Linear Random Process Model-Based EEG Classification Using Machine Learning Techniques

Mykhailo Fryz [a], Leonid Scherbak [b], Bogdana Mlynko [a], and Taras Mykhailovych [a]

[a] *Ternopil Ivan Puluj National Technical University, Ruska st. 56, Ternopil, 46001, Ukraine*
[b] *General Energy Institute of NAS of Ukraine, Antonovycha st. 172, Kyiv, 03150, Ukraine*

Abstract

The electroencephalogram (EEG) modelling and classification methods are very important in medical diagnostics and in creating complex information systems using brain-computer interface-based solutions for Industry 4.0. The mathematical model of the EEG signal has been presented in the paper in the form of a linear random process. The corresponding estimation procedure using the autoregressive model has been considered. The new informative features have been justified as the downsampled kernel of linear random process model representation. The comparative analysis of binary classification machine learning techniques has been performed based on autoregressive coefficients and new extracted informative features. The improvement of classification metrics has been shown.

Keywords [1]

Information system, electroencephalogram, signal, mathematical model, linear random process, autoregressive model, kernel, features extraction, estimation, binary classification.

1. Introduction

Electroencephalography (EEG) signals are recordings that capture the electrical potentials of the brain's activity [1, 2]. These measurements are generally non-invasive, using electrodes placed on the scalp, although intracranial electrodes may also be used. Examining the EEG data can yield important information about various brain disorders that affect a wide area of the brain.

Another great area of EEG analysis is related to the creation of complex information systems based on the brain-computer interface (BCI) concept [3, 4], including utilizing the classification of steady-state visual evoked potentials [5]. Several studies, related to EEG-based BCI applications for Industry 4.0 have been presented recently, including e.g. [6, 7]. The authors [6] categorized the possible industrial applications of safety at work, adaptive training, and device control.

Usually, EEG analysis for both medical applications and BCI applications includes mathematical modelling, features extraction, and classification methods. Linear models and techniques are significant in the field of EEG signal processing [1]. Discrete-time univariate and multivariate linear random sequences in the form of autoregressive moving average models are especially important and widely used [8, 9, 10]. Moreover, the continuous-time linear random process [11, 12] is the biophysically reasonable mathematical model of EEG representing the signal as the sum of many stochastically independent excitatory and inhibitory postsynaptic potentials which are generated by the pyramidal neurons in the cerebral cortex in response to the spike train [2] that can be considered as the Poisson arrival process. Another useful EEG signal model is a conditional linear random process considering the stochastic dependence between the postsynaptic potentials forming the EEG [2, 12]. All the above linear models can be analyzed using the characteristic function methods.

CITI'2023: 1st International Workshop on Computer Information Technologies in Industry 4.0, June 14–16, 2023, Ternopil, Ukraine
EMAIL: mykh.fryz@gmail.com (M. Fryz); prof_scherbak@ukr.net (L. Scherbak); mlynko@ukr.net (B. Mlynko); tarquas@gmail.com (T. Mykhailovych)
ORCID: 0000-0002-8720-6479 (M. Fryz); 0000-0002-1536-4806 (L. Scherbak); 0000-0003-0780-5365 (B. Mlynko); 0000-0002-8138-3642 (T. Mykhailovych)

Staying within the linear paradigm it is important to extract the new diagnostic features and find modern machine-learning classification techniques to improve the classification performance. Because of main elements of a linear random process is the kernel of its representation and generative white noise, the idea is to consider the first one as the source of informative characteristics for EEG analysis. We consider only binary classification problems in our paper.

The primary objective of the paper is to justify the informative features of EEG signals based on the kernel of its representation as a linear random process and to investigate the machine learning classification techniques based on the extracted features.

2. Methodology of modelling, features extraction and estimation

Continuous-time linear random process $\xi(\omega,t), t \in (-\infty,\infty)$, given on some probability space $\{\Omega, \mathfrak{F}, \mathbf{P}\}$ is defined using the following integral representation [12]:

$$\xi(\omega,t) = \int_{-\infty}^{\infty} \varphi(\tau,t) d\eta(\omega,\tau), \omega \in \Omega, t \in (-\infty,\infty), \tag{1}$$

where $\eta(\omega,\tau), \tau \in (-\infty,\infty)$, $\mathbf{P}(\eta(\omega,0)=0)=1$ is a stochastically continuous random process with independent increments;

$\varphi(\tau,t), \tau,t \in (-\infty,\infty)$ is a nonrandom function (kernel of representation (1)) such that the following conditions hold (the integral (1) is convergent in the mean-square sense under that conditions): $\int_{-\infty}^{\infty} |\varphi(\tau,t)|^p d\kappa_p(\tau) < \infty, \forall t, p = 1,2$, where $\kappa_p(\tau)$ is a cumulant function of the p^{th} order of given process with independent increments $\eta(\omega,\tau)$.

The linear random process (1) is the mathematical model of continuous-time EEG signal, representing it as a sum of large number of excitatory and inhibitory postsynaptic potentials, occurring at Poisson time moments. In the context of the model the process $\eta(\omega,\tau)$ is compound Poisson process, and $\varphi(\tau,t)$ represents the time-dependent properties of postsynaptic potentials.

The discrete-time counterpart of linear random process (LRP) (1) is denoted by $\xi_t(\omega), \omega \in \Omega, t \in \mathbf{Z}$ which is defined as a random sequence in the following form:

$$\xi_t(\omega) = \sum_{\tau=-\infty}^{\infty} \varphi_{\tau,t} \zeta_\tau(\omega), \tag{2}$$

where $\zeta_\tau(\omega)$, $\tau \in \mathbf{Z}$ is a sequence of independent infinitely divisible random variables (infinitely divisible strict-sense white noise) with finite expectation and variance $\mathbf{E}\zeta_\tau(\omega) = a_\tau$, $\mathrm{Var}[\zeta_\tau(\omega)] = \sigma_\tau^2$; $\varphi_{\tau,t}$, $\tau,t \in \mathbf{Z}$ is a nonrandom function (kernel) satisfying the following conditions:

$$\sum_{\tau=-\infty}^{\infty} \left|\varphi_{\tau,t} a_\tau\right| < \infty, \quad \sum_{\tau=-\infty}^{\infty} \left|\varphi_{\tau,t}\right|^2 \sigma_\tau^2 < \infty, \quad \forall t \in \mathbf{Z}$$

(the conditions are important to guarantee the mean-square convergence of the stochastic series (2)).

The expectation $\mathbf{E}\xi_t(\omega)$ and covariance function R_{t_1,t_2}, $t_1,t_2 \in \mathbf{Z}$ of linear random process (2) are represented by the following expressions:

$$\mathbf{E}\xi_t(\omega) = \sum_{\tau=-\infty}^{\infty} \varphi_{\tau,t} a_\tau,$$

$$R_{t_1,t_2} = \sum_{\tau=-\infty}^{\infty} \varphi_{\tau,t_1} \varphi_{\tau,t_2} \sigma_\tau^2.$$

If $\zeta_\tau(\omega)$, $\tau \in \mathbf{Z}$ is strict-sense stationary white noise, and the kernel $\varphi_{\tau,t}$ depends only from the difference of its arguments, that is $\varphi_{\tau,t} = \varphi_{t-\tau}$, then LRP (2) is strict-sense stationary and can be represented in the following form:

$$\xi_t(\omega) = \sum_{\tau=-\infty}^{\infty} \varphi_{t-\tau} \zeta_\tau(\omega) = \sum_{\tau=-\infty}^{\infty} \varphi_\tau \zeta_{t-\tau}(\omega), \quad \sum_{\tau=-\infty}^{\infty} \varphi_\tau^2 < \infty.$$

The expectation and covariance function of the stationary LRP are represented as:

$$\mathbf{E}\xi_t(\omega) = a\sum_{\tau=-\infty}^{\infty}\varphi_\tau, \quad R_s = \sigma^2\sum_{\tau=-\infty}^{\infty}\varphi_\tau\varphi_{\tau+s}, s\in\mathbf{Z}, \tag{3}$$

where $\mathbf{E}\zeta_\tau(\omega) = a$, $\mathrm{Var}\left[\zeta_\tau(\omega)\right] = \sigma^2$.

Among linear random processes the autoregressive models are the most important for applied problems of signal processing, especially when it is necessary to identify, predict, simulate, or classify the stochastic signals [13, 14].

In most important for practical application cases, a stationary discrete-time linear random process can be represented as a linear stochastic difference equation which is known as autoregressive moving average model of the order (p,q):

$$\xi_t(\omega) = -\sum_{k=1}^{p}a_k\xi_{t-k}(\omega) + \sum_{k=0}^{q}b_k\zeta_{t-k}(\omega), \quad t\in\mathbf{Z}, \tag{4}$$

where $a_k, k=\overline{1,p}$; $b_k, k=\overline{0,q}$ are the real coefficients.

If $q=0$ and $b_0=1$ we obtain the autoregressive (AR) model of the order p, which is represented in the following form:

$$\xi_t(\omega) = -\sum_{k=1}^{p}a_k\xi_{t-k}(\omega) + \zeta_t(\omega), \quad t\in\mathbf{Z}, \tag{5}$$

Autoregressive model is stationary if the complex roots z_k, $k=\overline{1,p}$ of the equation $1+a_1z^{-1}+a_2z^{-2}+\ldots+a_pz^{-p}=0$ (where z is complex variable) satisfy the conditions $|z_k|<1$, $k=\overline{1,p}$.

It is well known [13, 14] that covariance function of the stationary autoregressive model satisfies the following recurrent relation:

$$R_\tau = \begin{cases} -\sum_{k=1}^{p}a_kR_{-k} + \sigma^2, & \tau=0, \\ -\sum_{k=1}^{p}a_kR_{\tau-k}, & \tau>0. \end{cases}$$

Representing the above equation for $\tau=\overline{0,p}$ the system of linear equations is obtained, the matrix expression of the system is called Yule-Walker equations for AR model:

$$\begin{pmatrix} R_0 & R_{-1} & R_{-2} & \cdots & R_{-p} \\ R_1 & R_0 & R_{-1} & \cdots & R_{-p+1} \\ R_2 & R_1 & R_0 & \cdots & R_{-p+2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ R_p & R_{p-1} & R_{p-2} & \cdots & R_0 \end{pmatrix}\begin{pmatrix} 1 \\ a_1 \\ a_2 \\ \vdots \\ a_p \end{pmatrix} = \begin{pmatrix} \sigma^2 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}. \tag{6}$$

The Yule-Walker equations system (6) is the base for the EEG features extraction and estimation using the model of a linear random process. And autoregressive model coefficients are used successfully in the problems of EEG classification.

Our idea is to use the samples of the kernel of stationary discrete-time LRP representation of EEG, because, as it follows from the biophysical reasoning, the kernel represents the integral properties of the postsynaptic potentials forming EEG signal.

The general methodology of EEG analysis then consists of the following steps:

- EEG signal acquisition and preprocessing,
- order estimation of AR model [15],
- covariance function R_τ, $\tau=\overline{0,p}$ estimation,
- solving the linear equations (6) to find the AR coefficients,
- estimating the samples of the kernel of LRP representation of EEG signal as an impulse response of corresponding linear stationary recursive filter with parameters equal to coefficients of autoregressive model.

The extracted and estimated parameters can be used then in the algorithms of statistical hypotheses testing or machine learning classification [16]. The last approach is considered in the next section of the paper. We evaluate and compare the performance of binary classification algorithms for the case of using AR parameters as informative features versus set of samples of the kernel of LRP representation while keeping the same dimensionality of the features space.

3. Results

Several binary classification techniques have been compared using the dataset which is described in [17]. The length of each EEG realization is 23.6 sec, and it consists of 4097 samples. The realizations have passed the wide sense stationary test, and the eye movement and muscle activity artefacts have been removed. The binary classification tests have been analyzed based on healthy persons with open eyes (Class A) and closed eyes (Class B). The dataset consists of 100 elements in each of the classes. The representatives of Class A and Class B have been shown in Figure 1.

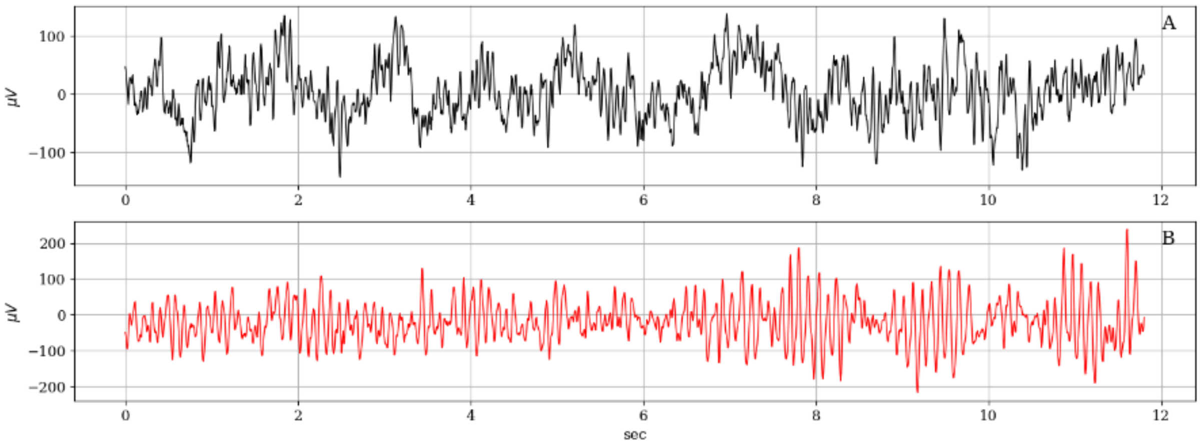

Figure 1: Open eyes EEG (A) and closed eyes EEG (B)

The covariance function estimations of open eyes EEG and closed eyes EEG have been represented in Figure 2. The above realizations have been utilized to calculate the estimations.

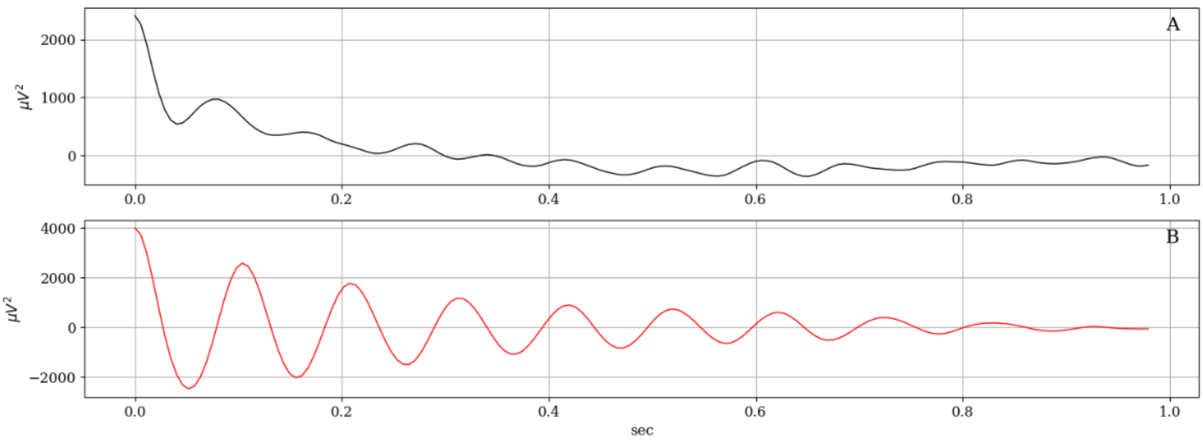

Figure 2: Covariance function estimations of open eyes EEG (A) and closed eyes EEG (B)

The kernel of EEG representation in the form of discrete-time linear random process has been estimated as the impulse response of corresponding recursive filter related to autoregressive model of the order 4. The coefficients of autoregressive model have been estimated using Yule-Walker approach. The estimations (based on realizations which have been shown in the Figure 1) of the kernel for open eyes EEG and closed eyes EEG have been represented in Figure 3.

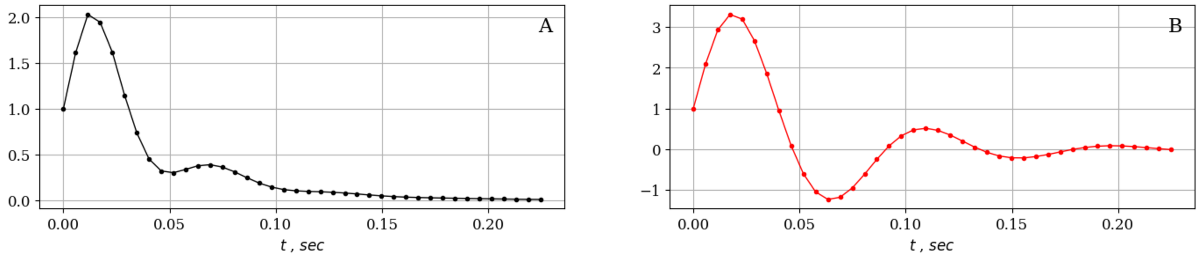

Figure 3: Kernel estimations of LRP representation of open eyes EEG (A) and closed eyes EEG (B)

The performance measures of different binary classification algorithms [18, 19, 20] based on AR parameters as informative features have been calculated using a 10-fold cross-validation approach and represented in Table 1. The best results are highlighted.

Table 1
Performance of the classifiers using AR parameters

Model	Accuracy	Recall	Precision	F1 score
Quadratic Discriminant Analysis	0.9286	0.9143	0.9446	0.9261
Linear Discriminant Analysis	0.9071	0.9143	0.9210	0.9092
Extra Trees Classifier	0.9071	0.9143	0.9121	0.9066
Random Forest Classifier	0.8929	0.9143	0.8918	0.8948
Ada Boost Classifier	0.8929	0.9143	0.8899	0.8941
Gradient Boosting Classifier	0.8857	0.9143	0.8821	0.8890
K Neighbors Classifier	0.8571	0.8143	0.9127	0.8455

The performance measures of the same classifiers based on the set of kernel samples $\{\varphi_1, \varphi_5, \varphi_9, \varphi_{13}\}$ as informative features have been calculated using a 10-fold cross-validation approach and represented in Table 2. The best results are highlighted.

Table 2
Performance of the classifiers using downsampled kernel

Model	Accuracy	Recall	Precision	F1 score
Quadratic Discriminant Analysis	0.8643	0.7714	0.9500	0.8471
Linear Discriminant Analysis	0.8143	0.7143	0.9086	0.7833
Extra Trees Classifier	0.9357	0.9429	0.9403	0.9354
Random Forest Classifier	0.9000	0.9143	0.9000	0.9032
Ada Boost Classifier	0.8929	0.8857	0.9300	0.8978
Gradient Boosting Classifier	0.9143	0.9143	0.9264	0.9145
K Neighbors Classifier	0.9214	0.8429	1.0000	0.9128

Comparing the above results, we can conclude that Extra Trees Classifier based on the second features set outperforms the other considered algorithms based on the first features set.

4. Discussion

The analysis of the above results shows that the kernel of linear random process representation of EEG signal can be used as an informative feature in the problems of its binary classification using machine learning techniques. We used the number of kernel samples equal to the order of the corresponding autoregressive model to compare the classification performance of different feature sets of the same dimensionality. The study of the dependence of classification accuracy on the number of informative features is a goal of future research. Still, we suppose that the dimensionality of the feature space should be equal to the order of the autoregressive model, but the kernel samples included in this space should be justified.

The main result of our research is the justification of new information features and corresponding binary classifier (the extra trees classifier) that can be used in medical diagnostics and in the problems of brain-computer interface based on EEG analysis. The results can be further extended to the alternative case of the autoregressive process with random coefficients [2], taking into account the nonlinear dynamics of EEG.

Being an ensemble method, the extra trees classifier combines predictions from multiple decision trees to make the final classification. This ensemble approach improves generalization and reduces the risk of overfitting, leading to better performance and more reliable results. There are also other advantages, such as fast training speed, efficient parallelization, and robustness to noisy data. It is important in the context of this paper that the extra trees classifier provides a measure of feature importance, gaining insights into the relevance and contribution of each feature to the classification task. This information can be useful for future research for feature selection, identifying the most influential kernel samples.

5. Conclusion

Methodology of features extraction and their estimation for EEG analysis based on the mathematical model in the form of linear random process has been considered. The set of informative characteristics has been identified consisting of the samples of the kernel of linear random process.

The new set of informative features has been compared with the traditional classification characteristics which are the autoregressive coefficients. Several binary classification machine learning algorithms have been analyzed for the task of classification of open eyes EEG and closed eyes EEG. Extra Trees Classifier has appeared to perform the best with the set of new informative features.

The prospective research should be related to the formal justification of the task of certain samples selection of the kernel which should be used as the classification features.

6. References

[1] S. Sanei, J. A. Chambers, EEG Signal Processing and Machine Learning, Wiley, 2021.

[2] M. Fryz, Conditional linear random process and random coefficient autoregressive model for EEG analysis, Proceedings of the 2017 IEEE First Ukraine Conference on Electrical and Computer Engineering (UKRCON), IEEE, Kyiv Ukraine, 2017, pp. 305–309. doi: 10.1109/UKRCON.2017.8100498.

[3] D. J. McFarland, J. R. Wolpaw, EEG-based brain–computer interfaces, Curr. Opin. Biomed. Eng. 4 (2017) 194–200. doi:10.1016/j.cobme.2017.11.004.

[4] S. Lupenko, R. Butsiy, N. Shakhovska, Advanced Modeling and Signal Processing Methods in Brain-Computer Interfaces Based on a Vector of Cyclic Rhythmically Connected Random Processes, Sensors 23.2 (2023). doi:10.3390/s23020760.

[5] M. Stadnyk, M. Fryz, L. Scherbak, The feature extraction and estimation of a steady-state visual evoked potential by the Karhunen-Loeve expansion, Eastern-European Journal of Enterprise Technologies 1.4 (2017) 56–62. doi:10.15587/1729-4061.2017.91861.

[6] K. Douibi, S. Le Bars, A. Lemontey, L. Nag, R. Balp, G. Breda, Toward EEG-Based BCI Applications for Industry 4.0: Challenges and Possible Applications, Front. Hum. Neurosci. 15 (2021). doi:10.3389/fnhum.2021.705064.

[7] D. Schmidt, J. Villalba Diez, J. Ordieres-Meré, R. Gevers, J. Schwiep, M. Molina, Industry 4.0 Lean Shopfloor Management Characterization Using EEG Sensors and Deep Learning, Sensors 20.10 (2020). doi:10.3390/s20102860.

[8] Y. Zhang, B. Liu, X. Ji, D. Huang, Classification of EEG Signals Based on Autoregressive

Model and Wavelet Packet Decomposition, Neural Process. Lett. 45.2 (2017) 365–378. doi: 10.1007/s11063-016-9530-1.

[9] Z. Liu, L. Wang, S. Xu, K. Lu, A multiwavelet-based sparse time-varying autoregressive modeling for motor imagery EEG classification, Comput. Biol. Med. 155 (2023) 106196. doi: 10.1016/j.compbiomed.2022.106196.

[10] C.-S. Ouyang, R.-C. Yang, C.-T. Chiang, R.-C. Wu, L.-C. Lin, EEG autoregressive modeling analysis: A diagnostic tool for patients with epilepsy without epileptiform discharges, Clin. Neurophysiol. 131.8 (2020) 1902–1908. doi:10.1016/j.clinph.2020.04.172.

[11] M. Fryz, Mixing property and ergodicity of linear random processes, Proceedings of the 2009 IEEE International Workshop on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications, IEEE, Rende Italy, 2009, pp. 343-346. doi: 10.1109/IDAACS.2009.5342967.

[12] V. Babak, A. Zaporozhets, Y. Kuts, M. Myslovych, M. Fryz, L. Scherbak, Models and Characteristics of Identification of Noise Stochastic Signals of Research Objects, in: Proceedings of the 2nd International Workshop on Information Technologies: Theoretical and Applied Problems (ITTAP 2022), Ternopil Ukraine, 2022, pp. 349-362. URL: https://ceur-ws.org/Vol-3309/paper22.pdf

[13] G. E. P. Box, G. M. Jenkins, G. C. Reinsel, G. M. Ljung, Time Series Analysis: Forecasting and Control, 5th ed., Wiley, 2015.

[14] P. J. Brockwell, R. A. Davis, Introduction to Time Series and Forecasting, 3rd ed., Springer Cham, 2016.

[15] S. Ross, Introduction to Probability Models, 12th ed., Academic Press, 2019.

[16] J. J. Li, X. Tong, Statistical Hypothesis Testing versus Machine Learning Binary Classification: Distinctions and Guidelines, Patterns 1.7 (2020) 100115. doi:10.1016/j.patter.2020.100115.

[17] R. G. Andrzejak, K. Lehnertz, F. Mormann, C. Rieke, P. David, C. E. Elger, 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 64.6 (2021) 061907. doi: 10.1103/PhysRevE.64.061907.

[18] A. Tharwat, Classification assessment methods, Appl. Comput. Informatics 17.1 (2021) 168–192. doi:10.1016/j.aci.2018.08.003.

[19] S.-C. Huang, T.-H. Le, Chapter 3 - Binary classification problem, in: Principles and Labs for Deep Learning, S.-C. Huang, T.-H. Le, Eds., Academic Press, 2021. doi:10.1016/B978-0-323-90198-7.00003-3.

[20] T. Marwala, W. T. Mongwe, R. Mbuvha, 2 - Sampling benchmarks and performance metrics, in: Hamiltonian Monte Carlo Methods in Machine Learning, T. Marwala, W. T. Mongwe, R. Mbuvha, Eds., Academic Press, 2023. doi:10.1016/B978-0-44-319035-3.00014-8.