

Diagnosing Epilepsy using Entropy Measures and Embedding Parameters of EEG Signals

Fatemeh Valipour (PhD Candidate)¹, Zahra Valipour (PhD Candidate)², Mani Garousi (MSc)³, Ali Khadem (PhD)^{1*}

ABSTRACT

Background: Manual analysis of electroencephalography (EEG) for epilepsy diagnosis can be subjective and time-consuming, leading to potential errors. An automatic classification system with high detection accuracy is essential for improving diagnostic efficiency and reliability.

Objective: This study aimed to evaluate a comprehensive set of entropy measures, along with embedding parameters, to identify the most effective single measure for epilepsy diagnosis.

Material and Methods: This analytical study used EEG data from the University of Bonn, including healthy controls (HCs) with open eyes and epileptic seizure patients, each with 100 single-channel segments. Discrete wavelet transform was applied, extracting ten entropy measures and two embedding parameters. Statistical tests evaluated feature significance, and a linear discriminant analysis (LDA) classifier was used for classification. Robustness was assessed by introducing Gaussian noise at varying signal-to-noise ratios (SNRs) and analyzing classification performance.

Results: Our findings indicated that embedding parameters, permutation entropy, fuzzy entropy, sample entropy, norm entropy, sure entropy, log entropy, and threshold entropy significantly differentiated epileptic patients from HCs. Among these, sample entropy, norm entropy, sure entropy, log entropy, threshold entropy, and embedding delay achieved classification accuracies between 97% and 100% using LDA classifier. Furthermore, even with substantial Gaussian noise, the classifier maintained an accuracy above 84%, demonstrating the robustness of these features in noisy conditions.

Conclusion: This study demonstrated that embedding-based and entropy-based features can serve as effective individual measures for discriminating epileptic EEG signals from HCs. These findings underscore the potential of these measures in automated epilepsy diagnosis systems, resulting in a robust and reliable tool for clinical applications.

Keywords

Epilepsy; Electroencephalography (EEG); Entropy; Embedding parameters; Classification; Signal to Noise Ratio (SNR)

Introduction

Epilepsy is a common neurological disorder caused by disruptions in brain electrophysiology, leading to recurrent, unpredictable seizures. These seizures can result in loss of awareness, whole-body convulsions, or, in severe cases, even death, highlighting the need

¹Department of Biomedical Engineering, Faculty of Electrical Engineering, K. N. Toosi University of Technology, Tehran, Iran

²Department of Biomedics and Biomedical Engineering, School of Advanced Technologies in Medicine, Isfahan University of Medical Sciences, Isfahan 81746-73461, Iran

³Department of Biomedical Engineering, Faculty of Electrical and Computer Engineering, Tarbiat Modares University, Tehran, Iran

*Corresponding author:
Ali Khadem
Department of Biomedical Engineering, Faculty of Electrical Engineering, K. N. Toosi University of Technology, Tehran, Iran
E-mail:
alikhadem@kntu.ac.ir

Received: 5 February 2024
Accepted: 9 March 2025

for precise evaluation [1]. Electroencephalography (EEG) is the primary noninvasive tool for detecting and monitoring neurological disorders, including epilepsy. It records brain electrical activity through scalp-mounted electrodes, each capturing the combined postsynaptic activity of numerous neurons. With its high temporal resolution, EEG provides direct insights into brain function and is widely used in clinical epileptology. However, visual EEG interpretation is subjective, time-consuming, and dependent on expert evaluation, making it prone to errors. Furthermore, subtle but clinically relevant EEG abnormalities may be overlooked [2,3]. To address these limitations, automated EEG-based epilepsy detection systems are essential for improving diagnostic accuracy and efficiency.

Distinguishing epileptic patients from healthy individuals fundamentally is a pattern recognition challenge, where feature extraction plays a crucial role in diagnostic accuracy. The quality and relevance of extracted features significantly impact classification performance. Existing literature categorizes features used in epilepsy diagnosis into several groups based on different analytical approaches, as follows: 1) time-domain features: statistical features [4,5], autoregressive (AR) parameter estimation [6,7], Burg's method [8], and time-domain power band features [9], 2) frequency-domain features: average variance of instantaneous frequencies [8], Fourier transform-based features [10], and higher order spectra [11]), 3) time-frequency domain features: wavelet variances [12, 13], relative wavelet energy [14,15], empirical mode decomposition (EMD) [16], EMD combined with discrete wavelet transform (DWT) [17], and multi-wavelet transform [18]), and 4) complexity features: Higuchi fractional dimension, Hurst exponent, approximate entropy, and sample entropy [11,18], spectral entropy [15,19], log-energy entropy [17,20], Sure entropy [20], embedding entropy, Kalmogorov—Sinai entropy, approximate

entropy [19], recurrence quantification analysis (RQA) measures [21,22], wavelet packet entropy [23], Reni entropy, Tallis entropy [24], Lyapunov exponent [25], fractional linear prediction [26], and permutation entropy [27]).

Biological signals, particularly EEG, are inherently nonlinear and non-periodic. As a result, traditional linear analysis methods, such as the Fast Fourier Transform (FFT), often fail to effectively distinguish between EEG signals from healthy individuals and those with neurological disorders. Therefore, nonlinear analysis techniques are essential for capturing the complex dynamics of EEG signals and improving classification accuracy [28].

State-space reconstruction is a dynamic analysis technique used for estimating embedding measures efficiently. It represents the underlying dynamics of a time series by reconstructing its phase space [21]. While phase-space reconstruction has been successfully applied to EEG analysis [29-31], it has not, to our knowledge, been specifically used to differentiate between epileptic and healthy EEG signals. Investigating the feasibility of this approach could provide valuable insights into the discriminative power of embedding parameters for epilepsy detection. Furthermore, although various entropy measures have been explored in epilepsy classification, no study, to our knowledge, has systematically compared the effectiveness of different entropy measures and classifiers using the same EEG dataset. Addressing this gap could enhance the reliability and accuracy of automated epilepsy detection.

In this study, we evaluate the accuracy of embedding delay, embedding dimension, and various entropy measures in distinguishing epileptic EEG signals from normal ones using a linear discriminant analysis (LDA) classifier, both in the presence and absence of noise.

Material and Methods

This study is an analytical research aimed at distinguishing epileptic EEG signals from normal ones. The following sections outline

the dataset and methodological approach.

EEG data

In this study, the EEG signals were collected by Andrzejak et al. [32] at the University of Bonn, Germany. All signals were recorded using a 128-channel amplifier system. The dataset consists of five EEG segment sets obtained from five healthy individuals under eyes-closed (Set Z) and eyes-open (Set O) conditions, as well as five epileptic patients in different seizure stages: interictal (Set N), ictal (Set F), and seizure (Set S). Each set contains 100 single-channel EEG segments. For this study, we selected two specific sets: healthy control group (Set O), EEG recordings from individuals with open eyes, epileptic seizure group (Set S), and EEG recordings from patients experiencing seizures.

Definition of selected features

State-Space Reconstruction

The behavior of nonlinear dynamic systems can be represented as a trajectory in phase space, where each point describes the system's state at a given instant (e.g., R). Phase space is a mathematical construct defined by the system's dynamic variables. If a system consists of n dynamic variables, its state at any given time is represented as a point in R^n -dimensional Euclidean space. As these variables evolve over time, the trajectory of this point forms an attractor, characterizing the system's dynamics.

To reconstruct the state-space representation, we utilized Takens' embedding theorem, a fundamental method in nonlinear time-series analysis. Given a time series $x(n)$; $n=1,2,\dots,N$, the state-space vectors are constructed using time-delayed embeddings, defined as:

$$\vec{X}_t = (x(t), x(t+\tau), x(t+2\tau), \dots, x(t+(m-1)\tau)) \quad t=1, \dots, N-(m-1)\tau \quad (1)$$

Where X_t represents the reconstructed state-space vector, m is the embedding dimension, and τ denotes the time delay.

A critical aspect of state-space reconstruc-

tion is determining the optimal values of τ and m . The ideal time delay (τ) ensures that each independent axis in the m -dimensional phase space retains the signal's information with minimal correlation between dimensions, preventing trajectory intersections [33]. Several methods exist for selecting τ and m . In this study, we determined time delay (τ) using the first minimum of average mutual information (AMI) and embedding dimension (m) using the false nearest neighbor (FNN) method [34].

Average Mutual Information (AMI)

Fraser and Swinney proposed that the first local minimum of the average mutual information (AMI) function provides the optimal time delay (τ) for state-space reconstruction [34].

AMI quantifies the predictable information shared between a time-series value and its delayed counterpart, measuring how much knowledge of a past value helps in predicting future values. The function is computed for different values of τ as follows:

$$I(\tau) = \sum_{n=1}^{N-\tau} P(x(n), x(n+\tau)) \times \log\left(\frac{P(x(n), x(n+\tau))}{P(x(n)) \times P(x(n+\tau))}\right) \quad (2)$$

Where $P(x_i, x_{i+\tau})$ represents the joint probability distribution of values at time i and $i+\tau$, and $P(x_i)$, $P(x_{i+\tau})$ are the marginal probability distributions. The optimal time delay (τ) is selected at the first minimum of $I(\tau)$, ensuring minimal redundancy while preserving the system's dynamics.

False Nearest Neighbor (FNN)

The False Nearest Neighbors (FNN) method determines the optimal embedding dimension (m) by analyzing the behavior of point distances in phase space as the dimensionality increases [34].

In this method, the distance between two points in phase space is examined as the spatial dimension D increases to $D+1$. If the distance between two neighboring points in dimension D significantly changes when projected into dimension $D+1$, the points are classified as false neighbors. This indicates that the embedding dimension D is insufficient to properly reconstruct the system's dynamics.

The distances between a point $X(t)$ and its r th nearest neighbor $X(t_r)$ in dimensions D and $D+1$ are estimated as follows:

$$R_D^2(t, r) = \|X(t) - X(t_r)\|^2 = \sum_{k=0}^{D-1} [x_{t-k\tau} - x_{t_r-k\tau}]^2 \quad (3)$$

$$R_{D+1}^2(t, r) = R_D^2(t, r) + [x_{t-D\tau} - x_{t_r-D\tau}]^2 \quad (4)$$

Where R_D and R_{D+1} represent the Euclidean distances in dimensions D and $D+1$, respectively. If the ratio exceeds a predefined threshold (R_{tot}), the points are considered false neighbors, indicating the need for a higher embedding dimension. The optimal embedding dimension (m) is the smallest D , where the fraction of false neighbors approaches zero, ensuring a well-reconstructed phase space (see equation 5).

$$\left[\frac{R_{D+1}^2(t, r) - R_D^2(t, r)}{R_D^2(t, r)} \right]^{\frac{1}{2}} > R_{tot} \quad (5)$$

Entropy

Entropy is a nonlinear measure that quantifies the complexity of a signal. A decrease in entropy indicates a more regular time series, suggesting a higher information rate within the signal [35].

Among various entropy measures, approximate entropy (AppEn), sample entropy (SampEn), and fuzzy entropy (FuzzyEn) are widely used. These methods estimate the predictability of a signal by analyzing the conditional probability of similarity between sequences. Specifically, if two sequences in a time series remain similar for m data points, these entropy measures evaluate the likelihood that they will also be similar at $m+1$ points.

These entropy metrics provide valuable insights into the underlying dynamics of EEG signals, aiding in distinguishing normal from pathological patterns. The following section provides a brief introduction to ten types of entropy measures for signal complexity analysis:

Approximate entropy (AppEn)

To identify specific patterns within a time series, approximate entropy (AppEn) establishes a relationship between probabilities, measur-

ing the degree of similarity between different segments of the signal. This similarity is determined using a tolerance threshold (r), allowing for the quantification of signal complexity [36]. AppEn is calculated as follows:

$$C_i^m(r) = n_i^m(r) / (N - m + 1) \quad (6)$$

$$\varphi^m(n) = (N - m + 1)^{-1} \sum_{i=1}^{N-m+1} \ln C_i^m(r) \quad (7)$$

$$ApEn(m, r) = \varphi^m(r) - \varphi^{m+1}(r) \quad (8)$$

Where $n_i^m(r)$ represents the number of m -dimensional similar patterns whose pairwise distance is less than r .

Sample Entropy (SampEn)

Sample Entropy (SampEn) is an improved version of AppEn that addresses its limitations. Unlike AppEn, SampEn is less sensitive to signal length, disregards self-matches, and provides an unbiased estimate of complexity, making it a more reliable measure [37]. SampEn is defined as:

$$SampEn(m, r, N) = -\ln(C^{m+1}(r) / C^m(r)) \quad (9)$$

Where $C^m(r)$ represents the number of m -point sequences that remain similar within a distance less than r .

Fuzzy Entropy (FuzzyEn)

Fuzzy Entropy (FuzzyEn) is a relatively recent method for quantifying the fuzziness and uncertainty in a time series by defining the similarity between vectors using a fuzzy approach. Unlike some traditional entropy measures, FuzzyEn is independent of data length, making it a robust complexity measure [38]. FuzzyEn is computed as follows:

$$FEn(m, n, r, N) = \ln(\varphi^m(n, r) - \varphi^{m+1}(n, r)) \quad (10)$$

$$FEn(m, n, r, N) = \ln(\varphi^m(n, r) - \varphi^{m+1}(n, r)) \quad (11)$$

Where D_{ij}^m identifies the similarity between two vectors with a distance less than r , embedding dimension m , and gradient boundary n .

In this study, the parameter values were set as follows: $n=2$, $r=0.15$ of the standard deviation of the time series, and $m=2$.

Shannon entropy (ShanEn)

This entropy quantifies a set of relational

parameters that vary linearly with the logarithm of the number of probabilities. It is mathematically expressed as:

$$ShanEn(x) = -\sum_k P(x_k) \times \log_2(P(x_k)) \quad 0 < M < N-1 \quad (12)$$

Where $P(x_k)$ represents the probability of x_k , and M denotes the number of levels of the discrete-valued random variable X .

Spectral entropy (SEn)

Spectral Entropy (SEn) is a normalized form of Shannon entropy that evaluates the spectral complexity of a signal by analyzing the amplitude components of its power spectrum. It provides insights into the distribution of frequency components within the signal [3]. SpecEn is computed as follows:

$$SEn = -\sum_k P(f_k) \times \log(P(f_k)) \quad 1 < k < M \quad (13)$$

Where $P(f_k)$ represents the normalized power spectral density at frequency f_k , and M is the total number of frequency components.

Permutation Entropy (PermEn)

Permutation entropy (PermEn) quantifies the complexity of a time series by identifying couplings between successive data points, capturing the presence or absence of specific permutation patterns in the signal. Given a time series x with an embedding dimension of m and time delay of τ , the reconstructed sequence is defined as follows [39]:

$$X(N-(m-1)\tau) = \{x(N-(m-1)\tau), x(N-(m-2)\tau), \dots, x(N)\} \quad (14)$$

PermEn is given by:

$$PermEn = -\sum_{j=1}^n p_j \log_2 p_j \quad (15)$$

Where p_j represents frequency related with each possible sequence pattern, and n denotes the permutation order, with $n \geq 2$.

PermEn is an effective measure for analyzing nonstationary, nonlinear, and chaotic time series, even in the presence of dynamical noise. It demonstrates robustness and computational efficiency, producing reliable results with minimal sensitivity to noise. Due to its low computational complexity, PermEn is particularly well-suited for the analysis of large datasets, making it a valuable tool in various signal processing applications [39].

In this study, we selected an embedding dimension of 3 and a time delay of 1 to effectively capture the temporal dynamics of the signal.

Furthermore, wavelet packet decomposition was applied to compute the following five entropy measures, which are defined as follows [40]:

Norm entropy (NormEn)

$$NormEn = \sum_i |x_i|^P, P \geq 1 \quad (16)$$

Threshold entropy (ThreshEn)

$$ThreshEn = \{i \text{ such that } |x_i| \geq P\} \quad (P > 0) \quad (17)$$

ThreshEn quantifies the number of time instants, at which the signal amplitude exceeds a predefined threshold. In this study, the threshold value was set to 0.2.

Sure entropy (SureEn)

$$E_{SEn} = N - \#\{i \text{ such that } |x_i| \leq \varepsilon\} + \sum_i \min(x_i^2, \varepsilon^2) \quad (18)$$

Log Energy entropy (LogEn)

$$E_{LgEn} = \sum_{i=1}^N \log(x_i^2) \quad (19)$$

Where x represents the signal, and x_i denotes the coefficients of x in the orthonormal basis. Additionally, p , P , and N correspond to power, the threshold value, and the signal length, respectively.

Classification

Linear Discriminant Analysis (LDA)

Linear Discriminant Analysis (LDA) is a supervised classification technique that projects data onto a new feature space by maximizing the separation between predefined groups. It achieves this by transforming the original predictor variables into a single discriminant variable that maximizes the variance between classes while minimizing the variance within each class.

LDA assumes that the independent variables follow a normal distribution and that different classes share a common covariance structure. The algorithm calculates the mean vector for each class and assigns a new observation to the class whose mean vector is closest in the

discriminant space. By ensuring the greatest possible separation between class means, LDA enhances classification performance in a statistically optimal manner [41].

Proposed method

Figure 1 indicates a flowchart outlining the analytical framework of this study.

In this study, EEG datasets from two groups, healthy individuals with open eyes and patients with epilepsy, were preprocessed using a Butterworth filter with a cutoff frequency range of 0.5 to 40 Hz to remove unwanted noise and artifacts. The signals were then decomposed into frequency sub-bands using Discrete Wavelet Transform (DWT), a linear time-frequency analysis method particularly well-suited for nonstationary signals like EEG due to its high resolution in both time and frequency domains.

DWT was applied to decompose the EEG signals into five levels, utilizing the Daubechies 4 (Db4) wavelet filter. This decomposition produced approximation and detail coefficients at each level, which were then used to extract relevant features. The extracted features included the embedding dimension and embedding delay of the state space, along with ten types of entropy measures, namely: AppEn, SampEn, ShanEn, FuzzyEn, SEn, PermEn, ThreshEn, NormEn, SureEn, and LogEn.

Subsequently, statistical analysis was performed to evaluate the significance of differences in the extracted features between the two groups. The Mann-Whitney U test was

employed to compare the *P*-values between the epileptic and healthy groups, assessing the statistical significance of each feature. Additionally, the mean and standard deviation of the extracted features were computed for both groups to facilitate a comparative analysis.

The classification of EEG signals was performed using an LDA classifier. To enhance robustness and prevent overfitting, *K*-fold cross-validation (*K*=10) was applied. The classifier's performance was evaluated using key metrics, including classification accuracy, specificity, and sensitivity.

Finally, to assess the robustness of the features that achieved the highest accuracy on clean EEG data, Gaussian noise was introduced into the signals. The classification performance was then analyzed across a range of signal-to-noise ratios (SNRs) from 1 to 40 dB, identifying the features that maintained their accuracy under varying noise conditions.

Results

The statistical test results for embedding parameters and the ten predefined entropy measures comparing the healthy and epileptic groups are summarized in Table 1. For each feature, the mean and standard deviation were computed for both groups. The results revealed that the healthy group exhibited lower values for embedding parameters compared to the epileptic group, whereas entropy measures were generally higher in the healthy group.

The analysis further demonstrated that Embedding Parameters, PermEn, FuzzyEn,

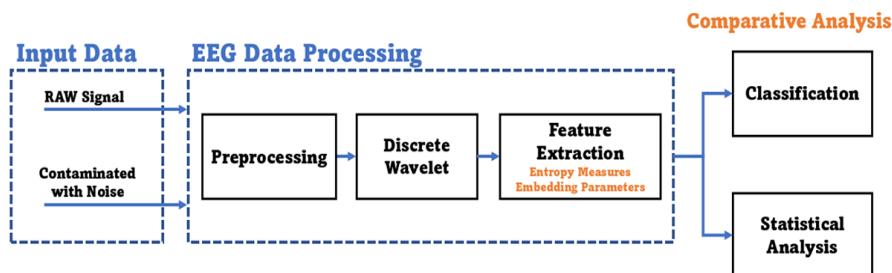


Figure 1: Block diagram of the proposed method for distinguishing epileptic electroencephalography (EEG) signals from normal EEG data.

SampEn, NormEn, SureEn, LogEn, and ThreshEn exhibited statistically significant differences between the two groups. The *P*-values for these features indicated their potential for significant differentiation (*P*-value<0.05), suggesting their relevance in distinguishing epileptic EEG signals from normal EEG activity.

The performance metrics of the LDA classifier are summarized in Table 2. These results were obtained by investigating each feature individually. The LDA classifier achieved 100%

accuracy for SampEn, LogEn, and ThreshEn, and over 97% accuracy for SureEn, NormEn, and Embedding Delay.

Comparative analysis

The performance comparison of this study with existing research on epileptic seizure detection is summarized in Table 3. Table 3 provides an overview of various methodologies, datasets, classifiers, and extracted features used in previous studies, highlighting how our approach compares in terms of accuracy,

Table 1: Mean and Standard Deviation (SD) of entropy measures and embedding parameters extracted from the healthy and epileptic patient groups.

Features	Healthy		Patient		<i>P</i> -Value
	Mean	SD	Mean	SD	
Approximate Entropy	0.89	0.10	0.90	0.11	1.02×10^{-1}
Sample Entropy	0.06	0.01	0.00	0.00	2.8×10^{-2}
Permutation Entropy	1.37	0.04	1.04	0.09	4×10^{-3}
Fuzzy Entropy	0.03	0.01	0.00	0.00	2.56×10^{-3}
Shannon Entropy	-5.9×10^{-8}	2×10^{-8}	-6×10^{-8}	5×10^{-9}	5.76×10^{-2}
Spectral Entropy	0.77	0.00	0.76	0.01	4.8×10^{-2}
Norm Entropy	7.7×10^{-3}	2×10^{-1}	6.17×10^{-2}	0.1	1.81×10^{-5}
Threshold Entropy	5.7×10^{-5}	3.4×10^{-3}	1.7×10^{-3}	0.26	9.01×10^{-6}
Log Entropy	9.3×10^{-5}	1.6×10^{-3}	6.7×10^{-3}	0.19	3.27×10^{-6}
Sure Entropy	8.4×10^{-2}	2.4×10^{-1}	9.37×10^{-1}	0.21	2.76×10^{-5}
Embedding delay	6.90	1.04	1.25×10^{-1}	6.75	4.48×10^{-24}
Embedding dimension	7.14	0.53	8.02	1.05	2.21×10^{-4}

Table 2: Performance of the linear discriminant analysis (LDA) classifier for entropy measures and embedding parameters extracted from the healthy and epileptic patient groups.

Performance criteria	AppEn	SampEn	PermEn	FuzzyEn	ShanEn	SpectralEn	NormEn	ThreshEn	LogEn	SureEn	Embedding dimension	Embedding Delay
Accuracy	86.11	100	81.67	90.56	77.78	83.33	98.89	100	100	98.89	95.85	97.78
Specificity	82.09	100	80.93	92.01	74.44	86.31	98.06	100	100	97.78	96.91	95.31
Sensitivity	90.35	100	82.93	89.74	82.51	90.72	100	100	100	100	94.79	100

Table 3: Comparison of the performance of our work with previous studies on epilepsy diagnosis using electroencephalography (EEG) signals.

Author	Feature	Classifier	EEG dataset	Accuracy (%)	Specificity (%)	Sensitivity (%)
Subasi et al. [4]	SD and average value	LR, MLPNN	[4]	-----	90.3, 91.4	89.2, 92.8
Subasi et al. [5]	Statistical features with ICA, PCA, LDA	SVM	University of Bonn [32]	99.50, 98.75, 100	99, 98.5, 100	100, 99, 100
Subasi et al. [6]	AR parameter estimation and maximum likelihood estimation	Wavelet neural networks and back propagation	[6]	93	92.4	93.6
Ubeyli et al. [7]	AR method	SVM	University of Bonn [32]	99.56	99.63	99.50
Faust et al. [8]	Frequency domain parameters, Burg's method	SVM	University of Bonn [32]	93.3	98.33	96.67
Donos et al. [9]	Time-domain and power band features	Random forest	EMU [42]	-----	-----	86.27
Polat et al. [10]	Fourier transform-based features	Decision tree	University of Bonn [32]	98.72	99.31	99.40
Acharya et al. [11]	HOS+Higuchi FD+Hurst EXPONENT+AppEn+SampEn	DT, PNN, KNN, Fuzzy, GMM, SVM	University of Bonn [32]	97.3, 98.1, 98.1, 100, 99, 99	94, 96, 96, 100, 98, 98	99, 99, 99, 100, 100, 99.5
Shengkuh et al. [12]	Wavelet variances	KNN	University of Bonn [32]	100	-----	-----
Orhan et al. [13]	Wavelet-based features	K-means clustering and MLP	University of Bonn [32]	96.67	97.98	94.12
Mortaga et al. [14]	Relative Wavelet Energy	ANN	University of Bonn [32]	95.2	92.12	98.17
Jia et al. [16]	Statistical features in the CEEMD domain	Random forest	University of Bonn [32]	98	99	100
Gandhi et al. [15]	DWT+(Spectral entropy, Energy)	PNN	University of Bonn [32]	100	-----	-----
Das et al. [17]	EMD-DWT method, log-energy entropy	KNN (Cityblock distance)	University of Bonn [32]	89.4	88.1	90.7
Guo et al. [18]	Multi wavelet transform+ ApproximateEn	MLPNN	University of Bonn [32]	98.2	95.50	99.00
Chandaka et al. [43]	Cross-correlation	SVM	University of Bonn [32]	95.96	100	92
Gupta et al. [20]	Cross corrEn, log energy En, SureEn	Least square-SVM (RBF kernel), KNN (Euclidean distance)	Bern Barcelona database [44]	94.41, 93.12	95.57, 95.15	93.25, 91.09
Kannathal et al. [19]	SpectralEn, EmbeddingEn, Kalmogorov—SinaiEn, ApproximateEn	ANFIS	University of Bonn [32]	92.2	-----	-----
Gruszczyńska et al. [21]	RQA measures	SVM	Medical University of Białystok [45]	86.8	-----	-----
Acharya et al. [22]	RQA measures	GMM, KNN	University of Bonn [32]	92.6, 95.2	92.2, 98.9	97.2, 98.3
Wang et al. [23]	Wavelet packet entropy	KNN	University of Bonn [32]	99.44	-----	-----

Author	Feature	Classifier	EEG dataset	Accuracy (%)	Specificity (%)	Sensitivity (%)
Redilico et al. [24]	ReniEn, TallisEn	LR	University of Bonn [32]	95, 94.5	94, 94	97, 97
Djemili et al. [25]	Lyapunov exponent	RNN	University of Bonn [32]	96.79	-----	-----
Joshi et al. [26]	Fractional linear prediction	SVM (RBF kernel)	University of Bonn [32]	96	95	95.33
Veisi et al. [27]	Permutation entropy	LDA	University of Bonn [32]	97	-----	-----
Polychronaki et al. [46]	Fractal dimension	KNN	Epilepsy Telemetry Unit, Department of Neurosurgery, University of Athens, 'Evangelismos' Hospital	-----	-----	100
Chua et al. [47]	Higher order statistics based features	GMM	University of Bonn [32]	93.1	92	97.67
Acharya et al. [48]		Deep convolutional neural network	University of Bonn [32]	88.67	90	95
Our study	LogEn, ThreshEn, SampEn, NormEn, SureEn, Embedding delay	LDA	University of Bonn [32]	100, 100, 100, 98.89, 98.89, 97.78	100, 100, 100, 98.06, 97.78, 95.31	100, 100, 100, 100, 100, 100

LR: Logistic Regression, MLPNN: Multilayer Perception Neural Networks, AR: Autoregressive, ANN: Artificial Neural Network, k-NN: K-Nearest Neighbor, LS: Least Squares, SVM: Support Vector Machine, GLM: Generalized Linear Model, HOS: Higher Order Spectra, PNN: Probabilistic Neural Network, ASE: Average Sample Entropy, AVIF: Average Variance Of Instantaneous Frequencies, RQA: Recurrence Quantification Analysis, RNN: Recurrent Neural Networks, GMM: Gaussian Mixture Model

sensitivity, and specificity.

Figure 2 illustrates the LDA classifier's performance for SampEn, LogEn, ThreshEn, SureEn, NormEn, and Embedding Delay across a range of signal-to-noise ratios (SNRs), from 1 to 40 dB. These features maintained an accuracy of over 84% for SNRs greater than 20 dB.

Discussion

Our results showed that measures of Embedding dimension, Embedding delay, PermEn, FuzzyEn, SampEn, NormEn, SureEn, LogEn, and ThreshEn effectively discriminate epileptic patients from normal subjects. The significant differentiation of EEG signals using these features can be attributed to the following factors:

Embedding Parameters

These parameters play a crucial role in phase space reconstruction, where the embedding dimension represents the minimum number of uncorrelated orientations necessary to re-

construct the system dynamics. While higher embedding dimensions can capture more information, excessive embedding introduces redundancy.

Two methods, AMI and FNN, are used to calculate the embedding delay and embedding dimension. Compared to singular value decomposition (SVD)-based approaches, the AMI method is more effective in capturing nonlinear interrelations, ensuring that the reconstructed state space consists of uncorrelated orientations [49].

As shown in Equation (1), a greater number of delayed time series contribute to the state-space vector for epileptic EEG signals, resulting in higher embedding parameter values in epileptic EEG compared to normal EEG, reflecting the transition from randomness to deterministic chaos during seizures [50] (see Table 1).

Entropy Measures

Entropy, a measure of signal complexity, is

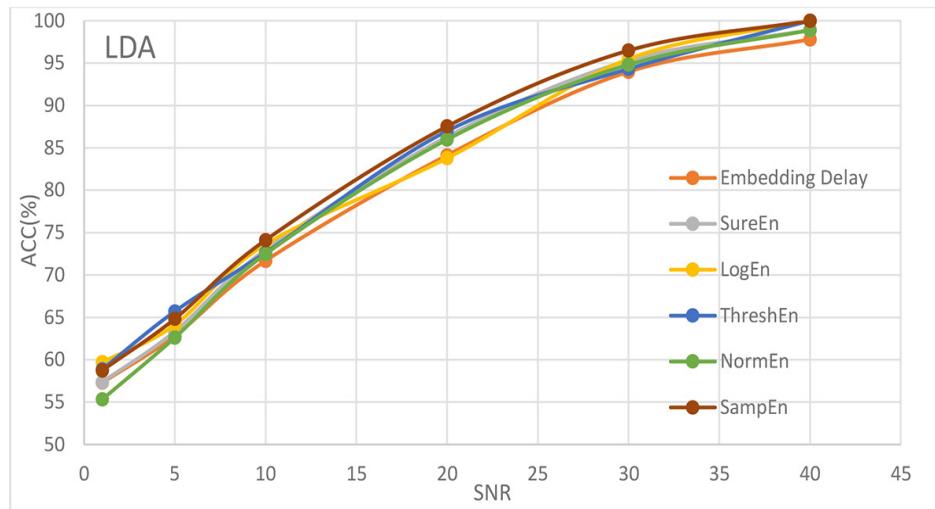


Figure 2: Linear discriminant analysis (LDA) classifier performance across a range of signal-to-noise ratios (SNRs), from 1 to 40 dB, using SampEn, LogEn, ThreshEn, SureEn, NormEn, and Embedding delay for EEG signal classification.

generally lower in epileptic EEGs compared to the healthy group due to the presence of more rhythmic and periodic patterns during seizures [50].

AppEn, SampEn, and FuzzyEn are particularly effective in characterizing nonlinear signals. Since epileptic EEG signals exhibit greater periodicity, they tend to have lower entropy values compared to normal EEG signals [50]. While AppEn is prone to bias and is highly sensitive to minor fluctuations [51], SampEn and FuzzyEn offer greater precision and robustness in capturing signal complexity [52].

PermEn and other entropy measures, including SpectralEn, NormEn, SureEn, LogEn, and ThreshEn are effective in detecting variability in nonstationary signals [53]. However, ShanEn has notable limitations, such as the potential overestimation of entropy and its inability to capture temporal dependencies in the signal. Given the increased predictability of epileptic EEG signals, these signals typically exhibit lower entropy values compared to normal EEGs [54].

Feature Dimensionality

In this study, all extracted features were

individually analyzed to assess their discriminatory power. The dimensionality of the feature space poses a significant challenge for classification algorithms. A higher-dimensional feature space increases model complexity, leading to greater computational costs for both training and testing. Additionally, an excessive number of features can introduce redundancy, which may degrade the estimation accuracy of model parameters. If a single feature can provide reliable classification performance, the need for multiple features is reduced, simplifying the model and enhancing computational efficiency.

Classification Performance

The LDA classifier demonstrated high classification accuracy when using individual features, particularly SampEn, NormEn, SureEn, LogEn, ThreshEn, and Embedding Delay. This finding indicates that these features possess strong discriminatory power, enabling the classification of EEG signals independently, even with a simple linear classifier like LDA.

These features effectively distinguished epileptic patients from healthy individuals, achieving reliable classification accuracy. Furthermore, our findings align with previous

studies, which have also identified significant alterations in these features among epileptic patients.

Robustness to Noise

This study further highlighted the robustness of the LDA classifier in distinguishing healthy and epileptic EEG signals, even in the presence of Gaussian noise. As illustrated in Figure 2, classification accuracy improves across the selected measures (SampEn, NormEn, SureEn, LogEn, ThreshEn, and Embedding Delay) as the SNR increases.

At low SNRs (e.g., 1 to 10 dB), the accuracy is generally lower, indicating that high noise levels negatively impact the classifiers' ability to distinguish between healthy and epileptic EEG signals. However, at SNR levels of 20 dB and above, significant improvements are observed in accuracy. Notably, SampEn and SureEn demonstrate strong performance at higher SNRs, with SampEn reaching over 87% accuracy at 20 dB, which further improves at higher SNRs. Additionally, ThreshEn and LogEn demonstrate competitive classification performance, making these measures particularly suitable for EEG signal analysis in noisy environments. These findings emphasize the resilience of the selected entropy measures in preserving classification accuracy under varying noise conditions.

Conclusion

In this study, embedding parameters and entropy measures from each wavelet sub-band were individually fed into the LDA classifier to classify EEG signals into two groups: healthy individuals and epileptic patients. By comparing the sensitivity, specificity, and accuracy of the classifier, it yielded reliable results, effectively discriminating the EEG signals of epileptic patients from those of normal subjects. Additionally, the effect of additive Gaussian noise on discrimination performance was evaluated, demonstrating that certain measures, such as SampEn, LogEn, ThreshEn, SureEn, NormEn, and Embedding

delay, maintained high classification accuracy even under varying noise conditions, thereby confirming the robustness of these features in noisy environments. Notably, if a single feature can guarantee reliable classification performance, it avoids the challenges associated with high-dimensional feature vectors, such as increased model complexity, time consumption, and redundancy. This analysis provides a valuable framework for the quantification of classification reliability and the identification of abnormal EEG activity. Future research could explore the generalizability of these findings across different datasets and noise models.

Acknowledgment

We would like to appreciate Anderzejak et al. from the University of Bonn, Germany for making the raw EEG data accessible. Finally, we thank Ali Pouresmaeil and Fatemeh Hasanzadeh for their comments which greatly improved the manuscript.

Authors' Contribution

Z. Valipour and M. Garousi contributed equally to this work. Data analysis and methodology: F. Valipour, Z. Valipour, and M. Garousi; conceptualization: F. Valipour, Z. Valipour, and M. Garousi; investigation: M. Garousi and Z. Valipour; writing – original draft: F. Valipour, Z. Valipour, and M. Garousi; writing – review & editing: A. Khadem; supervised development of work: A. Khadem; data interpretation: A. Khadem and F. Valipour; manuscript evaluation: A. Khadem. All of the authors have read, made modifications, and given approval for the final version of the manuscript.

Ethical Approval

The dataset utilized in this study is publicly available and was originally collected by Andrzejak et al. at the University of Bonn, Germany. The use of this data for research purposes was pre-approved by the original

authors. Additionally, a preprint of this study is archived at <https://arxiv.org/abs/2401.07258>.

Funding

This study received no specific funding from funding organizations in the public, commercial, or non-profit sectors.

Conflict of Interest

None

References

1. Hesdorffer DC, Beck V, Begley CE, Bishop ML, Cushner-Weinstein S, et al. Research implications of the Institute of Medicine Report, Epilepsy Across the Spectrum: Promoting Health and Understanding. *Epilepsia*. 2013;54(2):207-16. doi: 10.1111/epi.12056. PubMed PMID: PMC3566357.
2. Sivasankari N, Thanushkodi K. Automated epileptic seizure detection in EEG signals using FastICA and neural network. *Int J Adv Soft Comput Appl*. 2009;1(2):91-104.
3. Fell J, Röschke J, Mann K, Schäffner C. Discrimination of sleep stages: a comparison between spectral and nonlinear EEG measures. *Electroencephalogr Clin Neurophysiol*. 1996;98(5):401-10. doi: 10.1016/0013-4694(96)95636-9. PubMed PMID: 8647043.
4. Subasi A, Erçelebi E. Classification of EEG signals using neural network and logistic regression. *Comput Methods Programs Biomed*. 2005;78(2):87-99. doi: 10.1016/j.cmpb.2004.10.009. PubMed PMID: 15848265.
5. Subasi A, Gursoy MI. EEG signal classification using PCA, ICA, LDA and support vector machines. *Expert Systems with Applications*. 2010;37(12):8659-66. doi: 10.1016/j.eswa.2010.06.065.
6. Subasi A, Alkan A, Koklukaya E, Kiymik MK. Wavelet neural network classification of EEG signals by using AR model with MLE preprocessing. *Neural Netw*. 2005;18(7):985-97. doi: 10.1016/j.neunet.2005.01.006. PubMed PMID: 15921885.
7. Übeyli ED. Least squares support vector machine employing model-based methods coefficients for analysis of EEG signals. *Expert Systems with Applications*. 2010;37(1):233-9. doi: 10.1016/j.eswa.2009.05.012.
8. Faust O, Acharya UR, Min LC, Suthu BH. Automatic identification of epileptic and background EEG signals using frequency domain parameters. *Int J Neural Syst*. 2010;20(2):159-76. doi: 10.1142/S0129065710002334. PubMed PMID: 20411598.
9. Donos C, Dümpelmann M, Schulze-Bonhage A. Early Seizure Detection Algorithm Based on Intra-cranial EEG and Random Forest Classification. *Int J Neural Syst*. 2015;25(5):1550023. doi: 10.1142/S0129065715500239. PubMed PMID: 26022388.
10. Polat K, Güneş S. Classification of epileptiform EEG using a hybrid system based on decision tree classifier and fast Fourier transform. *Applied Mathematics and Computation*. 2007;187(2):1017-26. doi: 10.1016/j.amc.2006.09.022.
11. Acharya UR, Sree SV, Alvin AP, Yanti R, Suri JS. Application of non-linear and wavelet based features for the automated identification of epileptic EEG signals. *Int J Neural Syst*. 2012;22(2):1250002. doi: 10.1142/S0129065712500025. PubMed PMID: 23627588.
12. Xie S, Krishnan S. Wavelet-based sparse functional linear model with applications to EEGs seizure detection and epilepsy diagnosis. *Med Biol Eng Comput*. 2013;51(1-2):49-60. doi: 10.1007/s11517-012-0967-8. PubMed PMID: 23054383.
13. Orhan U, Hekim M, Ozer M. EEG signals classification using the K-means clustering and a multilayer perceptron neural network model. *Expert Systems with Applications*. 2011;38(10):13475-81. doi: 10.1016/j.eswa.2011.04.149.
14. Mortaga M, Brenner A, Kutafina E. Towards Interpretable Machine Learning in EEG Analysis. *Stud Health Technol Inform*. 2021;283:32-8. doi: 10.3233/SHTI210538. PubMed PMID: 34545817.
15. Gandhi TK, Chakraborty P, Roy GG, Panigrahi BK. Discrete harmony search based expert model for epileptic seizure detection in electroencephalography. *Expert Systems with Applications*. 2012;39(4):4055-62. doi: 10.1016/j.eswa.2011.09.093.
16. Jia J, Goparaju B, Song J, Zhang R, Westover MB. Automated identification of epileptic seizures in EEG signals based on phase space representation and statistical features in the CEEMD domain. *Biomedical Signal Processing and Control*. 2017;38:148-57. doi: 10.1016/j.bspc.2017.05.015.
17. Das AB, Bhuiyan MI. Discrimination and classification of focal and non-focal EEG signals using entropy-based features in the EMD-DWT domain. *Biomedical Signal Processing and Control*. 2016;29:11-21. doi: 10.1016/j.bspc.2016.05.004.
18. Guo L, Rivero D, Pazos A. Epileptic seizure detection using multiwavelet transform based approximate entropy and artificial neural networks. *J Neurosci Methods*. 2010;193(1):156-63. doi: 10.1016/j.jneumeth.2010.08.030. PubMed PMID: 21030300.

- 20817036.
19. Kannathal N, Choo ML, Acharya UR, Sadasivan PK. Entropies for detection of epilepsy in EEG. *Comput Methods Programs Biomed*. 2005;80(3):187-94. doi: 10.1016/j.cmpb.2005.06.012. PubMed PMID: 16219385.
 20. Gupta V, Priya T, Yadav AK, Pachori RB, Acharya UR. Automated detection of focal EEG signals using features extracted from flexible analytic wavelet transform. *Pattern Recognition Letters*. 2017;94:180-8. doi: 10.1016/j.patrec.2017.03.017.
 21. Gruszczyńska I, Mosdorff R, Sobaniec P, Żochowska-Sobaniec M, Borowska M. Epilepsy identification based on EEG signal using RQA method. *Adv Med Sci*. 2019;64(1):58-64. doi: 10.1016/j.advms.2018.08.003. PubMed PMID: 30476729.
 22. Acharya UR, Sree SV, Chattopadhyay S, Yu W, Ang PC. Application of recurrence quantification analysis for the automated identification of epileptic EEG signals. *Int J Neural Syst*. 2011;21(3):199-211. doi: 10.1142/S0129065711002808. PubMed PMID: 21656923.
 23. Wang D, Miao D, Xie C. Best basis-based wavelet packet entropy feature extraction and hierarchical EEG classification for epileptic detection. *Expert Systems with Applications*. 2011;38(11):14314-20. doi: 10.1016/j.eswa.2011.05.096.
 24. Redelico FO, Traversaro F, García MD, Silva W, Rosso OA, Risk M. Classification of normal and pre-ictal eeg signals using permutation entropies and a generalized linear model as a classifier. *Entropy*. 2017;19(2):72. doi: 10.3390/e19020072.
 25. Djemili R, Djemili I. Nonlinear and chaos features over EMD/VMD decomposition methods for ictal EEG signals detection. *Comput Methods Biomech Biomed Engin*. 2024;27(15):2091-110. doi: 10.1080/10255842.2023.2271603. PubMed PMID: 37861376.
 26. Joshi V, Pachori RB, Vijesh A. Classification of ictal and seizure-free EEG signals using fractional linear prediction. *Biomedical Signal Processing and Control*. 2014;9:1-5. doi: 10.1016/j.bspc.2013.08.006.
 27. Veisi I, Pariz N, Karimpour A. Fast and robust detection of epilepsy in noisy EEG signals using permutation entropy. In 7th International Symposium on Bioinformatics and BioEngineering; Boston, MA, USA: IEEE; 2007. p. 200-3.
 28. Gil LM, Nunes TP, Silva FH, Faria AC, Melo PL. Analysis of human tremor in patients with Parkinson disease using entropy measures of signal complexity. *Annu Int Conf IEEE Eng Med Biol Soc*. 2010;2010:2786-9. doi: 10.1109/IEMBS.2010.5626365. PubMed PMID: 21095968.
 29. Jacob JE, Cherian A, Gopakumar K, Iype T, Yohannan DG, Divya KP. Can Chaotic Analysis of Electroencephalogram Aid the Diagnosis of Encephalopathy? *Neurol Res Int*. 2018;2018:8192820. doi: 10.1155/2018/8192820. PubMed PMID: 30002930. PubMed PMCID: PMC5996471.
 30. Rieke C, Mormann F, Andrzejak RG, Kreuz T, David P, Elger CE, Lehnertz K. Discerning nonstationarity from nonlinearity in seizure-free and preseizure EEG recordings from epilepsy patients. *IEEE Trans Biomed Eng*. 2003;50(5):634-9. doi: 10.1109/TBME.2003.810684. PubMed PMID: 12769439.
 31. Talebi N, Nasrabadi AM. Recurrence plots for identifying memory components in single-trial EEGs. In: International conference on brain informatics; Berlin, Heidelberg: Springer; 2010. p. 124-32.
 32. Andrzejak RG, Lehnertz K, Mormann F, Rieke C, David P, Elger CE. Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: dependence on recording region and brain state. *Phys Rev E Stat Nonlin Soft Matter Phys*. 2001;64(6 Pt 1):061907. doi: 10.1103/PhysRevE.64.061907. PubMed PMID: 11736210.
 33. Banbrook M, McLaughlin S, Mann I. Speech characterization and synthesis by nonlinear methods. *IEEE Transactions on Speech and Audio Processing*. 1999;7(1):1-7. doi: 10.1109/89.736326.
 34. Rosenstein MT, Collins JJ, De Luca CJ. Reconstruction expansion as a geometry-based framework for choosing proper delay times. *Physica D: Nonlinear Phenomena*. 1994;73(1-2):82-98. doi: 10.1016/0167-2789(94)90226-7.
 35. Pincus SM. Approximate entropy as a measure of system complexity. *Proc Natl Acad Sci U S A*. 1991;88(6):2297-301. doi: 10.1073/pnas.88.6.2297. PubMed PMID: 11607165. PubMed PMCID: PMC51218.
 36. Richman JS, Moorman JR. Physiological time-series analysis using approximate entropy and sample entropy. *Am J Physiol Heart Circ Physiol*. 2000;278(6):H2039-49. doi: 10.1152/ajpheart.2000.278.6.H2039. PubMed PMID: 10843903.
 37. Zhang X, Zhou P. Sample entropy analysis of surface EMG for improved muscle activity onset detection against spurious background spikes. *J Electromyogr Kinesiol*. 2012;22(6):901-7. doi: 10.1016/j.jelekin.2012.06.005. PubMed PMID: 22800657. PubMed PMCID: PMC3514830.
 38. Shi J, Zhao P, Cai Y, Jia J. Classification of hand motions from surface electromyography

- with rough entropy. *Journal of Medical Imaging and Health Informatics*. 2015;5(2):328-34. doi: 10.1166/jmihi.2015.1396.
39. Höller Y, Nardone R. Quantitative EEG biomarkers for epilepsy and their relation to chemical biomarkers. *Adv Clin Chem*. 2021;102:271-336. doi: 10.1016/bs.acc.2020.08.004. PubMed PMID: 34044912.
40. Han J, Dong F, Xu YY. Entropy feature extraction on flow pattern of gas/liquid two-phase flow based on cross-section measurement. *J Phys: Conf Ser*. 2009;147(1):012041. doi: 10.1088/1742-6596/147/1/012041.
41. Fielding AH. Cluster and classification techniques for the biosciences. Cambridge University Press; 2006.
42. Cogan D, Birjandtalab J, Nourani M, Harvey J, Nagaraddi V. Multi-Biosignal Analysis for Epileptic Seizure Monitoring. *Int J Neural Syst*. 2017;27(1):1650031. doi: 10.1142/S0129065716500313. PubMed PMID: 27389004.
43. Chandaka S, Chatterjee A, Munshi S. Cross-correlation aided support vector machine classifier for classification of EEG signals. *Expert Systems with Applications*. 2009;36(2):1329-36. doi: 10.1016/j.eswa.2007.11.017.
44. Nonlinear Time Series Analysis Group. The Bern-Barcelona EEG database. 2013. Available from: <http://ntsa.upf.edu/downloads>.
45. Klem GH, Lüders HO, Jasper HH, Elger C. The ten-twenty electrode system of the International Federation. The International Federation of Clinical Neurophysiology. *Electroencephalogr Clin Neurophysiol Suppl*. 1999;52:3-6. PubMed PMID: 10590970.
46. Polychronaki GE, Ktonas PY, Gatzonis S, Siatouni A, Asvestas PA, Tsekou H, et al. Comparison of fractal dimension estimation algorithms for epileptic seizure onset detection. *J Neural Eng*. 2010;7(4):046007. doi: 10.1088/1741-2560/7/4/046007. PubMed PMID: 20571184.
47. Chua KC, Chandran V, Acharya UR, Lim CM. Automatic identification of epileptic electroencephalography signals using higher-order spectra. *Proc Inst Mech Eng H*. 2009;223(4):485-95. doi: 10.1243/09544119JEIM484. PubMed PMID: 19499838.
48. Acharya UR, Oh SL, Hagiwara Y, Tan JH, Adeli H. Deep convolutional neural network for the automated detection and diagnosis of seizure using EEG signals. *Comput Biol Med*. 2018;100:270-8. doi: 10.1016/j.combiomed.2017.09.017. PubMed PMID: 28974302.
49. Marwan N, Romano MC, Thiel M, Kurths J. Recurrence plots for the analysis of complex systems. *Physics Reports*. 2007;438(5-6):237-29. doi: 10.1016/j.physrep.2006.11.001.
50. Li P, Karmakar C, Yan C, Palaniswami M, Liu C. Classification of 5-S Epileptic EEG Recordings Using Distribution Entropy and Sample Entropy. *Front Physiol*. 2016;7:136. doi: 10.3389/fphys.2016.00136. PubMed PMID: 27148074. PubMed PMCID: PMC4830849.
51. Chen W, Zhuang J, Yu W, Wang Z. Measuring complexity using FuzzyEn, ApEn, and SampEn. *Med Eng Phys*. 2009;31(1):61-8. doi: 10.1016/j.medengphy.2008.04.005. PubMed PMID: 18538625.
52. Valipour F, Esteki A. Pattern Classification of Hand Movement Tremor in MS Patients with DBS ON and OFF. *J Biomed Phys Eng*. 2022;12(1):21-30. doi: 10.31661/jbpe.v0i0.1028. PubMed PMID: 35155289. PubMed PMCID: PMC8819264.
53. Hussain L, Saeed S, Awan IA, Idris A. Multiscaled complexity analysis of EEG epileptic seizure using entropy-based techniques. *Arch Neurosci*. 2018;5(1):1-1. doi: 10.5812/archneurosci.61161.
54. Wu BF, Wang KC. Robust endpoint detection algorithm based on the adaptive band-partitioning spectral entropy in adverse environments. *IEEE Transactions on speech and audio processing*. 2005;13(5):762-75. doi: 10.1109/TSA.2005.851909.