew = \frac{1}{M}\sum_{i=1}^{N}(\frac{X(w) - \mu_w}{\sigma_w})^3$$

where $X(w)$ is sample in frequency domain,μ_w is the mean value of the frequency domain and σ_w is the standard deviation in the frequency domain.

- Kurtosis

Kurtosis is the same as skew but raised to the power 4 as follows:

$$Kurtosis = \frac{1}{M}\sum_{i=1}^{N}(\frac{X(w) - \mu_w}{\sigma_w})^4$$

B. Nonlinear Features

Non-linear analysis of EEG signal exhibit description of the non-stationary nature of the signals. Different features are used by different researchers in the literature. They used many features from information theory, nonlinear dynamical analysis, and stochastic processes analysis. Non-linear features showed promising results in both detection and prediction for epileptic seizures [10]. In this study, different nonlinear features were examined as follows:

- Approximate Entropy (ApEn)

Approximate entropy is a probabilistic method developed by Steve M. Pincus [11]. It measures how ordered or disordered a given EEG signal is. A small output value indicates regularity in the input EEG signal, and on the contrary, as the EEG gets more irregular, the higher the output value becomes [12]. The dataset is divided into overlapping subsequents, where m is the length of each subsequent. $S(i) = [x(i), x(i+1), ..., x(i+m-1)]$, where i =1,2,...,N-m + 1. Then, the algorithm searches for matched patterns by calculating the distance between each subsequent and every other subsequent. Finally, it compares this distance with a certain tolerance r. If the distance is less than the tolerance, the patterns are considered matched which supports the decision of having a regular predictable EEG and vice versa. A distance function $d[x(i), x(j)]$ between each subsequent and every other subsequent. The correlation $log\ C_i^m(r)$ is then calculated by counting the distances that are smaller than a tolerance r and then divided by the number of subsequents N-m + 1. Finally, the logs of these values are summed together and formulating Approximate Entropy as follows:

$$\phi^m(r) = \frac{1}{N-m+1}\sum_{i=1}^{N-m+1} log(C(r))$$

Finally the Approximate Entropy can be calculated as follows:

$$ApEn = \phi^m(r) - \phi^{m+1}(r)$$

- Shannon Entropy

Shannon Entropy is a measure for information that the system exhibits. It estimates the number of bits required to encode a string of symbols based on their frequencies [13]. Continuous values of EEG signals are quantized . Then, the frequency of each symbol is calculated to get Shannon Entropy as follows:

$$H(x) = -\sum_{i=1}^{N} P(x_i).log(P(x_i))$$

where $P(x_i)$ is the probability of the symbol x_i.

- Permutation Entropy

Permutation Entropy, as other entropies, measures how disordered the EEG signal is. However, it is computed independent of the values of the samples. First, a mapping function is applied to generate windows of length n. Probability of a given permutation is given as:

$$P(\pi) = \frac{\#\, of\, windows\, of\, permutation\, \pi}{T-n+1}$$

$$H_n^* = -\sum P(\pi).log(P(\pi))$$

- Renyie Entropy

Renyie Entropy generalizes Shannon Entropy as the parameter α gives extra degree of freedom for the distributions. It is calculated as follows:

$$H(x) = -\frac{1}{1-\alpha}log(\sum_{i=1}^{N} P_i^{\alpha})$$

- Hurst Exponent

Hurst Exponent is a measure of whether the data is pure white noise or it contains information. If H is equal to 0.5, then the time series is is purely random. However, if it is larger than 0.5, then it contains some trends. It is calculated for a given time series with length t from the rescaled range series (R/S) which is calculated from the standard deviation S and the range series R. Finally, a line fitting is done between log(R/S) and log(t) to get the Hurst exponent value [14].

- Modified Hurst Exponent

The Hurst exponent is the slope of the linear fit of the log-log graph. Another simpler implementation for the Hurst Exponent was using the below equation.

$$H = \frac{log(\frac{R}{S})}{log(T)}$$

where R is the maximum deviation from the mean and the minimum deviation from the mean, S is the standard deviation, $\frac{R}{S}$ is the rescaled value and T is the sample duration. In this implementation it is assumed that this linear fit will always pass through the origin.

- Fractal Dimension

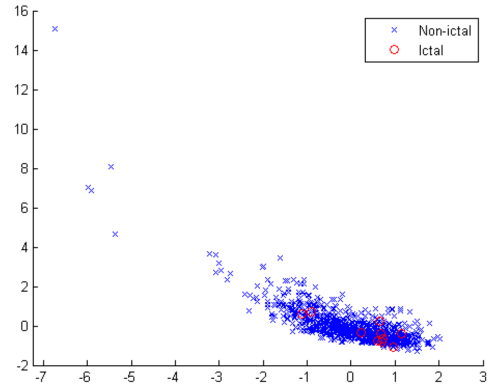
Fractal Dimension (FD) is based on fractal geometry. Higuchi's algorithm with k= 5 is used to calculate the fractal dimension [15].

III. Simulation Setup

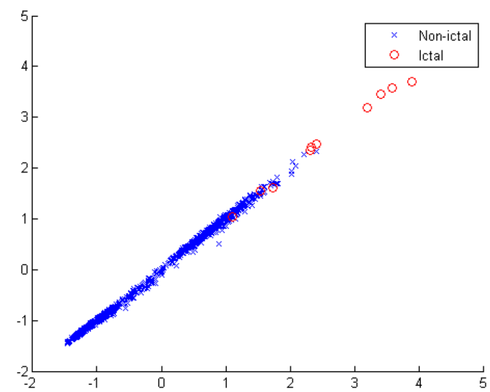
The dataset used was collected at the Children's Hospital Boston from subjects with intractable seizures. Recordings were collected from 22 subjects (5 males, and 17 females). The age of the subjects was from 3 to 22 in males and from 1.5 to 19 in females. The signals were sampled at 256 sample per second with 16-bit resolution. For each subject, 23 channels were recorded from different electrodes. The dataset comes with labeling on the epileptic sessions for different patients [16].

An implementation of all features and classifiers was done first using MATLAB2016a. After implementing the 20 linear and nonlinear features, different combinations of these features are used and tested along with linear kernel SVM. The performance metrics -sensitivity, specificity and accuracy- are extracted from each combination and compared.

The model could be built using two methods, either all-in then backward elimination according to p-value or try all possible combinations for a fixed number of features. In this work, the second solution was adopted. The decision was made to use three features in each combination based on many work done in the literature [5]. A total of 1140 combinations are tested and compared.



(a) Nonlinearly separable features



(b) Linearly separable features

Figure 2: Features data points

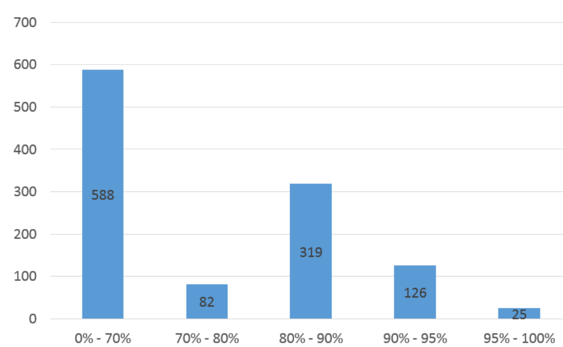


Figure 3: Features Combinations' sensitivity results

IV. Results and Discussion

As shown in Figure 2, Some features are linearly separable while others aren't.

Performance metrics were computed such as accuracy, specificity and sensitivity of classifier. The sensitivity is the

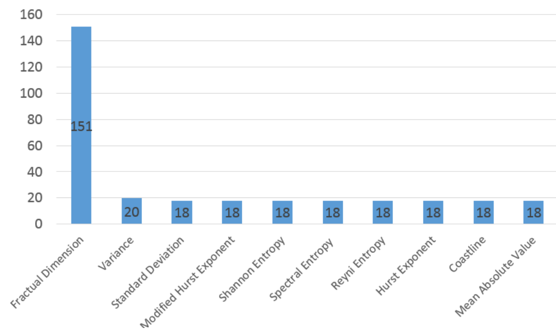


Figure 4: Most frequent features contribute in the best performing combinations

true positive rate and the specificity is the true negative rate, while accuracy is is as follows:

$$Sensitivity = \frac{TP}{TP + FN}$$

$$Specificity = \frac{TN}{TN + FP}$$

$$Accuracy = \frac{TP + TN}{TP + TP + FP + FN}$$

The implemented linear and non-linear functions are combined into combinations of three and used for seizure detection. The dataset is dominated by samples representing non seizure EEG and consequently the model is biased for detecting non seizure signals. Hence, the specificity and accuracy were higher than 96.4% for all combinations. That is why in the analysis, the sensitivity is the main point of interest as it varies from 0% up to 96.7% using SVM classifier with linear kernel. Using 20 features in total and grouping them in groups of 3, this results in 1140 combination to experiment. The 1140 combinations are classified into groups according to their sensitivity as shown in Figure 3. The results shows that out of these 1140 combinations, 25 combinations give sensitivity higher than 95%. Moreover, 126 combinations give sensitivities between 90% and 95 %. By more analysis of these combinations, the features that appeared in the best combinations are showed in descending order in Figure 4. It is obvious that Fractal Dimension is a very important feature in neural seizure detection. Also, the combination of Fractal dimension, Hurst exponent and Coastline is the best combination that gives sensitivity of 96.7%, specificity of 97.9% and accuracy of 97.9%.

V. Conclusion

Feature selection is a key metrics in enhancing the performance of SVM classifier. More than 1100 combinations, each consists of 3 features, are tested with linear kernel SVM. 126 combinations of them give sensitivity between 90 and 95%. 25 combinations of them give sensitivity more than 95%, While the specificity and accuracy are more than 96% for all combinations. Fractal dimension, Hurst exponent and coastline combination is the best combination.

Acknowledgment

This research was partially funded by ONE Lab at Cairo University, Zewail City of Science and Technology, and KAUST.

References

[1] A. H. Hassan, H. Mostafa, Y. Ismail, and A. M. Soliman, "A low-power high-efficiency inductive link power supply for neural recording and stimulation system-on-chip," *Journal of Low Power Electronics*, vol. 14, no. 1, pp. 129–139, 2018.

[2] S. Omar, H. Mostafa, T. Ismail, and S. Gabran, "Low-power implantable seizure detection processor," in *Electronics, Circuits, and Systems (ICECS), 2015 IEEE International Conference on*. IEEE, 2015, pp. 496–497.

[3] K. A. Helal, A. Y. A. Elmkarem, A.-M. B. Refaat, T. S. Kamel, K. A. Mohamed, M. M. Kamal, M. M. Abdelrahman, H. Mostafa, and Y. Ismail, "Low-power high-accuracy seizure detection algorithms for neural implantable platforms," in *Microelectronics (ICM), 2017 29th International Conference on*. IEEE, 2017, pp. 1–4.

[4] G. S. Maximous, A. M. El-Gunidy, H. Mostafa, T. Ismail, and S. Gabran, "A new sensitivity-specificity product-based automatic seizure detection algorithm," in *Electronics, Communications and Computers (JAC-ECC), 2017 Japan-Africa Conference on*. IEEE, 2017, pp. 107–110.

[5] Q. Yuan, W. Zhou, S. Li, and D. Cai, "Epileptic eeg classification based on extreme learning machine and nonlinear features," *Epilepsy research*, vol. 96, no. 1-2, pp. 29–38, 2011.

[6] M. Ashraf, H. Mostafa, and A. A. El-Adawy, "A low-power area-efficient design and comparative analysis for high-resolution neural data compression," in *Microelectronics (ICM), 2016 28th International Conference on*. IEEE, 2016, pp. 217–220.

[7] M. Alsenwi, M. Saeed, T. Ismail, H. Mostafa, and S. Gabran, "Hybrid compression technique with data segmentation for electroencephalography data," in *Microelectronics (ICM), 2017 29th International Conference on*. IEEE, 2017, pp. 1–4.

[8] T. Das, A. Ghosh, S. Guha, and P. Basak, "Classification of eeg signals for prediction of seizure using multi-feature extraction," in *Electronics, Materials Engineering and Nano-Technology (IEMENTech), 2017 1st International Conference on*. IEEE, 2017, pp. 1–4.

[9] T. Zhang and W. Chen, "Lmd based features for the automatic seizure detection of eeg signals using svm," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 25, no. 8, pp. 1100–1108, 2017.

[10] V. Sakkalis, *Modern Electroencephalographic Assessment Techniques*. Springer, 2015.

[11] S. Pincus, "Approximate entropy as an irregularity measure for financial data," *Econometric Reviews*, vol. 27, no. 4-6, pp. 329–362, 2008.

[12] S. M. Pincus, "Approximate entropy as a measure of system complexity." *Proceedings of the National Academy of Sciences*, vol. 88, no. 6, pp. 2297–2301, 1991.

[13] P. R. Pal, N. P. Mohanty, and T. Gandhi, "Entropy based detection & evaluation of epileptic seizure," *International Journal of Applied*, vol. 4, no. 1, pp. 73–77, 2011.

[14] V. Vijith, J. E. Jacob, T. Iype, K. Gopakumar, and D. G. Yohannan, "Epileptic seizure detection using non linear analysis of eeg," in *Inventive Computation Technologies (ICICT), International Conference on*, vol. 3. IEEE, 2016, pp. 1–6.

[15] T. Higuchi, "Approach to an irregular time series on the basis of the fractal theory," *Physica D: Nonlinear Phenomena*, vol. 31, no. 2, pp. 277–283, 1988.

[16] A. H. Shoeb, "Application of machine learning to epileptic seizure onset detection and treatment," Ph.D. dissertation, Massachusetts Institute of Technology, 2009.

[17] Y. Wang, Z. Li, L. Feng, H. Bai, and C. Wang, "Hardware design of multiclass svm classification for epilepsy and epileptic seizure detection," *IET Circuits, Devices & Systems*, vol. 12, no. 1, pp. 108–115, 2017.

[18] Q. Yuan, W. Zhou, Y. Liu, and J. Wang, "Epileptic seizure detection with linear and nonlinear features," *Epilepsy & Behavior*, vol. 24, no. 4, pp. 415–421, 2012.