PAPER • OPEN ACCESS

Optimization of pathology diagnosis by applying chaos theory and fractal analysis to EEG-signal processing

View the article online for updates and enhancements.

IOP ebooks™

Bringing you innovative digital publishing with leading voices to create your essential collection of books in STEM research.

Start exploring the collection - download the first chapter of every title for free.

Optimization of pathology diagnosis by applying chaos theory and fractal analysis to EEG-signal processing

Vasily Kornilov, Maxim Ostroukhov and Andrey Dmitriev

National Research University Higher School of Economics, 33 Kirpichnaya St., 105187 Moscow, Russia

E-mail: a.dmitriev@hse.ru, vkornilov@hse.ru

Abstract. By 2018, there are more than 70 million people suffering from various forms of one of the most common neurological diseases – epilepsy. In fact, epipelsy is a central nervous system (neurological) disorder, manifesting itself in anomalous brain activity. Nowadays, most of the patients do not have the ability to foresee the onset of an attack in advance, which is due to the complex symptoms that are difficult to predict. One of the best way to analyze such kind of behaviour is electroencephalography (EEG). This research paper considers the problem of EEG epileptic seizures prediction from the point of the theory of nonlinear dynamical systems. According to the study, signals of cerebral cortex neural networks corresponds to multifractal nature which gives an opportunity to analyze changes between states in phase space during the abnormal electrical activity. Therefore, monitoring the preictal stage and early indication of an attack may help patients to avoid problems related to sudden seizures. This research will provide valuable information regarding the mechanism of epilepsy onset and introduce a prediction model using machine learning algorithm based on fractal analysis of RQA.

1. Introduction

Epilepsy is understood as a sudden excessive excitation of neurons in the cerebral cortex, the external manifestation of which consists in repeated convulsive attacks, manifested in violations of the nervous and cardiovascular systems of the patient, loss of consciousness, partial or complete loss of control over their own movements and, as a result, the inability to control their condition during the attack [1]. According to the survey published by the World Health Organization (WHO), the incidence or the number of new cases of epilepsy is increasing every year, ranging from 30 to 50 per 100,000 in high-income countries and 80-100 in low and middle-income countries, which form about 80% of all epilepsy cases.

By its complex nature, the disease is almost untreatable. Antispasmodic drugs which are aimed at reducing the probability of the seizure onset, often lead to severe side effects such as disruption of the pancreatic system, the cardiovascular system, etc. Moreover, after brain surgery, which can be helpful in some cases, 30% of people continue falling into periodic seizures. As a result, people suffering from this disease are constantly in a state of anxiety because of the possibility of being in a state of seizure at the wrong time from the medical and social points of view.

One of the most common ways to monitor current state of disease is to analyse brain activity within electroencephalography (EEG) due to its sensitivity and neurological nature of pathology. Considering EEG as time series research scientists distinguish four stages of disease: 1) interictal stage - stable brain activity within normal limits, 2) preictal stage – slight increase in brain activity, also called the

aura, 3) ictal stage - epileptic seizure, and 4) postictal stage - recovery from an attack, weakness, reduced brain and physical activity [2].

An important approach to support the patient's life cycle is to detect the presence of an epileptic attack. A positive aspect of this method is the timely notification of others in order to create the necessary conditions at the time of epileptic seizures. In other words, the main goal of this article is an early identification of the preictal stage.

2. Practical implementation

2.1. Dataset description
CHB-MIT Scalp EEG Database was chosen as a source data set [3, 4]. The data collected at Boston children's hospital consists of EEG records of patients with resistant epilepsy. Each patient was followed up for 1-3 days, some of which were given anticonvulsant drugs to determine the type of epilepsy and evaluate the candidate for surgery to remove the part of the brain that provoked the attack. All signals were recorded at the standard frequency of 256 Hz with 16-bit resolution. A distinctive feature of this set is the presence of epileptic seizures' periods labelling, which allows to track the dynamics of the brain activity on time series up to a second.

2.2. Initial fractal analysis
To begin with, it's needed to test the hypothesis that EEG data is an example of a nonstationary data set with long memory which is a simple indicator of correspondence to nonlinear dynamics.
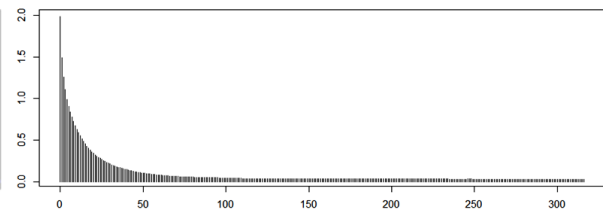To test the assumption, autocorrelation function (ACF) was constructed (figure 1):

$$R(\tau) = \frac{E[(X_t - \mu)(X_{t+\tau} - \mu)]}{\sigma^2} \#(1)$$

where τ is the lag at which the correlation is considered, μ-expected value of the quantity under consideration, σ^2 - the variance.
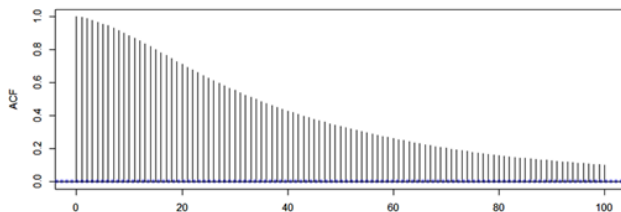
| **Figure 1.** Autocorrelation function. | **Figure 2.** Average Mutual Information. |

The value of the autocorrelation function has a rather weak downtrend and slowly converges to zero, which indicates the presence of characteristic features of the model with long memory and the possibility of applying the theory of nonlinear dynamic systems to the set.

An alternative metric for ACF is Average Mutual Information (AMI) - the average mutual information calculated through the entropy of two values:

$$I(X;Y) = H(X) - H(X|Y) = H(X) + H(Y) - H(X,Y)\#(2)$$

This metric (figure 2) gives an ACF-like result and demonstrates a slow decline to higher benefits, which gives additional reason to believe that the series has a long memory.

Moving to the stochastic characteristics of the time series to understand the structure of the available data correlation dimension is calculated by the formula:

$$C(r) = \frac{1}{\frac{m(m-1)}{2}} \sum_{i=0}^{m-2} \sum_{j=i+1}^{m-1} \theta\left(r - p(x_i, x_j)\right) \#(3)$$

where Θ is the Heaviside function: $\Theta(\alpha) = \begin{cases} 1, \alpha \geq 0, \\ 0, \alpha < 0, \end{cases}$

p – distance in n-dimensional phase space, m is the number of points on the attractor.

Due to the fact that there is no precise information about the phase space, the Takens' method [5] was used for the attractor's reconstruction. Eventually, correlation dimension equal to 8.5 was received.

To check the system for the presence of persistence, we calculate the Hurst index (H):

$$E\left[\frac{R(n)}{S(n)}\right] = Cn^{H}, n \to \infty \quad \#(4)$$

where $R(n)$ the scope of n-values of the time series, $S(n)$ – the standard deviation, n-the value of the time interval, $E[...]$ –expected value. The numerical estimation of the indicator in this case can be calculated as the slope coefficient of the line describing the change in relation to n, constructed in logarithmic coordinates. When the parameter H<0.5, the system is considered to be antipersistent, i.e. no trend tends to be saved, but on the contrary – it changes to the opposite. The resulting estimate-0.2427±0.1838 gives the opportunity to confidently state the absence of any persistence in the data and the presence of a long-term trend to vary between high and low values of the series, constantly changing the relative direction.
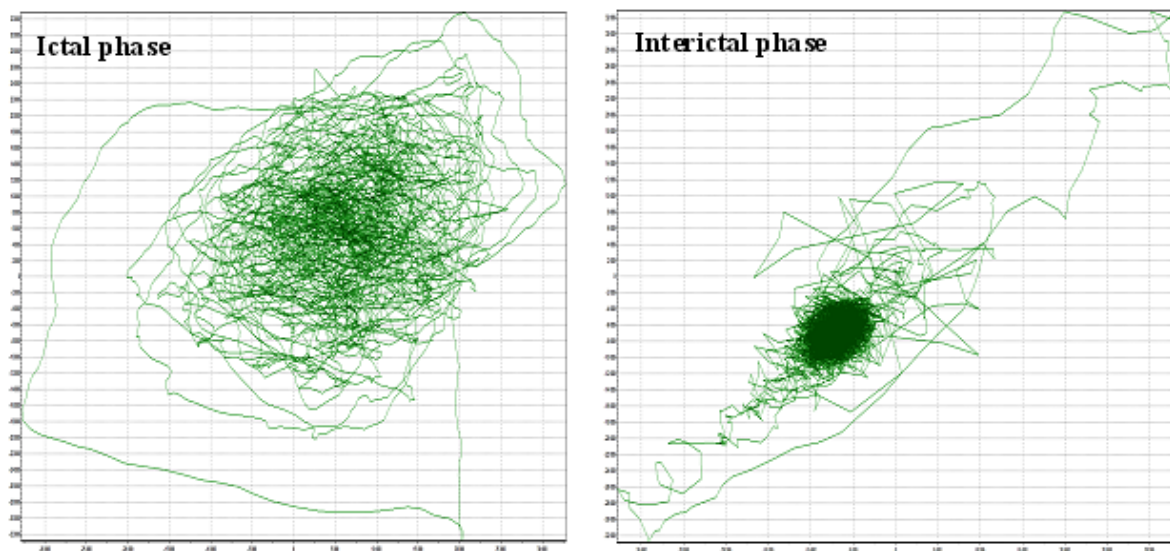
Figure 3. Attractors for the ictal and interictal stages.

Figure 3 shows the attractors built separately for the ictal phase and the phase of normal activity. To build such an attractor, windows were chosen, covering ~1 minute of EEG recording. The window size was chosen based on the average duration of the attack in the patient. Estimates of the resulting set of trajectories are presented in table 1.

Table 1. Comparison of fractal measures between ictal and interictal stages.

	Seizure	Dormant state
Correlation dimension	4.856	8.775
Phase space dimension	6	12

The obtained data shows the confirmation of formulated hypothesis [6] about significant decrease of correlation dimension during turbulence and allow us to conclude that the metrics of nonlinear dynamics for different phases of attack can vary.

2.3. Recurrence Quantification Analysis (RQA)

The main objective of this work was exposed to the application of the RQA approach to the prediction of seizures. It is important to note that RQA is a parametric method and depends on several factors: embedding dimension, time delay, ε radius of neighborhood, distance metric and the minimum length of the linear structures. The selection of each of these parameters can be devoted to the whole article, so for the purposes of this article the classical approaches to find optimal parameters were chosen.

- **Embedding dimension**: according to Takens' theorem [5], embedding dimension is $2m+1$, where m is the correlation dimension, will be sufficient to preserve the properties of the original phase space.
- **Delay time:** common heuristic for calculating the delay time is to select the first lag of the autocorrelation function whose value is zero or reduced to 1/e from the original value.
- **Radius ε of the neighbourhood:** the key parameter is the allowable radius, beyond which it can be assumed that the system has moved to another state.
- **Distance metric:** Euclidean distance is commonly used, but there are lots of different options for different tasks.
- **Minimum length of linear structures:** selected exclusively by manual selection for the purpose of high-quality visualization of statistics.

It is necessary to pay attention to the fact that the above parameters are selected on the basis of the interictal phase to find non-standard patterns of behavior during the transition of EEG into the phase of attack.

Using the approach:
- **Embedding dimension**: 12. It is calculated as the average of several calculated investments at different stages of rest.
- **Delay time**: 70.
- **The radius of the neighborhood** ε: 350.
- **Metric**: Euclidean distance.
- **Minimum length of linear structures**: 8.

The results of the calculated statistics are shown in figure 4.
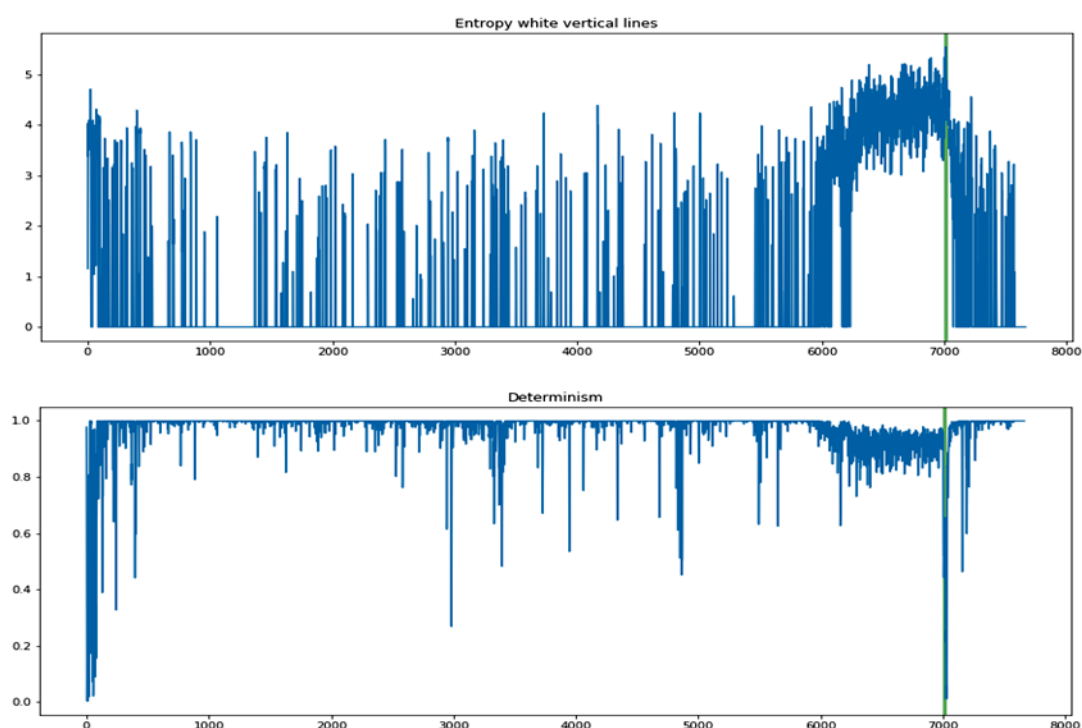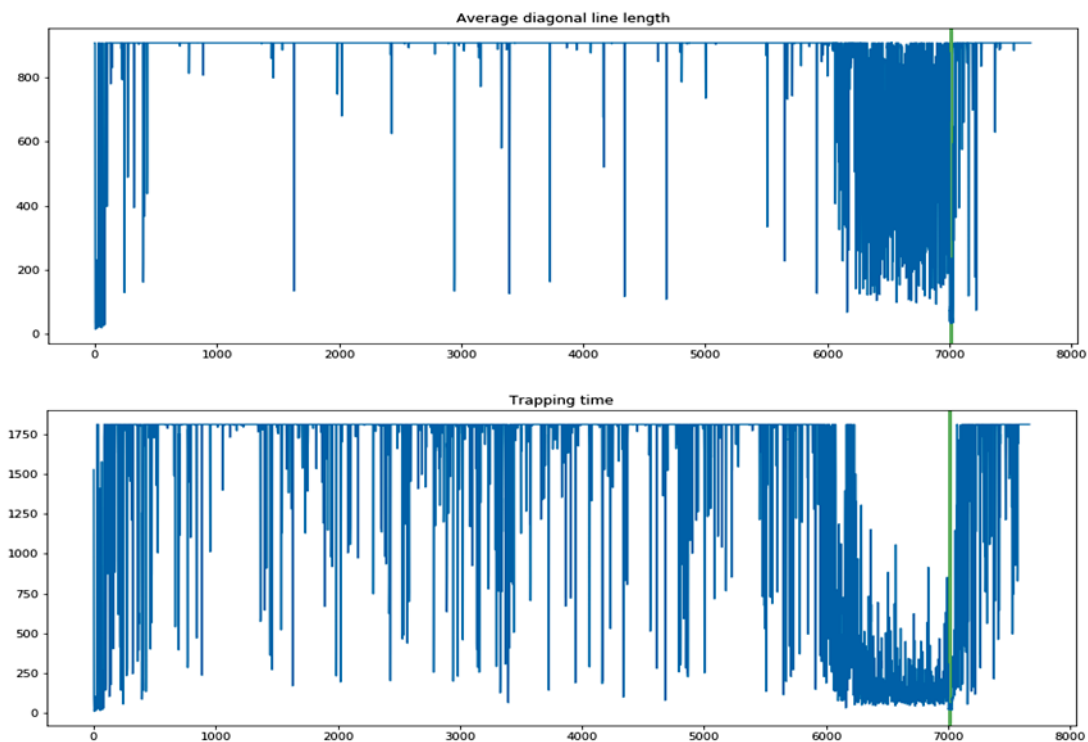
Figure 4. RQA metrics at 4 hour period.

The vertical line on the charts shows the seizure period. It is clearly visible that RQA reflects the changes in brain activity during preictal stage, which allows to confirm the hypothesis of the presence of structural changes in the state of the system obtained on the basis of RQA for a certain period of time before the attack. The practical value of the results is that, using such statistics, basic algorithm that detects the onset of an attack in advance can be implemented. Thus, it can be concluded that an approach based on the study of RQA metrics utilizing rolling window for some types of seizures can provide information about the presence of preictal phase. However, it should be noted that this approach cannot guarantee 100% accuracy, due to sufficient number of artifacts on in data which potentially leads to false positives.

In [7], published using the same data set, the authors propose an algorithm for classifying outliers based on the application of the support vector machine (SVM), whose purpose is to detect the preictal phase using spectral, spatial and temporal features for estimation. The model was evaluated using a leave-one-out cross-validation based on a data set consisting of a set of two-hour pairs so that one hour refers to the data in which the seizure is not recorded, and the other necessarily contains the seizure. Three metrics were chosen for measuring algorithm's quality:

- Sensitivity – percentage of predicted seizures, recall.
- Specificity–true negative rate.
- Delay- the delay between the expert evaluation of the onset of the attack and the evaluation of the algorithm.

The peculiarity of this implementation is that predictive model were trained on each person individually. In the context of this study, it is interested to make a comparison with an algorithm that would not depend on a particular patient. For the experiment integrity, the same SVM model was trained, formed according to the description of the article, supplemented with RQA metrics and tested on three patients. Since it is impossible to restore the implemented models exactly, a comparison with the local model was made, which gives the results not much worse than those given in the article. In table 2 you can see the comparative statistics of the applied model with and without new features.

Table 2. Evaluation of RQA features using SVM

	Basic SVM	SVM with RQA
Sensitivity	73%	84%
Specificity	88%	86%
Delay	6,3 s.	5,1 s.

The main achievement is a significant increase in the recall of the detection preictal stages and a slight decrease in the average delay in the determination of emissions. Analyzing the obtained results, it can be assumed that the model became more "confident" in classification which has been achieved at the expense of new quantitative characters. This "confidence" is also expressed in a slight decrease in the specificity of the model, since the increase in the "sensitivity" of the model always leads to an increase in the number of false positives.

3. Conclusions

Summing up the results of the research, it can be said that the approach based on the application of chaos theory to EEG is a good addition to existing studies, the quality of which is justified by the multifractal nature of the signal. At the same time, it is worth noting that the task of building a predictive model, expanding the existing one was achieved mainly through the approach based on the analysis of the quantitative characteristics of the Recurrence Plot with manually selected hyperparameters.

From the point of practical application, an important achievement of the study is a significant increase in the model recall, which gives rise to rethink the current methods of detecting epileptic seizures used on real people in devices implanted surgically. The main advantage of this work is not only the growth of different metrics, but also the proof that the electrical signals inside the human brain can be analyzed and predicted by fractal analysis, which is a positive outlook for further research in this area.

Obviously, this article is just an initial step towards the development of an adaptive predictive algorithm utilizing chaos theory. In this regard, there are several ways which the future improvements can be devoted to such as: analysis of dependencies between different parts of the brain, additional RQA metrics and at the same time using advanced machine learning models which can be a great basis improvement for the final results and a significant push for the implementation of such models to the patients' medical devices.

4. References

[1]   Singh A and Trevick S 2016 *J. Neurol. Clin.* **34** 837–47
[2]   Świderski B, Osowski S, Cichocki A and Rysz A 2007 *Adaptive and Natural Computing Algorithms* ed B Beliczynski, A Dzielinski, M Iwanowski and B Ribeiro (Warsaw: Springer) vol 4432 pp 373-81
[3]   Goldberger A L, Amaral L A, Glass L, Hausdorff J M, Ivanov P C, Mark R G, Mietus J E, Moody G B, Peng C K and Stanley H E 2000 Physiobank, physiotoolkit, and physionet: components of a new research resource for complex physiologic signals *J. Circulation.* **101** 215-20
[4]   Shoeb A H 2009 Application of Machine Learning to Epileptic Seizure Onset Detection and Treatment (Boston: Massachusetts Institute of Technology)
[5]   Takens F 1981 Detecting strange attractors in turbulence *Lecture Notes in Mathematics* vol. 898, ed D A Rand and L-S Young (New York: Springer-Verlag) pp 366-381
[6]   Babloyantz A and Destexhe A 1986 *J. Proc. Natl. Acad. Sci.* **83** 3513-7
[7]   Hamadene W, Peyrodie L and Seidiri H 2006 *8th international Conference on Signal Processing* (Beijing: IEEE)