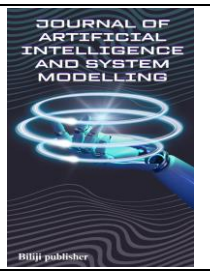
Original/Research Article

OPEN ACCESS

Journal of Artificial Intelligence and System Modelling

e-ISSN:3041-850X
Volume 02, Issue 02, 2024

Journal Web Page: https://jaism.bilijipub.com



A Differential Biomarker Based on Recurrence Quantification Analysis of EEG Signal and Genetic Algorithm for Epilepsy Diagnosis

Vanithamani Palanisamy [1], A.Ranichitra [2 *], Radhamani Ellapparaj [2]

[1] *Department of Computer Science, PPG College of Arts and Science, Coimbatore, Tamil Nadu 641035, India*
[2] *Department of Computer Science, College of Sri S. Ramasamy Naidu Memorial, Sattur, Tamil Nadu, 626203, India*

Highlights

- RQA with Bayesian classifiers and genetic algorithms distinguished ictal and normal EEGs effectively.
- Recurrence plots used five distance norms and ten thresholds, yielding 100 samples per signal category.
- Longest diagonal line, transitivity, and recurrence rate features achieved up to 100% accuracy.
- Minimum norm with ε = 0.4 achieved 100% discrimination, making transitivity a top seizure detection biomarker.

Article Info

Received: 25 May 2024
Received in revised: 15 June 2024
Accepted: 30 June 2024
Available online: 30 Jnue 2024

Keywords

Electroencephalogram,
Epilepsy,
Recurrent Quantification Analysis,
Genetic Algorithm,
Biomarker,
Machine Learning.

Abstract

Selecting features with strong discriminatory capabilities is crucial for data classification challenges. Recurrence Quantification Analysis (RQA) is a promising technique for detecting seizures without assuming stationary conditions, accommodating various signal and noise sizes. In this study, RQA was used to distinguish between ictal and normal EEGs, utilizing a combination of Bayesian classifiers and genetic algorithms to select optimal RQA features. Recurrence plots were generated using five different distance norms (Mahalanobis, maximum, minimum, and Manhattan) and 10 threshold levels (εmin = 0.1, εmax = 1, Δε = 0.1) for each signal category, totaling one hundred samples. Examining the participation rate of each feature in all experiments showed that each feature appeared on average in 52% of repetitions, among which transitivity and determinism features had the highest and lowest participation in the feature selection stage with 64% and 33%, respectively. Among 12 calculated RQA features from EEGs, the features of longest diagonal line, transitivity and the recurrence rate with 6, 4 and 3 numbers of 100% accuracy in separating normal and epileptic EEGs yielded better results than other recurrence features. On the other hand, the features of divergence, trapping time and longest vertical line without occurrence of 100% accuracy yielded the poorest results. Experimental results showed that using the minimum norm and ε = 0.4 achieved a 100% discrimination rate for seizure detection. The transitivity recurrence feature proved highly effective in classifying normal and epileptic EEGs, making it an excellent biomarker for seizure detection with high diagnostic value.

1. Introduction

Epilepsy is a neurological disorder characterized by recurrent and unprovoked seizures, affecting individuals of all ages. It has a significant global prevalence, with an estimated 65 million people worldwide living with the condition [1]. Epilepsy can have a profound impact on an individual's quality of life, leading to increased mortality and morbidity risks [2,3]. While it is generally not a life-threatening condition, mortality rates among people with epilepsy are higher compared to the general population [4,5]. These rates are primarily attributed to accidents, sudden unexpected deaths in epilepsy (SUDEP), and

* Corresponding Author: A.Ranichitra
Email: ranichitra@srnmcollege.ac.in

https://doi.org/10.22034/jaism.2024.459400.1046

associated comorbidities. Diagnosing epilepsy often involves the analysis of electroencephalogram (EEG) recordings, which capture the electrical activity of the brain. By examining EEG data, healthcare professionals can identify abnormal brainwave patterns and markers that are indicative of epilepsy [6,7]. EEG analysis plays a crucial role in providing an accurate diagnosis, enabling appropriate treatment planning, and improving the overall management of individuals with epilepsy [8,9]. Nowadays, the design of recognition and detection systems for epileptic seizures is one of the fields of interest for medical and engineering researchers. On the other hand, it is necessary to choose methods that can reveal the basic dynamic features of the brain system [10]. The issue of the return of states has been studied for a long time. In 1987, Ekman introduced a method to show the repetition of system states in phase space. The repetition of system states means that the system states approach each other after the passage of time, and this is one of the basic characteristics of deterministic systems. Using recurrent mapping, a proper representation of the system routes is created in two-time dimensions [11].

Due to the reduction of EEG signal irregularity during epileptic seizures, the system deals with an almost similar neighborhood in the phase space. As a result, using this method for detecting seizures can be meaningful. In addition, the main advantage of a recurrence plot is its ability to be used with short and even non-stationary data [12]. In 2008, Ouyang and colleagues used recurrence quantification analysis (RQA) to diagnose the three states of normal, pre-ictal and ictal in mice with genetic absence epilepsy. The results of this research showed that there is a significant difference between epileptic data and two other categories according to recurrent features [13]. The study [14] reported an accuracy of 95.6% in the classification of epileptic and normal states through EEG analysis by applying 10 features from recurrence plots to the SVM classifier. In a study [15], the EEG signal of normal, inter-ictal and ictal human states was decomposed into delta, theta, alpha and gamma sub-bands, and then the recurrent features of these sub-bands were extracted and used as input to the learning machine. The authors of this study reported an overall accuracy of 98.67% in the classification of these three states using recurrent features.

Marwan showed that the appropriate choice of the distance norm in the RQA is effective in the results and can provide the possibility of analyzing some of the recurrent features and increase the calculation speed of the recurrent patterns [16]. For different distance norms, different degrees of neighborhood are considered and the preservation of details related to each plot is different. Furthermore, a small threshold ε causes the loss of recurrent points. In contrast, if ε is chosen too large, any two points in the space are neighbors and the error increases. On the other hand, using all features in the classification problem does not necessarily lead to the best result. Feature selection algorithms are divided into open-loop and closed-loop depending on their evaluation processes [17–19]. It has been shown that the closed-loop algorithms usually obtain better results [20]. This method performs the search in the space of subsets based on the estimation of accuracy resulting from the selection of a specific subset under the conditions of the used classification algorithm (as a criterion of the optimality of that subset) [21]. In these methods, the most important issue is the search algorithm used. During the past decade, researchers have focused on evolutionary search algorithms such as genetic algorithm [22], ant algorithm [23], and particle/ant swarm algorithm [24]. Among them, the genetic algorithm's ability to quickly and comprehensively search a large space makes it suitable for feature selection [20].

Although RQA has been used for epileptic EEG analysis in past studies, important details related to this method that have a significant effect on classification performance have not been investigated yet. The main contribution of this research is to investigate all the details of RQA to extract distinctive features from EEG signals and optimize the results using machine learning approaches to determine a potential biomarker for epilepsy diagnosis. The main policy in this article is to evaluate the efficiency of a subset of features of recurrence plots using a combination of a genetic algorithm and Bayesian classifier. To the best of our knowledge, no valid biomarker for the automatic diagnosis of epilepsy has been introduced yet. Therefore, the aim of this paper is to introduce a potential biomarker based on the nonlinear analysis of the EEG signal.

2. Methods

In this section, the EEG database used, signal processing and feature selection methods, as well as the classification model, are described.

2.1. Dataset

The data used for testing and evaluation was related to the University of Bonn, Germany [25]. The used data set consists of two data categorizations: the normal data recorded from five healthy people with open eyes, and the epileptic data recorded from five patients with right hippocampus epilepsy. EEG data contain 100 epochs with a sampling frequency of 173.61 Hz and a duration of 23.6 s. EEG recording was done through 19 electrodes according to

the 10-20 international system. Following a 12 bit analog-to-digital transformation, the recordings were written continually onto the disk of a data acquisition computer system at a sampling rate of 173.61 Hz. Band-pass filter settings were 0.53–40 Hz. Exemplary EEG recordings are shown in Figure 1. As can be seen, the morphology of the signal in different states has obvious differences, which can be caused by different brain dynamics in the production of each of these EEG signals.

2.2. *EEG preprocessing*

To enhance signal quality and eliminate interference from city electricity, a 4-th order Butterworth band-pass filter ranging from 0.5 to 40 Hz and a 50 Hz notch filter were implemented. The stopband frequencies were positioned outside the passband to effectively diminish noise and artifacts, with the Butterworth filter stopband spanning from 0-0.5 Hz to 40-120 Hz. This resulted in a filter bandwidth of 39.5 Hz and a center frequency of 20.25 Hz; the Q-factor for the band-pass filter was established at 1. Additionally, passband ripple ranged between 0.1 dB to 0.5 dB with a stopband attenuation of at least 40 dB. Furthermore, independent component analysis (ICA) was utilized to eliminate biological artifacts like eye movements by statistically isolating independent components viewed as linear combinations of EEG electrodes, each characterized by a distinct topography and time period. Post ICA application, semi-automatic selection of independent components of the EEG for artifact correction (SASICA) was employed to enhance objectivity and reproducibility in preprocessing reports through the use of quantitative parameters and thresholds. A total of 20 components were designated for extraction, and automatic component rejection was executed using joint probability-based thresholding within the 0.5-40 Hz frequency range.

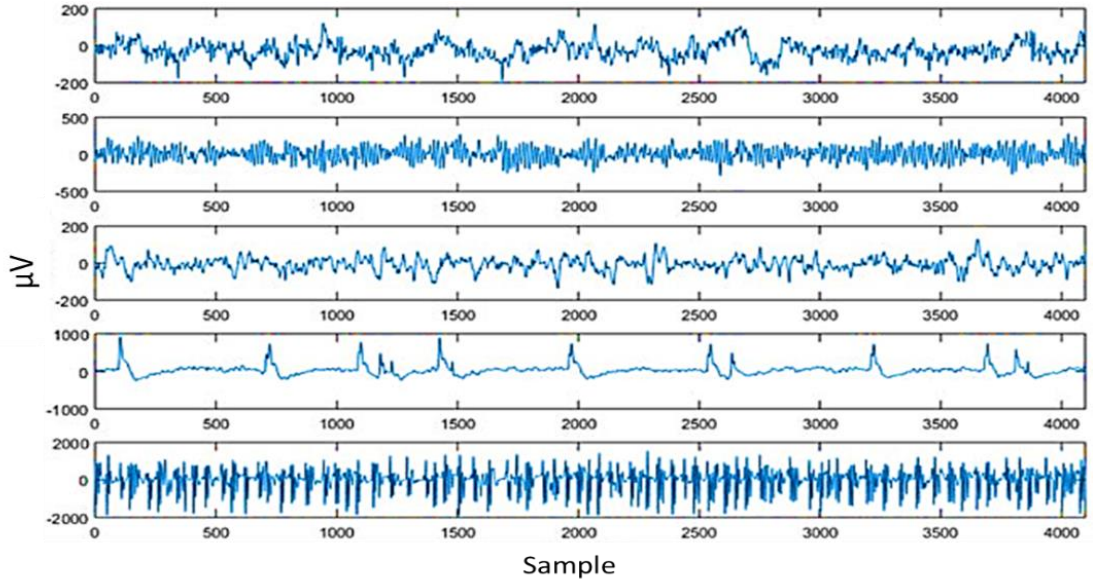


Fig 1. Example of EEGs from Bonn University Database used in this research.

2.3. *Recurrence quantification analysis*

This method, which was first implemented in 1998 by Ekman and colleagues, is able to show the details of the non-stationary EEG signal at different times [26,27]. RQA is a powerful technique used to analyze and characterize the patterns and dynamics of complex systems, particularly in time series data. It is employed to explore the presence and properties of recurrent behaviors within a given dataset. RQA is based on the concept of recurrence plots, which visually represent the recurrence of states in a phase space. The main idea of this method is to return the state of the system to the same regions of the phase space where there was a path before [28]. The repetition of a state that occurred at time i, at different times j is determined by a square matrix containing 0 and 1 (black and white points in the diagram), both of which are the axes of the time diagram. Mathematically speaking, arecurrence plot can be created as follows:

$$R(i.j) = \theta(\varepsilon(i) - \|x(i) - x(j)\|).\ x \in R^m.\ i.j = 1 \ldots . N \quad (1)$$

where N represent the length of the time series, θ denotes the step function, ||.|| denotes the distance norm, and ε(i) denotes the radius that should be chosen for the neighborhood of each point x_i. By selecting appropriate features from the recurrence plot, hidden patterns and structural changes in the dynamics can be revealed. Table 1 shows the features extracted from the recurrence plots resulting from the RQA.

Table 1. Estimated features from the recurrence plots of ictal and normal EEGs for epilepsy diagnosis.

Feature	Description
Recurrence rate	The proportion of recurrence points within a certain radius around each data point
Determinism	The proportion of recurrent points that form diagonal lines in the plot
Entropy	The degree of randomness based on the distribution of recurrent points in the plot
Trend	An indicator of the vertical trend within the recurrent patterns captured in the plot
Longest diagonal line	The length of the longest continuous diagonal line segment in the plot
Divergence	The measure of the spreading or dispersal of recurrent points in the plot
Transitivity	The degree of connectedness among recurrent points in the plot
Trapping time	A measure of the average time that a trajectory remains trapped or confined in a specific region of the phase space
Laminarity	The linearity or interconnectedness of recurrent points or clusters in the plot; the proportion of recurrence points that form vertical or horizontal line structures in the plot
Average diagonal line	The prevalence or density of diagonal line structures in the plot
Longest vertical line	The vertical line of consecutive recurrence points that spans the greatest number of rows in the plot
Recurrence time entropy	The unpredictability or entropy of the time intervals between recurrences in the recurrence plot

As mentioned, this method is suitable for non-stationary signal processing. Figure 2 shows the first and second order moments of two groups of normal and epileptic EEG signals over time, which proves the non-stationarity of the brain signal due to the time-varying moments.

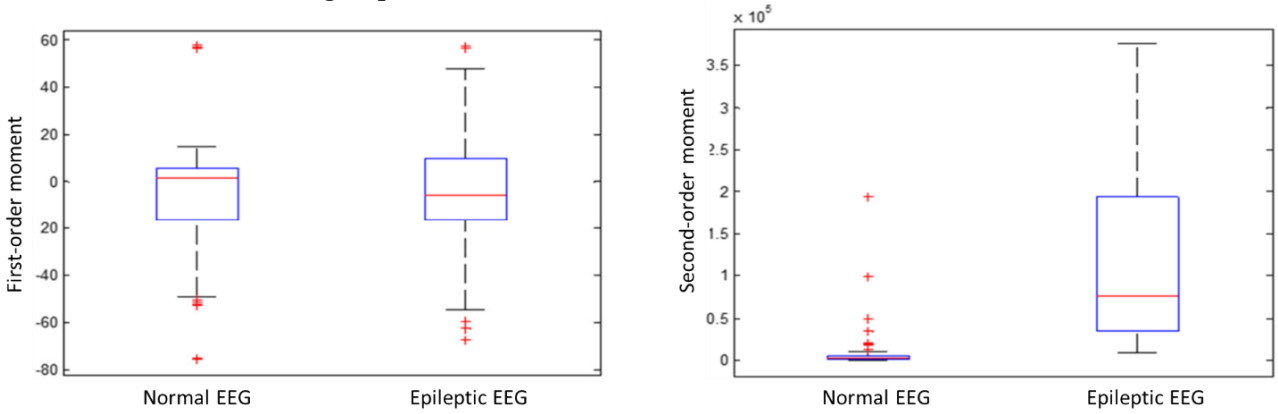


Fig 2. Box plots of first- and second-order moments for normal and epileptic EEGs.

2.4. Genetic algorithm

Genetic algorithms have a good place among other methods due to their high ability to solve optimization problems. This algorithm provides fully parallel search methods for complex optimization problems. Genetic algorithms have fundamental differences with common search and optimization methods [29]. This algorithm does not have special mathematical requirements and solves optimization problems regardless of the internal performance of the problem. The structure of genetic algorithm operators enables this algorithm to succeed in finding the overall optimal solution. Genetic Algorithms provide high flexibility to be combined with innovative techniques and thus enable an efficient and effective solution of a problem [30]. In a genetic algorithm, several parameters play crucial roles in determining the performance and behavior of the algorithm. The population size refers to the number of individuals or candidate solutions in each generation. A larger population size increases the exploration capability of the algorithm but also increases the computational cost. The selection mechanism determines which individuals will be chosen as parents for reproduction. It can be based on fitness proportionate selection, tournament selection, or other approaches. The crossover rate defines the probability of performing crossover, which is the recombination of genetic material between two parent individuals. A higher crossover rate allows for more exploration of the solution space. The mutation rate represents the probability of altering individual genes randomly. It introduces diversity into the population and helps avoid premature convergence. The termination condition specifies the stopping criterion for the algorithm, such as reaching a maximum number of generations or achieving a desired fitness level. Choosing appropriate values for these

parameters is crucial for achieving effective and efficient optimization using genetic algorithms [31–33]. In this study, the initial population is the number of chromosomes in each generation and is of a discrete type. To produce a new generation, the selection operator, single-point crossover and mutation were used, and in each step, the number of chromosomes of the new generation was equal to the number of chromosomes of the initial population (Table 2).

Table 2. Selected parameters for genetic algorithm to determine the best subset of computed features for classification of normal and epileptic EEGs.

Parameters	Value
Initial population	50
Chromosome size	12
Minimum gene value	1
Maximum gene value	12
Crossover rate	0.8
Mutation rate	0.05
Selection rate	0.15

2.5. Bayesian Classifier

In statistics, the Bayesian classifier is an optimal classifier that minimizes the average error. In other words, this classifier tries to find the main minimum in the error function. In this classifier, the feature vector X is belonged to the class whose probability is $P(\omega_i|X)$, which is generally expressed as follows [34].

$$if\ X \in \omega_i \Rightarrow P(\omega_i|X) > P(\omega_j|X) \quad for\ all\ i \neq j \tag{2}$$

where ω_i represents the i^{th} class. On the other hand, according to Bayes theorem,

$$P(\omega_i|X) = \frac{P(\omega_i)P(\omega_i|X)}{P(X)} \tag{3}$$

where $P(\omega_i|X)$ represents the probability density class ω_i, $P(\omega_i)$ represents the probability of occurrence of class ω_i, and P(X) represents the probability of occurrence of class X, which is usually omitted. Hence,

$$g_i(X) = P(\omega_i)P(\omega_i|X) \tag{4}$$

Assuming that the feature vector density function follows the normal distribution, we have:

$$P(X|\omega_i) = \frac{1}{(2\pi)^{\frac{n}{2}}|\sum_i|^{\frac{1}{2}}} e^{-\frac{1(X-\mu_i)^T(X-\mu_i)}{2\sum_i}} \tag{5}$$

where μ_i and $\sum_i$ represent mean and covariance of testing feature vectors of each class, respectively. So, according to this relationship, the decision-making function for the i^{th} class is equal to:

$$g_i(X) = -\frac{1}{2}(X-\mu_i)\sum_i^{-1}(X-\mu_i)^T - \frac{1}{2}log(|\sum_i|) + log(P(\omega_i)) \tag{6}$$

Vector X will belong to the class whose decision function is maximized. Figure 3 shows the structure of a Bayes classifier.

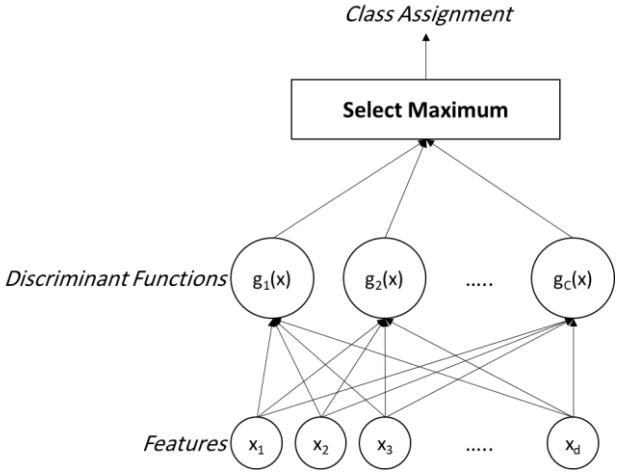


Fig 3. Bayesian classifier structure used in this study for classification of normal and epileptic EEGs.

2.6. Proposed framework

First, the recurrence plots are reconstructed based on the five distance norms of Euclidean, Mahalanobis, maximum norm, minimum norm and Manhattan, as well as 10 thresholds with 0.1 intervals, and recurrence features are computed. Then the Bayesian classifier is trained based on the features determined by the genetic algorithm and the

available training data at each stage. The refinement of the genes that match the features continues until the best subset of features that results in the best classification rate is selected. Figure 4 shows the block diagram of the proposed framework for epilepsy diagnosis from EEG signals.

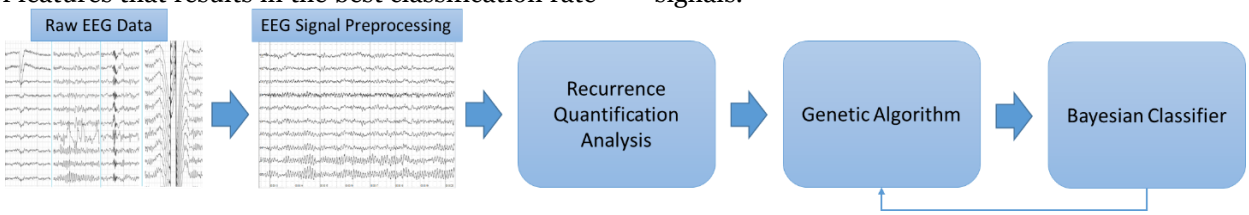


Fig 4. Block diagram of the proposed framework for epilepsy diagnosis from EEG signals.

3. Results

In the first step, recurrence plot reconstruction was performed for five distance norms and 10 threshold levels (ε_{min} = 0.1, ε_{max} = 1, $\Delta\varepsilon$ = 0.1). The time delay τ and the embedding dimension m were calculated for all the samples using the average mutual information and false nearest neighbor methods, respectively. To determine the optimal time delay, the average function of mutual information was calculated according to the time delay for each EEG epoch. The optimal τ values corresponding to the occurrence of the first minimum of the average function of mutual information were obtained in the range of 4 to 11. Also, the embedding dimension was calculated by the false nearest neighbor method for different embedding dimensions for each EEG epoch. The embedding dimension was obtained between 5 and 10 for all data. Figure 5 shows the pattern of recurrence plots for the five distance norms for ε = 0.5. For different criteria, different degrees of neighborhood were considered, which led to preserving different details in each plot. As shown, the percentage of black points increased from plot A to E. In the next step, the recurrence features (mentioned in Table 1) were calculated using the CRP Toolbox for the considered states and applied to the recognition system.

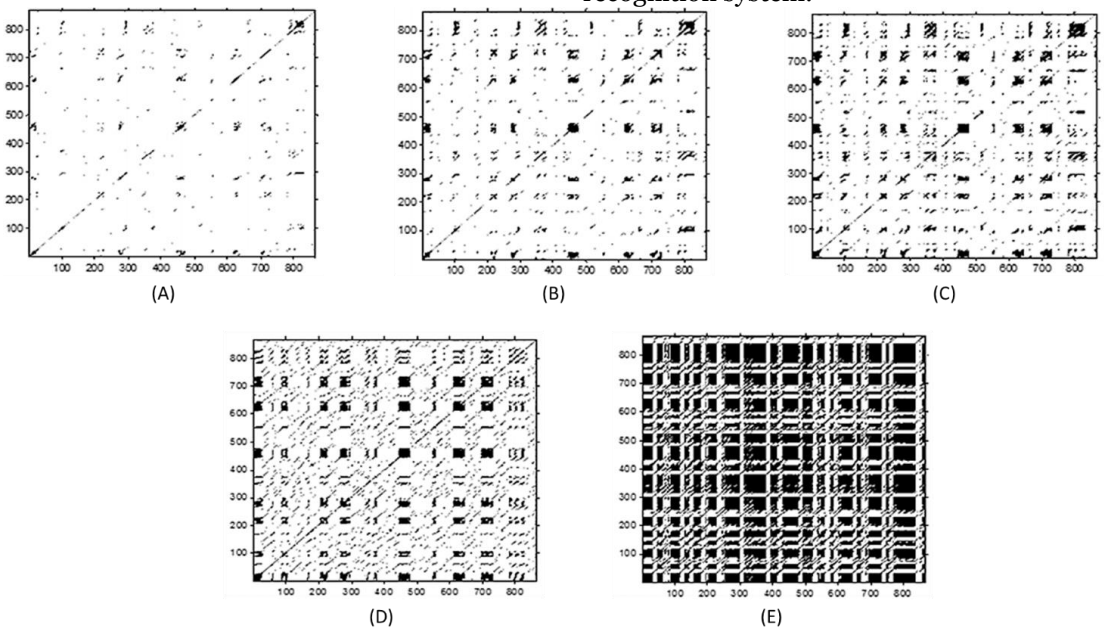


Fig 5. Examples of recurrence plots of epileptic EEG for distance norms of (A) Euclidean, (B) maximum norm, (C) minimum norm, (D) Mahalanobis, and (E) Manhattan for ε = 0.5.

In the feature selection phase, the genetic algorithm was implemented with 50 iterations, the results of which are mentioned in Table 3. Accuracy values of the classifier are reported in this table. When evaluating classifier performance, accuracy plays a crucial role in understanding the behavior of the classifier. The accuracy measures the overall correctness of the classifier by indicating the proportion of correctly classified instances (true positives and true negatives) among the total instances. Figure 6 shows that an average of 6.27 features were simultaneously involved in the feature selection stage, in which the Euclidean criterion was the least and the Mahalanobis criterion was the most necessary to achieve appropriate separation accuracy.

Table 3. Mean classification accuracy of normal and epileptic EEGs after 50 iterations of the genetic algorithm for five distance norms and 10 threshold levels.

Distance norm	ε = 0.1	ε = 0.2	ε = 0.3	ε = 0.4	ε = 0.5	ε = 0.6	ε = 0.7	ε = 0.8	ε = 0.9	ε = 1
Maximum norm	100%	100%	100%	100%	100%	98.1%	100%	99.8%	99.9%	99.9%
Minimum norm	98.9%	100%	100%	100%	100%	100%	100%	100%	100%	99.9%
Euclidean	100%	100%	100%	100%	100%	100%	99.7%	99.9%	99.8%	99.8%
Mahalanobis	98.7%	98.7%	98.8%	100%	98.4%	98.5%	98.4%	98.3%	98.4%	100%
Manhattan	99.6%	99.7%	99.4%	99.7%	98.4%	99.4%	98.7%	98.6%	98%	100%

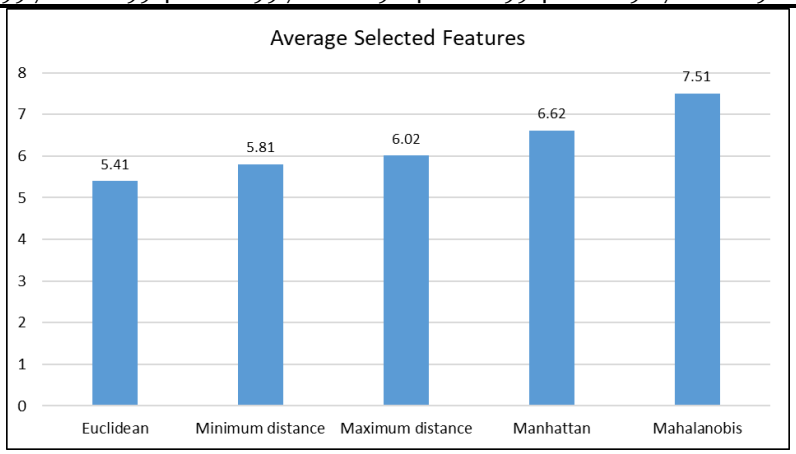


Fig 6. The average number of features utilized in 500 iterations (50×10) of the genetic algorithm.

In order to determine the optimal feature in the RQA, three criteria were studied: the participation rate of the feature in all repetitions, the discriminability of the feature, and the occurrence of 100% accuracy for each feature. Examining the participation rate of each feature in all experiments showed that each feature appeared on average in 52% of repetitions, among which transitivity and determinism features had the highest and lowest participation in the feature selection stage with 64% and 33%, respectively (Figure 7).

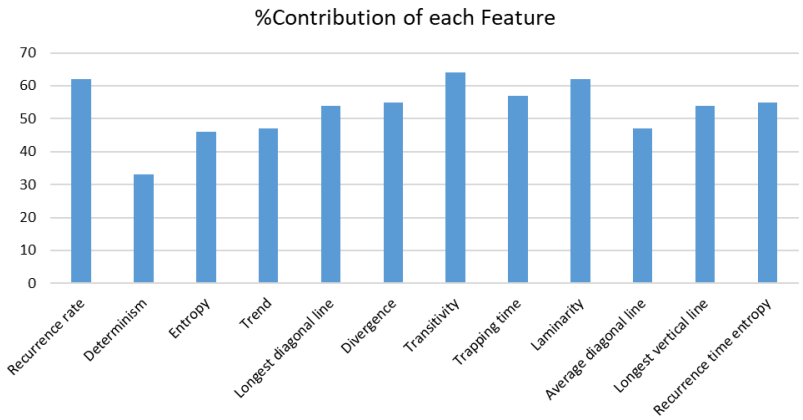


Fig 7. The participation rate of each feature in 2500 iterations (50×10×5) of the genetic algorithm.

An independent t-test was used to check the discriminability of each feature [35–38]. Figure 8 shows the obtained P-values for each feature to differentiate between two groups of normal and epileptic EEG. P-value calculation using independent t-test showed that the features of transitivity and the longest vertical line have the most and the least discriminability, respectively. In addition, Table 4 shows the results for the number of occurrences with 100% accuracy in separating normal and epileptic EEGs for each feature. As shown, the features of longest diagonal line, transitivity and recurrence rate with 6, 4 and 3 numbers of 100% accuracy in separating normal and epileptic EEGs yielded better results than other recurrence features. On the other hand, the features of

divergence, trapping time and longest vertical line without occurrence of 100% accuracy yielded the poorest results.

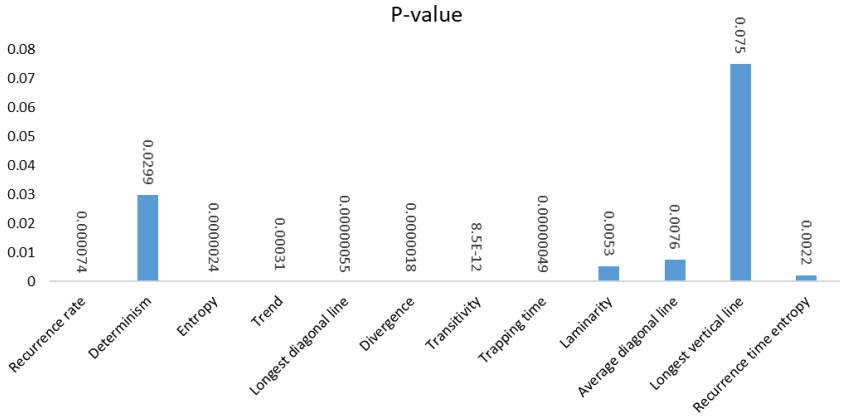


Fig 8. The obtained P-values through independent t-test for each recurrence feature to differentiate between two groups of normal and epileptic EEG.

Table 4. The number of occurrences with 100% accuracy in separating normal and epileptic EEGs for each recurrence feature.

Feature	Distance norm	Threshold level	Number of 100% accuracies	Total number of 100% accuracies
Recurrence rate	Mahalanobis	1	3	3
Determinism	Mahalanobis	1	1	1
Trend	Euclidean	0.1	1	2
		0.2	1	
Entropy	Mahalanobis	1	1	1
Longest diagonal line	Minimum norm	0.2	2	6
	Maximum norm	0.2	1	
	Euclidean	0.2	3	
Divergence	-	-	-	-
Transitivity	Mahalanobis	1	4	4
Trapping time	-	-	-	-
Laminarity	Mahalanobis	1	2	2
Average diagonal line	Euclidean	0.2	2	2
Longest vertical line	-	-	-	-
Recurrence time entropy	Manhattan	0.4	1	1

Table 5 summarizes the results of the three investigated criteria for selecting the best recurrence feature in separating normal and epileptic EEGs with the best accuracy. Based on this, it seems that the transitivity recurrence feature has a high ability to classify normal and epileptic EEG.

Table 5. The optimal recurrence features for classification of normal and epileptic EEG.

Evaluation criterion	Optimal feature
Participation rate	Transitivity
Discriminability	Transitivity
Occurrence of 100% accuracy	Longest diagonal line

4. Discussion

In this article, a new method for diagnosing epileptic EEGs based on recurrence quantification analysis and the combination of genetic algorithms and Bayesian classification is presented, and for the first time, the optimal subset of the recurrence plot for the diagnosis of epileptic seizures is obtained. The experimental results showed that this optimal subset of features is able to distinguish epileptic EEGs from normal ones with 100% diagnostic accuracy. Considering the high participation rate of the transitivity feature in the feature selection

process and its high ability to distinguish epileptic EEG from normal EEG, it can be considered an optimal feature and potential biomarker in the diagnosis of epileptic seizures. This feature represents the dependence of the number of recurrence points of lines parallel to the central diagonal line relative to their distance from this line [39]. In the context of recurrence plots, transitivity refers to the inherent property of interconnectedness or the occurrence of consecutive patterns within a time series. Specifically, it captures the notion of the formation of a triangular pattern among three data points in the time series. When constructing a recurrence plot, we examine pairs of data points and determine if their proximity in the phase space exceeds a predefined threshold. Transitivity extends this analysis by examining not just pairs, but triplets of points [40]. It considers three data points A, B, and C, where the distance between A and B, as well as the distance between B and C, are below the threshold. If the distance between points A and C also falls within the threshold, a transitive relationship is established, resulting in the formation of a recurrence triangle. The transitivity feature in recurrence plots provides valuable insights into the underlying dynamics and complexity of time series data. It characterizes the presence and interconnectedness of multiple recurrent patterns, revealing information about the system's regularity, periodicity, or chaotic behavior. By analyzing the distribution and abundance of recurrence triangles in a recurrence plot, researchers can quantify and compare the level of transitivity across different time series or investigate temporal changes in system dynamics [41]. This feature has found application in various fields, including physics, biology, neurology, and finance, offering a unique perspective to understand the complex relationships within time-dependent data [42–44].

Although maximum accuracy has been achieved in some of the recent research in seizure detection, the correlation of RQA with the nature of chaotic systems enables us to use it in recognition systems, as well as to interpret system and results from the chaos point of view. RQA offers several key advantages for EEG analysis. Firstly, RQA is a nonlinear method that allows the exploration of the complex and dynamic nature of brain activity, going beyond traditional linear techniques. It provides a comprehensive understanding of the recurrence patterns in EEG signals, unveiling hidden temporal dependencies and capturing nonlinear interactions. Furthermore, RQA is robust to noise and artifacts commonly present in EEG recordings, ensuring reliable and accurate analysis. It also offers a data-driven approach, removing the need to assume linearity or stationarity. RQA's ability to quantify various complexity measures, such as recurrence rate and transitivity, enables researchers to investigate changes in brain dynamics associated with different cognitive states or neurological disorders. Overall, RQA serves as a powerful tool to unravel the intricate dynamics of EEG signals and deepen our understanding of brain functioning [39].

Table 6 compares previous EEG findings in seizure diagnosis with current results. As shown, the results obtained in the present research are superior to many previous techniques. This indicates the importance of RQA in epileptic EEG analysis and the automatic diagnosis of this brain abnormality. Along with the strengths of this study in providing a technique with high diagnostic accuracy and introducing a potential biomarker for the diagnosis of epilepsy from the EEG signal, there are also limitations that should be mentioned. To effectively use this approach in clinical settings, it is crucial to obtain a wider range of EEG data sets specific to different epileptic seizures in a large population of epileptic patients of different ages and sexes. This is particularly significant for machine learning techniques as they necessitate extensive datasets to achieve optimal results. In addition, future studies should validate the proposed approach in longitudinal designs with a broad range of patients. However, considering that the performance of the proposed method is promising, it can serve as a CAD tool for clinical purposes. It is also worth mentioning that the presented framework offers advantages such as minimal labor, time efficiency, and decreased susceptibility to human errors in comparison to traditional epilepsy diagnosis approaches. Consequently, it can offer a swift and accurate diagnosis of seizures without the need for direct human involvement.

Table 6. Comparing previous EEG findings in seizure diagnosis with the current results.

Reference	EEG features	Classifier	Reported accuracy (%)
[45]	Coefficients of a nonlinear filter	Neural network	97.20
[46]	Entropy measures	ANFIS	92.22
[47]	Time-frequency features	Recurrent neural network	99.60
[48]	Discrete wavelet coefficients	Adaptive fuzzy network	85.90
[49]	Discrete wavelet coefficients and statistical features	Ensemble classifier	94.50
[50]	Spectral features	Decision tree	98.22
[51]	Time-frequency features	Neural network	97.20

[52]	Approximate entropy	Probabilistic neural network	100
[53]	Spectral features	Neural network	100
[54]	AR model coefficients	Decision tree	99.32
[55]	Entropy and discrete wavelet coefficients	ROC curve	96.65
[56]	Wavelet entropy coefficients	Neural network	95.20
Current research	RQA	Bayesian classifier	100

5. Conclusion

This research is one of the few studies that introduces a potential biomarker based on nonlinear EEG signal analysis (i.e., RQA technique) and machine learning (i.e., a combination of genetic algorithm and Bayesian classification) for epilepsy diagnosis. According to the results obtained in this research, it is recommended that future studies focus more on recurrence plots and transitivity feature. The transitivity feature is proved to be an excellent EEG biomarker for seizure detection with high diagnostic value. In clinical settings, such a biomarker could aid in the differentiation of epileptic seizures from other similar conditions, leading to more targeted and effective treatment strategies for patients. Additionally, it could potentially enable the monitoring of disease progression and treatment response, contributing to personalized medical interventions for individuals with epilepsy. However, the findings of this study require external validation using other epileptic EEG databases. For future research, it would be beneficial to conduct prospective clinical studies to validate the performance of this biomarker in diverse patient populations. Furthermore, investigating the potential integration of this biomarker with existing diagnostic tools and technologies, such as neuroimaging modalities or wearable devices, could enhance its clinical applicability. Additionally, exploring longitudinal data analysis to assess the predictive value of the biomarker in tracking the evolution of epilepsy over time would be an important direction for further investigation. Finally, considering the ethical and practical implications of implementing this biomarker in real-world clinical practice, including issues related to patient privacy, data security, and healthcare resource allocation, is crucial for its successful translation into routine medical care.

Conflicts of interests

There is no conflict of interest in this study.

Funding

Self-funded.

Acknowledgements

Not applicable.

Availability of data and materials

The datasets used during the current study are available from the corresponding author on request.

Using artificial intelligent chatbots

None.

References

[1] Shafie M, Darijani J, Mirsepassi Z, Razavi A, Mayeli M, Arbabi M, Aghamollaii V. Psychosis of Epilepsy: A 10-Year Iranian Clinical Survey. Iran J Psychiatry 2023;18:476.

[2] Mohammadi MR, Ahmadi N, Khaleghi A, Mostafavi SA, Kamali K, Rahgozar M, Ahmadi A, Hooshyari Z, Alavi SS, Molavi P. Prevalence and correlates of psychiatric disorders in a national survey of Iranian children and adolescents. Iran J Psychiatry 2019;14:1.

[3] Talepasand S, Mohammadi MR, Alavi SS, Khaleghi A, Sajedi Z, Akbari P, Lari M, Kasaeian R, Eskandaripour M, Rashti E. Psychiatric disorders in children and adolescents: prevalence and sociodemographic correlates in Semnan Province in Iran. Asian J Psychiatr 2019;40:9–14.

[4] Beghi E. The epidemiology of epilepsy. Neuroepidemiology 2020;54:185–91.

[5] Khaleghi A, Mohammadi MR, Zandifar A, Ahmadi N, Alavi SS, Ahmadi A, Mahmoudi-Gharaei J, Hooshyari Z, Mostafavi A, Fooladi MF. Epidemiology of psychiatric disorders in children and adolescents; in Tehran, 2017. Asian J Psychiatr 2018;37:146–53.

[6] Khaleghi A, Mohammadi MR, Shahi K, Nasrabadi AM. Computational neuroscience approach to psychiatry: a review on theory-driven approaches. Clinical Psychopharmacology and Neuroscience 2022;20:26.

[7] Khaleghi A, Mohammadi MR, Shahi K, Nasrabadi AM. Possible neuropathological mechanisms underlying the increased complexity of brain electrical activity in schizophrenia: a computational study. Iran J Psychiatry 2023;18:127.

[8] Khaleghi A, Mohammadi MR, Jahromi GP, Zarafshan H. New ways to manage pandemics: Using technologies in the era of covid-19: A narrative review. Iran J Psychiatry 2020;15:236.

[9] Khaleghi A, Mohammadi MR, Shahi K, Motie Nasrabadi A. A neuronal population model based on

cellular automata to simulate the electrical waves of the brain. Waves in Random and Complex Media 2024;34:1445–64.
[10] Chen S, Zhang X, Chen L, Yang Z. Automatic diagnosis of epileptic seizure in electroencephalography signals using nonlinear dynamics features. IEEE Access 2019;7:61046–56.
[11] Xu X, Lin M, Xu T. Epilepsy seizures prediction based on nonlinear features of EEG signal and gradient boosting decision tree. Int J Environ Res Public Health 2022;19:11326.
[12] Pitsik E, Frolov N, Hauke Kraemer K, Grubov V, Maksimenko V, Kurths J, Hramov A. Motor execution reduces EEG signals complexity: Recurrence quantification analysis study. Chaos: An Interdisciplinary Journal of Nonlinear Science 2020;30.
[13] Ouyang G, Li X, Dang C, Richards DA. Using recurrence plot for determinism analysis of EEG recordings in genetic absence epilepsy rats. Clinical Neurophysiology 2008;119:1747–55.
[14] Acharya UR, Sree SV, Chattopadhyay S, Yu W, Ang PCA. Application of recurrence quantification analysis for the automated identification of epileptic EEG signals. Int J Neural Syst 2011;21:199–211.
[15] Niknazar M, Mousavi SR, Vahdat BV, Sayyah M. A new framework based on recurrence quantification analysis for epileptic seizure detection. IEEE J Biomed Health Inform 2013;17:572–8.
[16] Marwan N, Romano MC, Thiel M, Kurths J. Recurrence plots for the analysis of complex systems. Phys Rep 2007;438:237–329.
[17] Khaleghi A, Birgani PM, Fooladi MF, Mohammadi MR. Applicable features of electroencephalogram for ADHD diagnosis. Research on Biomedical Engineering 2020;36:1–11.
[18] Khaleghi A, Sheikhani A, Mohammadi MR, Nasrabadi AM, Vand SR, Zarafshan H, Moeini M. EEG classification of adolescents with type I and type II of bipolar disorder. Australas Phys Eng Sci Med 2015;38:551–9.
[19] Mohammadi MR, Khaleghi A, Nasrabadi AM, Rafieivand S, Begol M, Zarafshan H. EEG classification of ADHD and normal children using non-linear features and neural network. Biomed Eng Lett 2016;6:66–73.
[20] Verma S, Pant M, Snasel V. A comprehensive review on NSGA-II for multi-objective combinatorial optimization problems. IEEE Access 2021;9:57757–91.
[21] Dwivedi R, Kumar R, Jangam E, Kumar V. An ant colony optimization based feature selection for data classification. Int J Recent Technol Eng 2019;7:35–40.
[22] Kayode AA, Akande NO, Asani EO. Implementation of a hybrid feature selection algorithm for improving classification of mammograms. Journal of Engineering and Applied Sciences 2019;14:4419–29.
[23] Hossain SKM, Ema SA, Sohn H. Rule-Based Classification Based on Ant Colony Optimization: A Comprehensive Review. Applied Computational Intelligence and Soft Computing 2022;2022:2232000.
[24] Chalabi NE, Attia A, Bouziane A, Akhtar Z. Particle swarm optimization based block feature selection in face recognition system. Multimed Tools Appl 2021;80:33257–73.
[25] 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 2001;64:061907.
[26] Luo Y, Wang Y. A statistical time-frequency model for non-stationary time series analysis. IEEE Transactions on Signal Processing 2020;68:4757–72.
[27] Gruszczyńska I, Mosdorf R, Sobaniec P, Żochowska-Sobaniec M, Borowska M. Epilepsy identification based on EEG signal using RQA method. Adv Med Sci 2019;64:58–64.
[28] Furman Ł, Duch W, Minati L, Tołpa K. Short-time Fourier transform and embedding method for recurrence quantification analysis of EEG time series. Eur Phys J Spec Top 2023;232:135–49.
[29] Rostami M, Berahmand K, Forouzandeh S. A novel community detection based genetic algorithm for feature selection. J Big Data 2021;8:2.
[30] Maleki N, Zeinali Y, Niaki STA. A k-NN method for lung cancer prognosis with the use of a genetic algorithm for feature selection. Expert Syst Appl 2021;164:113981.
[31] Lambora A, Gupta K, Chopra K. Genetic algorithm-A literature review. 2019 international conference on machine learning, big data, cloud and parallel computing (COMITCon), IEEE; 2019, p. 380–4.
[32] Katoch S, Chauhan SS, Kumar V. A review on genetic algorithm: past, present, and future. Multimed Tools Appl 2021;80:8091–126.
[33] Sohail A. Genetic algorithms in the fields of artificial intelligence and data sciences. Annals of Data Science 2023;10:1007–18.
[34] van de Schoot R, Depaoli S, King R, Kramer B, Märtens K, Tadesse MG, Vannucci M, Gelman A, Veen D, Willemsen J. Bayesian statistics and modelling. Nature Reviews Methods Primers 2021;1:1.
[35] Khaleghi A, Mohammadi MR, Moeini M, Zarafshan H, Fadaei Fooladi M. Abnormalities of alpha activity in frontocentral region of the brain as a biomarker to diagnose adolescents with bipolar disorder. Clin EEG Neurosci 2019;50:311–8.
[36] Khaleghi A, Sheikhani A, Mohammadi MR, Nasrabadi AM. Evaluation of cerebral cortex function in clients with bipolar mood disorder I (BMD I) compared with BMD II using QEEG analysis. Iran J Psychiatry 2015;10:93.

[37] Khaleghi A, Zarafshan H, Mohammadi MR. Visual and auditory steady-state responses in attention-deficit/hyperactivity disorder. Eur Arch Psychiatry Clin Neurosci 2019;269:645–55.
[38] Moeini M, Khaleghi A, Mohammadi MR. Characteristics of alpha band frequency in adolescents with bipolar II disorder: a resting-state QEEG study. Iran J Psychiatry 2015;10:8.
[39] Webber CL, Marwan N. Recurrence quantification analysis. Theory and Best Practices 2015;426.
[40] Lopes MA, Zhang J, Krzemiński D, Hamandi K, Chen Q, Livi L, Masuda N. Recurrence quantification analysis of dynamic brain networks. European Journal of Neuroscience 2021;53:1040–59.
[41] Hosni SMI, Borgheai SB, McLinden J, Zhu S, Huang X, Ostadabbas S, Shahriari Y. Graph-based recurrence quantification analysis of EEG spectral dynamics for motor imagery-based BCIs. 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), IEEE; 2021, p. 6453–7.
[42] Murugappan M, Zheng BS, Khairunizam W. Recurrent quantification analysis-based emotion classification in stroke using electroencephalogram signals. Arab J Sci Eng 2021;46:9573–88.
[43] Timothy LT, Krishna BM, Nair U. Recurrence quantification analysis of mci eeg under resting and visual memory task conditions. Biomed Eng (Singapore) 2019;31:1950025.
[44] Chavan RM, RC P, Menon MS. Recurrence Quantification Analysis of EEG signals for Children with ASD. J Sci Ind Res (India) 2021;80:438–48.
[45] Nigam VP, Graupe D. A neural-network-based detection of epilepsy. Neurol Res 2004;26:55–60.
[46] Kannathal N, Choo ML, Acharya UR, Sadasivan PK. Entropies for detection of epilepsy in EEG. Comput Methods Programs Biomed 2005;80:187–94.
[47] Srinivasan V, Eswaran C, Sriraam and N. Artificial neural network based epileptic detection using time-domain and frequency-domain features. J Med Syst 2005;29:647–60.
[48] Sadati N, Mohseni HR, Maghsoudi A. Epileptic seizure detection using neural fuzzy networks. 2006 IEEE international conference on fuzzy systems, IEEE; 2006, p. 596–600.
[49] Subasi A. EEG signal classification using wavelet feature extraction and a mixture of expert model. Expert Syst Appl 2007;32:1084–93.
[50] Polat K, Güneş S. Classification of epileptiform EEG using a hybrid system based on decision tree classifier and fast Fourier transform. Appl Math Comput 2007;187:1017–26.
[51] Tzallas AT, Tsipouras MG, Fotiadis DI. Automatic seizure detection based on time-frequency analysis and artificial neural networks. Comput Intell Neurosci 2007;2007:080510.
[52] Srinivasan V, Eswaran C, Sriraam N. Approximate entropy-based epileptic EEG detection using artificial neural networks. IEEE Transactions on Information Technology in Biomedicine 2007;11:288–95.
[53] Polat K, Güneş S. Artificial immune recognition system with fuzzy resource allocation mechanism classifier, principal component analysis and FFT method based new hybrid automated identification system for classification of EEG signals. Expert Syst Appl 2008;34:2039–48.
[54] Polat K, Güneş S. A novel data reduction method: Distance based data reduction and its application to classification of epileptiform EEG signals. Appl Math Comput 2008;200:10–27.
[55] Ocak H. Automatic detection of epileptic seizures in EEG using discrete wavelet transform and approximate entropy. Expert Syst Appl 2009;36:2027–36.
[56] Guo L, Rivero D, Seoane JA, Pazos A. Classification of EEG signals using relative wavelet energy and artificial neural networks. Proceedings of the first ACM/SIGEVO Summit on Genetic and Evolutionary Computation, 2009, p. 177–84.