Epileptic EEG Labeling with Anomaly Detection Machine Learning Algorithms

Matvey Khoymov
Baltic Center for Neurotechnology and Artificial Intelligence, Immanuel Kant Baltic Federal University,
Kaliningrad, Russia
matthewkhoimov@gmail.com

Vadim Grubov
Baltic Center for Neurotechnology and Artificial Intelligence, Immanuel Kant Baltic Federal University,
Kaliningrad, Russia
Neuroscience Research Institute, Samara State Medical University,
Samara, Russia
vvgrubov@gmail.com

Semen Kurkin
Baltic Center for Neurotechnology and Artificial Intelligence, Immanuel Kant Baltic Federal University,
Kaliningrad, Russia
Neuroscience Research Institute, Samara State Medical University,
Samara, Russia
kurkinsa@gmail.com

Alexey Rozhdestvensky
National Research Medical Center for Psychiatry and Neurology
St Petersburg, Russia
a.s.rozhdestvensky@gmail.com

***Abstract*—In this paper we considered the application of unsupervised machine learning for labeling epileptic human electroencephalogram. We tested the performance of several algorithms specifically designed for anomaly detection.**

***Index Terms*—epilepsy, electroencephalogram, Machine Learning, anomaly detection.**

I. Introduction

Epilepsy is the one of the most common neurological diseases characterized by uncontrollable recurrent seizures [1]. Epileptic seizures are usually accompanied by uncontrolled convulsions or absence state, which can be dangerous for both the patient and the others [2]. Thus, epilepsy treatment, including early diagnostics, is a task of great importance in modern medicine. Epilepsy is commonly diagnosed manually by medical experts who review electroencephalogram (EEG) in search of seizure episodes [3]. Random and sparse nature of epileptic seizures leads to a frequent situation when hour-long EEG recording contains only minutes of epileptic activity. Therefore, diagnostics demands a lot of time and effort from the expert, and automation of this process is of great demand. It is widely assumed that the cause of the seizures is spontaneous synchronous activity of neurons in the brain [4], [5]. Knowledge on features of this activity can be used to develop approaches for reducing expert's workload in epilepsy diagnostics [6]–[8]. One way to do it is to implement machine learning (ML) techniques for automated data labeling and classification [9], [10]. Today many researchers use supervised ML algorithms for this task [11]. However, this approach has certain limitations, that include dependence on prelabeled data and high possibility of overfitting. In this work we consider another approach — unsupervised ML; we tested the performance of of several unsupervised ML algorithms and assess the possibility of their application in real medical practice.

This work was supported by the Academic Leadership Program "PRIORITY-2030". VG thanks the President program (MK-2603.2022.1.6). SK thanks the President program (MD-590.2022.1.2).

II. Materials and Methods

A. Machine learning methods

Supervised ML methods require data labeling to learn relationship between features and labels. Unsupervised methods usually aim to separate normal data from outliers, and thus these methods are called anomaly detection algorithms. Epileptic seizures can be considered as certain type of outliers [12], [13], which makes unsupervised ML methods a great choice for automatic labeling of epileptic EEG.

In this work five popular unsupervised algorithms were considered: One-Class Support Vector Machine (OCSVM), k-Nearest Neighbors (kNN), Local Nearest Neighbors Distance (LNND), Local Outlier Factor (LOF) and Isolation Forest (IF). OCSVM takes features to the higher dimensional space and constructs hyperplanes capable of separating normal data from outliers [14]. kNN calculates distances between each point and its k-nearest neighbors [15], it then labels the nth fraction of data points with the largest average distance to neighbors as outliers. LNND works like kNN, but it divides distance from point a to its nth nearest neighbor (point b) by the distance from b to its nth nearest neighbor [16]. LOF calculates density arrangement of points around each point and chooses data with lowest density [17]. IF separates data space by a randomly

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

selected value trying to isolate a specified percentage of data points [18].

B. Dataset

Database with EEG records was provided by the National Medical and Surgical Center named after N. I. Pirogov of the Russian Healthcare Ministry. All patients gave their consent to participate in the experiment. The recording was made during the daily activities of patients with purpose to confirm epileptic status. The record's duration varied from 8 to 84 hours according to patient conditions and number of detected seizures needed to confirm the diagnosis. All seizures occurred spontaneously. The data was prelabeled by medical expert. Database contains EEG data of 80 patients with diagnosed focal epilepsy.

EEG records were obtained from 25 channels arranged according to"10-20" layout. Raw signals were preprocessed by continuous wavelet transform (CWT): wavelet power was calculated and averaged in 60-second time intervals and across all EEG channels. CWT is a powerful instrument of time-frequency analysis of EEG signals and it can be used to extract certain feautrres from EEG [19]–[21]. In the resulting data each 60-second interval was considered as an event characterized by a vector of 300 values. These values correspond to the spectrum of wavelet power in the 1–30 Hz band divided into bins with a 0.1 Hz step. To reduce the complexity of the data, we additionally implemented principal component analysis (PCA) [22]. Then some feature types were chosen for consideration:

- Raw 30 Hz — 300 values vector;
- Raw 3 Hz — subvector of the first 30 values;
- Mean 30 Hz — the mean of 300 values;
- Mean 3 Hz — the mean of the first 30 values;
- PCA 30 Hz — PCA-based feature obtained after decomposing 300 values and considering 4 principal components that explain 90% of the variance
- PCA 3 Hz — PCA-based feature obtained after decomposing 30 values and considering 4 principal components that explain 98% of the variance

C. Models training

Each ML algorithm has hyperparameters — parameters that are chosen by user to control learning process and don't change during training. Optimal combination of hyperparameters may improve model results. Therefore, before models training the hyperparameters tuning using Grid Search method from Scikit learn library was applied to each model and feature type (i.e., this process was done 30 times). It is important to note that there was a class imbalance in the data (the number of normal data instances greatly exceeded the number of outliers). Thus, F1 score [23] was chosen as a function for assessing the optimality of the combination of hyperparameters and the quality of the models:

$$F1score = \frac{2 * precision * recall}{precision + recall} * 100\% \quad (1)$$

$$precision = \frac{TP}{TP + FP} * 100\% \quad (2)$$

$$recall = \frac{TP}{TP + FN} * 100\% \quad (3)$$

where TP – true positive, FP – false positive, FN – false negative.

After hyperparameters tuning, individual model was trained for each patient because of high EEG variability. Performance of method was calculated as averaged F1 score of 80 models.

III. Results and discussion

Results obtained after training are showed in Table I.

It is clearly seen that obtained F1 scores are relatively low. For a complete understanding of the results, it is better to consider the violin plot, which shows the distribution of F1 score values in the group of models that showed the best results (Figure 1). This plot demonstrates that there is an imbalance in distribution of F1 score values in the group. There is a considerable amount of models with both high $(0.6 - 0.8)$ and very low (~ 0) F1 scores, which ultimately leads to mediocre group-average F1 scores. The most likely reason is that all unsupervised methods are based on the distance between data points. EEG signal contains artifacts with amplitude that can significantly exceed an epileptic seizure, therefore, the model labels artifacts as outliers. Of course, artifacts can be removed from the data before training, but this requires time on the part of the expert, which contradicts the idea of our work and unsupervised approach in general.

It is curious that kNN, LNND and LOF methods have approximately the same principle of operation, however, there are much more LNND and LOF models with an F1 score of 1 than kNN. Perhaps the reason is that kNN uses the common distance between points, while LNND and LOF calculate the local distance. Consequently, it is easier to distinguish a seizure from normal data and a fracture of artifacts by its local distance to its nearest neighbors. This means that there are more anomalous points near the artifacts than near the seizures. Based on this, it can be assumed that after averaging the data in shorter time intervals, these models would be able to better distinguish seizures from artifacts. But this only applies to cases where the models were trained on data with a frequency range of 30 Hz. In other cases, LNND and LOF are inferior to kNN. Most likely, the 3 Hz range is ill-suited for separation seizures from artifacts.

Poor OCSVM results suggest that there is no hidden difference in the EEG data between seizures and other points. Therefore, the transfer of data to a higher dimension didn't give a significant improvement in performance.

IF is partly based on randomness when it chooses a point and a feature to divide the data space. Therefore, its F1 scores are also low.

IV. Conclusion

The best result was showed by LNND (0.338). On first glance, this score is low, which leads to assumption that in this moment unsupervised ML methods can't be applied for

TABLE I
AVERAGED VALUE OF F1 SCORE WITH CONFIDENCE INTERVALS OF TRAINED MODELS WITH OPTIMAL HYPERPARAMETERS (BOLD TEXT INDICATES THE BEST RESULTS OF THE MODELS).

Feature	**Algorithm**				
	OCSVM	*kNN*	*LNND*	*LOF*	*IF*
3 Hz Raw	0.305 ± 0.057	**0.316 ± 0.055**	0.281 ± 0.055	0.297 ± 0.055	**0.304 ± 0.057**
30 Hz Raw	**0.307 ± 0.060**	0.312 ± 0.060	0.331 ± 0.056	0.330 ± 0.054	0.271 ± 0.073
3 Hz mean	0.255 ± 0.071	0.282 ± 0.074	0.116 ± 0.040	0.278 ± 0.073	0.282 ± 0.074
30 Hz mean	0.245 ± 0.071	0.270 ± 0.073	0.135 ± 0.046	0.254 ± 0.053	0.270 ± 0.073
3 Hz PCA	0.273 ± 0.072	0.300 ± 0.075	0.250 ± 0.071	0.292 ± 0.075	0.276 ± 0.073
30 Hz PCA	0.300 ± 0.058	0.313 ± 0.077	**0.338 ± 0.077**	**0.331 ± 0.080**	0.304 ± 0.061

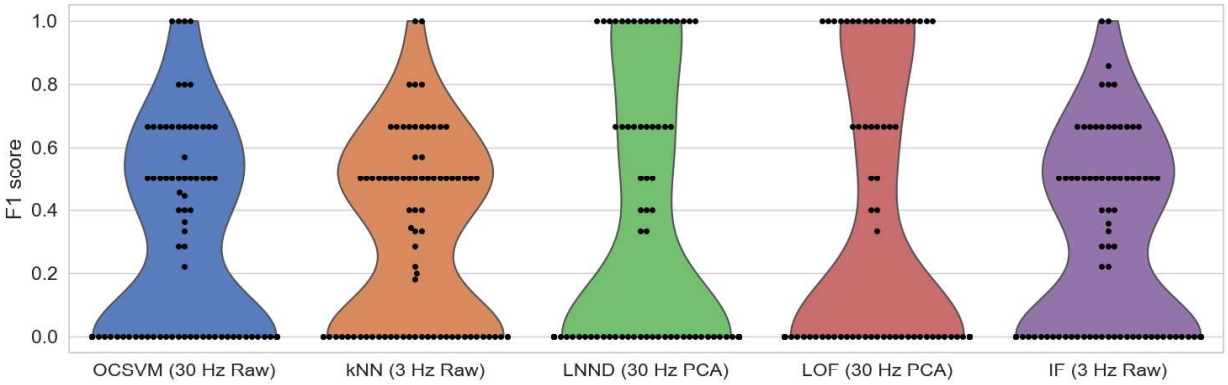

Fig. 1. Distribution of F1 score values in the group of models with the best average result.

raw epileptic EEG data labeling. However, presence of both high and low F1 scores in group (see Figure 1) allows to conclude, that there can be hidden differences between patients in the group. We assume that the results of ML labeling can be improved by finding a feature (or a combination of features) that would better characterize an epileptic seizure, separating it from normal data and artifacts.

Additionally, even the proposed approach can be implemented in clinical practice. F1 score is low in average, which limits possibility for fully automated EEG labeling. However, the fact that it works well for some patients, can justify its application in Clinical Decision Support System (CDSS). In CDSS ML algorithm performs data prelabeling, then this prelabeled data is cross-checked by the human expert, who makes the final decision. While such system still requires participation of the expert, it can be used to greatly (up to 95%) reduce the expert's workload [6].

REFERENCES

[1] E. Beghi, "The epidemiology of epilepsy," *Neuroepidemiology*, vol. 54, no. 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*, vol. 55, no. 4, pp. 475–482, 2014.

[3] W. O. Tatum IV, *Handbook of EEG interpretation*. Springer Publishing Company, 2021.

[4] 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*, vol. 7, no. 1, p. 2487, 2017.

[5] R. D. Thijs, R. Surges, T. J. O'Brien, and J. W. Sander, "Epilepsy in adults," *The Lancet*, vol. 393, no. 10172, pp. 689–701, 2019.

[6] 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.

[7] 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*, vol. 12, no. 1, p. 11474, 2022.

[8] 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*, vol. 13, no. 9, p. 5655, 2023.

[9] O. E. Karpov, E. N. Pitsik, S. A. Kurkin, V. A. Maksimenko, A. V. Gusev, N. N. Shusharina, and A. E. Hramov, "Analysis of publication activity and research trends in the field of ai medical applications: Network approach," *International Journal of Environmental Research and Public Health*, vol. 20, no. 7, p. 5335, 2023.

[10] 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*, vol. 918, pp. 1–133, 2021.

[11] 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*, vol. 232, no. 5, pp. 673–682, 2023.

[12] 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*, vol. 9, no. 1, p. 7243, 2019.

[13] 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*, vol. 103, no. 2, p. 022310, 2021.
[14] 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)*. IEEE, 2016, pp. 273–280.
[15] Z. Zhang, "Introduction to machine learning: k-nearest neighbors," *Annals of translational medicine*, vol. 4, no. 11, 2016.
[16] W. Zheng, L. Zhao, and C. Zou, "Locally nearest neighbor classifiers for pattern classification," *Pattern recognition*, vol. 37, no. 6, pp. 1307–1309, 2004.
[17] 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*, 2000, pp. 93–104.
[18] F. T. Liu, K. M. Ting, and Z.-H. Zhou, "Isolation forest," in *2008 eighth ieee international conference on data mining*. IEEE, 2008, pp. 413–422.
[19] A. Aldroubi and M. Unser, *Wavelets in medicine and biology*. Routledge, 2017.
[20] A. E. Hramov, A. A. Koronovskii, V. A. Makarov, V. A. Maximenko, A. N. Pavlov, and E. Sitnikova, *Wavelets in neuroscience*. Springer, 2021.
[21] H. Adeli, Z. Zhou, and N. Dadmehr, "Analysis of eeg records in an epileptic patient using wavelet transform," *Journal of neuroscience methods*, vol. 123, no. 1, pp. 69–87, 2003.
[22] R. Bro and A. K. Smilde, "Principal component analysis," *Analytical methods*, vol. 6, no. 9, pp. 2812–2831, 2014.
[23] C. Goutte and E. Gaussier, "A probabilistic interpretation of precision, recall and f-score, with implication for evaluation," in *Advances in Information Retrieval: 27th European Conference on IR Research, ECIR 2005, Santiago de Compostela, Spain, March 21-23, 2005. Proceedings 27*. Springer, 2005, pp. 345–359.