# Concept for epilepsy diagnostics with combined unsupervised outlier detection and supervised data labeling

Vadim Grubov
*Baltic Center for Neurotechnology and Artificial Intelligence,*
*Immanuel Kant Baltic Federal University,*
Kaliningrad, Russia
vvgrubov@gmail.com

***Abstract*—In the present work we proposed a concept for epileptic seizure detection based on combination of unsupervised and supervised machine learning algorithms. The first step in the method suggests preliminary EEG marking with unsupervised outlier detection based on extreme value theory. The second step includes more traditional supervised machine learning classifier to separate epileptic seizures from artifacts.**

***Index Terms*—epilepsy, electroencephalogram, machine learning, unsupervised, extreme value, supervised**

Epilepsy is one of the most common neurological diseases [1]. It is characterized by sudden recurrent seizures, that vary greatly in frequency of appearance and severity of symptoms [2], [3]. Commonly seizures manifest themselves as absences or periods of vigorous shaking, which leads to a state of incapacity dangerous for both patients and surrounding people. Additionally, epileptic patients are predisposed to cognitive and behavioral deficits [4]. Effects of epilepsy have a heavy impact on patient's everyday life, so antiepileptic treatment becomes crucial. One of the common approaches is antiseizure medicines, that can help for up to 70% of patients [5]. However, the treatment is more efficient when started at early stages of epilepsy, which leads to increasing importance of proper and early diagnostics.

Common approach to diagnostics is electroencephalogram (EEG) study, in which the patients are monitored for a prolonged period of time to detect and analyze epileptiform activity [6]. The method is fairly reliable, but requires much time and effort due to manual EEG data deciphering. Various brain conditions may lead to occurance of epileptic seizures, which includes injury, stroke, congenital disabilities, etc., but exact reason usually remains unknown [7]–[9]. This leads to high variability of epileptic activity and additionally complicates diagnostics. Thus, this area of study is in dire need of automated and reliable methods for diagnostics.

New methods for automatic diagnosis of epileptic activity based on wavelet analysis, including adaptive wavelets, have been proposed previously [10]–[13]. However, one of the more promising approaches to epilepsy diagnostics is machine learning (ML) [14]. Application of ML to this task usually results in development of classifiers capable to separate two classes: "seizures" and "non-seizures" (normal activity) [15]. There are two general categories of ML classifiers: supervised and unsupervised [14]. Supervised ML algorithms aim to learn the distinctive features of epileptic seizures to detect them on EEG signals. These classifiers undergo training using the pre-labeled data of some patients before labeling the data from a new patient [16]. Unsupervised ML algorithms, on other hand, use unlabeled data and perform its clusterization [17].

Review shows that the majority of existing approaches to epilepsy diagnostics rely on supervised ML algorithms [18]. While these classifiers commonly demonstrate high performance, they can suffer from certain limitations. Rare nature of seizures results in class imbalance when the number of examples for "non-seizures" class is much higher than for "seizures" class. One way to do it is to perform artificial class balancing for the training data. Alternatively, class imbalance can be managed by constructing feature space with long distances between classes. However, pushing this concept too far may result in overfitting, when the algorithm performs well on the training data but fails to classify the data of the other patients. To counter this one can consider a subspace of features with biomarkers of seizures common for the most patients. These reasonings also lead us to the problem of features interpretability which commonly occurs in ML.

In our recent studies we tried to address all aforementioned issues by using unsupervised ML methods. We evaluated unsupervised algorithms for the task of epilepsy diagnostics [19]–[21]. We specifically aimed to apply outlier detection techniques. This choice was supported by earlier findings: epileptic seizures on EEG can be described in terms of extreme value theory [9], [22]. We tested multiple outlier detection algorithms, including one-class support vector machine [23], k-nearest neighbors [24], local nearest neighbors distance [25], local outlier factor [26], isolation forest [27]. Features for ML algorithms were also based on our previous results. We showed that extreme behavior for seizures occur in certain

This research was funded by Academic Leadership Program PRIORITY'2030 of the Ministry of Science and Education of Russian Federation, the President program (grants MK-2603.2022.1.6 and NSh-589.2022.1.).

979-8-3503-0574-6/23/$31.00 ©2023 IEEE

frequency range, where epileptic EEG differs from the normal one [28]. Thus, we tested different features based on wavelet spectra of EEG signals. We were able to achieve sensitivity of $\sim 77\%$ and accuracy of $\sim 13\%$ for the dataset of 83 epileptic patients [20]. In terms of sensitivity this result is acceptable for unsupervised ML algorithm, but accuracy is rather low. We concluded that this approach can be used in clinical decision support system, where ML algorithm labels the data first and then it is checked by the human who makes final decision [19].

However, fully automated seizure detection is an attractive idea. In such system the described unsupervised algorithm can be used as a first step for data preparation. Indeed, if we consider results of classification from earlier, we can see a peculiar detail: there is noticeable number of patients with both $100\%$ and $0\%$ sensitivity. The patients from the dataset have prolonged recordings with several (up to 5) epileptic seizures. Sensitivity of $100\%$ suggests that all seizures were detected, while sensitivity of $0\%$ means the opposite — all seizures were missed. Coexistence of these two situations implies that the data is somehow different for cases with $100\%$ and $0\%$ sensitivity. Additionally, low accuracy suggests large number of false positive detections, that are also outliers as much as epileptic seizures. Both of these problems may be due to EEG contamination with artifacts or inherent differences in features of epileptic patterns. This situation can be addressed with subsequent application of supervised ML algorithms, that, as we mentioned earlier, excel in separating different classes.

We evaluated supervised algorithms for the task of epilepsy diagnostics by applying classification to the initial EEG data, however, the obtained results were comparable to the ones for unsupervised algorithms [29]. Thus, we suggest that the supervised algorithm should be used as a second step after the unsupervised algorithm. In this case the unsupervised algorithm performs data preparation and pre-labeling, detecting all outliers in the data. This also should eliminate the class imbalance issue. Then this new dataset of outliers is used to train supervised algorithm to separate actual epileptic seizures from artifacts and other types of outliers. At this step great attention should be paid to feature selection procedure — "seizures" and "non-seizures" should have some complicated differences in feature space, since unsupervised algorithm wasn't able to separate them. We believe this combined approach can become the first step in transition from clinical decision support system to fully automated diagnostics system.

## REFERENCES


[1] E. Beghi, "The epidemiology of epilepsy," *Neuroepidemiology* **54**(2), pp. 185–191, 2020.

[2] R. S. Fisher, C. Acevedo, A. Arzimanoglou, A. Bogacz, J. H. Cross, C. E. Elger, J. Engel Jr, L. Forsgren, J. A. French, M. Glynn, *et al.*, "Ilae official report: a practical clinical definition of epilepsy," *Epilepsia* **55**(4), pp. 475–482, 2014.

[3] R. D. Thijs, R. Surges, T. J. O'Brien, and J. W. Sander, "Epilepsy in adults," *The Lancet* **393**(10172), pp. 689–701, 2019.

[4] G. Motamedi and K. Meador, "Epilepsy and cognition," *Epilepsy & Behavior* **4**, pp. 25–38, 2003.

[5] J. W. Sander, "The use of antiepileptic drugs—principles and practice," *Epilepsia* **45**, pp. 28–34, 2004.

[6] D. E. Friedman and L. J. Hirsch, "How long does it take to make an accurate diagnosis in an epilepsy monitoring unit?," *Journal of Clinical Neurophysiology* **26**(4), pp. 213–217, 2009.

[7] G. D. Hammer, S. J. McPhee, and M.-H. Education, *Pathophysiology of disease: an introduction to clinical medicine*, McGraw-Hill Education Medical, 2014.

[8] E. M. Goldberg and D. A. Coulter, "Mechanisms of epileptogenesis: a convergence on neural circuit dysfunction," *Nature Reviews Neuroscience* **14**(5), pp. 337–349, 2013.

[9] A. Pisarchik, V. Grubov, V. Maksimenko, A. Lüttjohann, N. Frolov, C. Marqués-Pascual, D. Gonzalez-Nieto, M. Khramova, and A. Hramov, "Extreme events in epileptic eeg of rodents after ischemic stroke," *The European Physical Journal Special Topics* **227**(7), pp. 921–932, 2018.

[10] E. Sitnikova, A. E. Hramov, A. A. Koronovsky, and G. van Luijtelaar, "Sleep spindles and spike–wave discharges in eeg: their generic features, similarities and distinctions disclosed with fourier transform and continuous wavelet analysis," *Journal of neuroscience methods* **180**(2), pp. 304–316, 2009.

[11] V. A. Maksimenko, S. Van Heukelum, V. V. Makarov, J. Kelderhuis, A. Lüttjohann, A. A. Koronovskii, A. E. Hramov, and G. Van Luijtelaar, "Absence seizure control by a brain computer interface," *Scientific Reports* **7**(1), p. 2487, 2017.

[12] G. van Luijtelaar, A. Lüttjohann, V. V. Makarov, V. A. Maksimenko, A. A. Koronovskii, and A. E. Hramov, "Methods of automated absence seizure detection, interference by stimulation, and possibilities for prediction in genetic absence models," *Journal of neuroscience methods* **260**, pp. 144–158, 2016.

[13] A. E. Hramov, V. A. Maksimenko, and A. N. Pisarchik, "Physical principles of brain–computer interfaces and their applications for rehabilitation, robotics and control of human brain states," *Physics Reports* **918**, pp. 1–133, 2021.

[14] M. Mohri, A. Rostamizadeh, and A. Talwalkar, *Foundations of machine learning*, MIT press, 2018.

[15] M. K. Siddiqui, R. Morales-Menendez, X. Huang, and N. Hussain, "A review of epileptic seizure detection using machine learning classifiers," *Brain informatics* **7**, pp. 1–18, 2020.

[16] K. D. Tzimourta, A. T. Tzallas, N. Giannakeas, L. G. Astrakas, D. G. Tsalikakis, P. Angelidis, and M. G. Tsipouras, "A robust methodology for classification of epileptic seizures in eeg signals," *Health and Technology* **9**, pp. 135–142, 2019.

[17] J. Birjandtalab, M. B. Pouyan, and M. Nourani, "Unsupervised eeg analysis for automated epileptic seizure detection," in *First International Workshop on Pattern Recognition*, **10011**, p. 100110M, International Society for Optics and Photonics, 2016.

[18] B. Abbasi and D. M. Goldenholz, "Machine learning applications in epilepsy," *Epilepsia* **60**(10), pp. 2037–2047, 2019.

[19] O. E. Karpov, V. V. Grubov, V. A. Maksimenko, S. A. Kurkin, N. M. Smirnov, N. P. Utyashev, D. A. Andrikov, N. N. Shusharina, and A. E. Hramov, "Extreme value theory inspires explainable machine learning approach for seizure detection," *Scientific Reports* **12**(1), pp. 1–14, 2022.

[20] V. Grubov, N. Smirnov, and S. Kurkin, "Classifier for detecting outliers in epileptic seizures," *Bulletin of the Russian Academy of Sciences: Physics* **87**(4), pp. 532–536, 2023.

[21] O. E. Karpov, M. S. Khoymov, V. A. Maksimenko, V. V. Grubov, N. Utyashev, D. A. Andrikov, S. A. Kurkin, and A. E. Hramov, "Evaluation of unsupervised anomaly detection techniques in labelling epileptic seizures on human eeg," *Applied Sciences* **13**(9), p. 5655, 2023.

[22] N. S. Frolov, V. V. Grubov, V. A. Maksimenko, A. Lüttjohann, V. V. Makarov, A. N. Pavlov, E. Sitnikova, A. N. Pisarchik, J. Kurths, and A. E. Hramov, "Statistical properties and predictability of extreme epileptic events," *Scientific reports* **9**(1), pp. 1–8, 2019.

[23] E. Burnaev and D. Smolyakov, "One-class svm with privileged information and its application to malware detection," in *2016 IEEE 16th International conference on data mining workshops (ICDMW)*, pp. 273–280, IEEE, 2016.

[24] Z. Zhang, "Introduction to machine learning: k-nearest neighbors," *Annals of translational medicine* **4**(11), 2016.

[25] W. Zheng, L. Zhao, and C. Zou, "Locally nearest neighbor classifiers for pattern classification," *Pattern recognition* **37**(6), pp. 1307–1309, 2004.

[26] M. M. Breunig, H.-P. Kriegel, R. T. Ng, and J. Sander, "Lof: identifying density-based local outliers," in *Proceedings of the 2000 ACM SIGMOD international conference on Management of data*, pp. 93–104, 2000.

[27] F. T. Liu, K. M. Ting, and Z.-H. Zhou, "Isolation forest," in *2008 eighth ieee international conference on data mining*, pp. 413–422, IEEE, 2008.

[28] O. E. Karpov, V. V. Grubov, V. A. Maksimenko, N. Utaschev, V. E. Semerikov, D. A. Andrikov, and A. E. Hramov, "Noise amplification precedes extreme epileptic events on human eeg," *Physical Review E* **103**(2), p. 022310, 2021.
[29] O. E. Karpov, S. Afinogenov, V. V. Grubov, V. Maksimenko, S. Korchagin, N. Utyashev, and A. E. Hramov, "Detecting epileptic seizures using machine learning and interpretable features of human eeg," *The European Physical Journal Special Topics* , pp. 1–10, 2022.