Joanna Chwał, Arkadiusz Banasik, Radosław Dzik, Konrad Duraj, Szymon Sieciński, and Ewaryst Tkacz

Hybrid Machine Learning Approach for Seizure Detection Using Hjorth Parameters in EEG Signals

https://doi.org/10.1515/cdbme-2025-0185

Abstract: Epilepsy represents a chronic non-communicable brain disease which affects about 70 million people worldwide. The condition presents with recurring seizures which produce brief periods of unmanageable movement. EEG serves as the primary diagnostic instrument for epilepsy diagnosis to date. We examined seizure EEG signals from Physionet database based on Hjorth parameters (Activity, Mobility, Complexity), k-Nearest Neighbors (k-NN), Random Forest, and Decision Tree. The results (between 0.9827 and 0.9999 of accuracy) demonstrate that Hjorth parameters can be used for detecting seizure episodes, despite class imbalance in the dataset. The best-performing approach was the hybrid system based on the Random Forest (1.0000 of accuracy).

Keywords: Hjorth Parameters; Seizure; Electroencephalography; Machine learning

1 Introduction

Epilepsy is a chronic neurological disorder affecting approximately 70 million individuals globally, characterized by recurrent seizures due to excessive neuronal electrical discharges. Seizures range from brief lapses of attention to severe convulsions, varying in frequency and intensity. Diagnosis typically requires two or more spontaneous seizures [1].

Joanna Chwał, Arkadiusz Banasik, Radosław Dzik, Konrad Duraj, Szymon Sieciński*, Ewaryst Tkacz, Academy of Silesia, Department of Clinical Engineering, Rolna 43, 40-555 Katowice, Poland, e-mail: joanna.chwal@akademiaslaska.pl; arkadiusz.banasik@akademiaslaska.pl; radoslaw.dzik@akademiaslaska.pl; konrad.duraj@akademiaslaska.pl; szymon.siecinski@akademiaslaska.pl; ewaryst.tkacz@wst.pl
Joanna Chwał, Silesian University of Technology, Faculty of Biomedical Engineering, Department of Medical Informatics and Artificial Intelligence, e-mail: joanna.chwal@polsl.pl
Szymon Sieciński, University of Luebeck, Institute of Medical Informatics, Ratzeburger Allee 160, 23562 Lübeck, Germany, e-mail: szymon.siecinski@uni-luebeck.de
Szymon Sieciński, AGH University of Cracow, Faculty of Computer Science, Kawiory 21, 30-055 Kraków, Poland, e-mail: ssiecinski@agh.edu.pl

Despite advances in imaging, electroencephalography (EEG) remains critical for epilepsy diagnosis and classification [2]. EEG data reflect changes in spatial and temporal patterns indicative of different brain states: preictal (preceding seizures), ictal (during seizures), postictal (after seizures), and interictal (between seizures). Accurate identification of preictal states is essential for seizure prediction [3].

Machine learning has significantly improved seizure prediction by applying noise reduction, feature extraction, and classification methods [4, 5]. The features include spectral power [6], phase locking value [7], bag-of-waves [8], and zero-crossing measures [9]. Support Vector Machines (SVM) have frequently been used to distinguish preictal from interictal phases effectively [10, 11].

Recently, Convolutional Neural Networks (CNNs), have been applied for EEG-based seizure prediction due to superior feature extraction capabilities [12–14]. However, for real-time, mobile, or low-power applications, Hjorth parameters are preferred due to the lower computational cost and comparable performance (up to 97.9% accuracy in seizure detection) [15, 16].

In this study we evaluated the classification of seizures based on Hjorth parameters (Activity, Mobility, Complexity) obtained from EEG signals and three machine learning techniques: k-Nearest Neighbors, Random Forest, and Decision Tree.

2 Material and Methods

2.1 Hjorth Parameters

Hjorth parameters, also known as normalized slope descriptors (NSDs), are statistical functions that represent EEG signal properties in both time and frequency domains that were introduced by Bo Hjorth in 1970 [17, 18]. They include activity, mobility, and complexity [17], which approximate signal activity and mean power, mean frequency. and the signal bandwidth, respectively, and are calculated as follows [17, 18]:

$$Activity = \sigma_x^2 \quad (1)$$

$$Mobility = \sqrt{\frac{\sigma_d^2}{\sigma_x^2}} = \frac{\sigma_d}{\sigma_x} \quad (2)$$

$$Complexity = \sqrt{\frac{\frac{\sigma_{dd}^2}{\sigma_d^2}}{\frac{\sigma_d^2}{\sigma_x^2}}} = \frac{\frac{\sigma_{dd}}{\sigma_d}}{\frac{\sigma_d}{\sigma_x}} \quad (3)$$

where ${\sigma_x}^2$ is the variance of the EEG signal x, σ_x is the standard deviation of x, σ_d is the standard deviation of the first derivative of x, and σ_{dd} represents the standard deviation of the second derivative of x.

To calculate the collection of Hjorth parameters, we used an optimal sliding window and stride based on the method described in [19].

2.2 EEG signal dataset

To assess the proposed model's performance, we utilized the CHB-MIT dataset, which was acquired in partnership with MIT University and Boston Children's Hospital and is publicly available. This data set includes EEG recordings from 23 people who had uncontrolled seizures [20–22].

This dataset comprises EEG recordings from pediatric patients with intractable seizures. The patients were monitored over several days after discontinuing anti-seizure medication to understand their seizures better and evaluate their suitability for surgical treatment.

All signals were recorded at 256 Hz with a 16-bit resolution. Most files include 23 EEG signals, although some contain 24 or 26. The recordings were conducted using the International 10-20 system for EEG electrode placement and nomenclature [20–22].

Every EEG signal (and channel, respectively) were analyzed, and the descriptors were calculated accordingly. Using the annotations provided by PhysioNet for the analyzed database, we labeled '1' for positions where a seizure occurred and '0' where there was no seizure.

2.3 Analysis methods

The statistical significance of the Mobility, Activity, and Complexity descriptors was then assessed using ANOVA testing. Next, we used three machine learning algorithms to further evaluate the data, namely k-Nearest Neighbors (k-NN), Random Forest, and Decision Tree. Their performance was evaluated in terms of their ability to effectively categorize and predict using the Hjorth parameters as cross-validation loss, accuracy, precision, recall, and F1 score. The dataset was split into training and validation sets with an 80%/20% ratio.

3 Results

Table 1 presents the p-values obtained from the ANOVA analysis.

Tab. 1: ANOVA Results (p-values)

Activity	***Mobility***	***Complexity***
1.4354e-65	1.4029e-32	7.9495e-54

As observed, all values are close to zero, indicating substantial differences among the groups. To analyze the results thoroughly, we have prepared box plots to show the statistical meaning of the results, which are shown in Fig. 1.

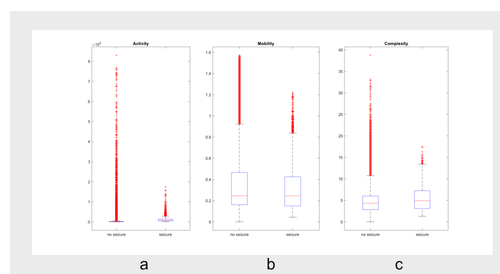

Fig. 1: Box plots of the Hjorth parameters. Graphs (a) presents Activity, (b) Moblility, and (c) Complexity, respectively.

Then, we implemented three classifiers to obtain final results (Fig. 2), showing confusion matrices.

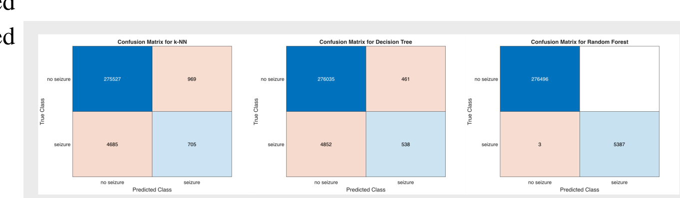

Fig. 2: Confusion Matrices. The section relates (a) to k-NN, (b) Decision Tree, and (c) Random Forest.

To ensure the validity of our findings, we implemented various metrics to evaluate the classifiers. The results are detailed in Table 2.

Tab. 2: Classifier Metrics

Metrics	Classifier		
	Decision Tree	***k-NN***	***Random Forest***
Cross-validation loss	0.0188	0.0201	<0.0001
Accuracy	0.9812	0.9799	1.0000
Precision	0.5385	0.4211	1.0000
Recall	0.1308	0.0998	0.9994
F1 score	0.1996	0.1684	0.9997

4 Discussion and Conclusions

In this study, we evaluated three machine learning models (k-NN, Random Forest, Decision Tree) for detecting seizures in EEG signals based on Hjorth parameters. Our results for the Random Forest (accuracy of 1.0000, precision of 1.0000, recall of 0.9994, and F1 score of 0.9997) show an improvement over approaches based on multi-view CNN (sensitivity of 0.93) [23], iterative channel selection, Haar wavelet and Support Vector Machine (accuracy of 0.94 and sensitivity of 0.96) [13], CNN + Bi-LSTM (bi-directional Long Short-Term Memory) and DCAE (deep convolutional autoencoder) + Bi-LSTM (all metrics ranging between 0.9960 and 0.9972) [24] and Graph Attention Network (GAT) and Temporal Convolutional Network (TCN) using the same dataset (accuracy of 0.9871, specificity of 0.9835, and a recall of 0.9907) [25], and wavelet with 1D convolutional layers and multi-head attention mechanism (0.9983 of accuracy) [26].

Rizal et al. report in [15] similar results (the best accuracy of 0.995) to those reported in our study using a similar approach. However, they rescaled the EEG signal into new signals using the coarse-graining procedure, where the new signal is the average of the closest sequential samples. Subsequently, the Hjorth parameters were calculated on these new signals. These features were later used for classification using different Support Vector Machines (SVMs).

Combining Hjorth parameters of the EEG signal with machine learning techniques for epilepsy seizure detection yielded superior results, especially with Random Forest method, forming a hybrid system with a low computational cost.

However, the precision of 0.1308 and 0.4211, recall of 0.1308 and 0.0998, and F1 score between 0.1996 and 0.1684 for Decision Tree and k-NN, respectively, despite an accuracy over 0.97 indicate the significant class imbalance between true positive (label 1) and true negative (label 0).

Based on the evaluation, the Random Forest classifier outperformed both the k-NN and Decision Tree algorithms. The model reached almost flawless results by producing an accuracy of 1.0000 and recall of 0.9994 and precision (PPV) of 1.000. The high performance level indicates that the model demonstrates excellent generalization capabilities to detect seizure events without generating incorrect positive results.

Importantly, Random Forest demonstrated strong resistance to class imbalance problem which affects EEG-based seizure detection because seizure events appear much less frequently than non-seizure data. Random Forest achieves its resilience through its ensemble structure which trains multiple decision trees on data subsets. This technique enables individual trees to learn minority-class patterns better because seizure events are frequently sampled during training which leads the ensemble to recognize these patterns accurately. The outcome produces a highly sensitive and specific model which effectively detects seizure activity in this particular unbalanced class distribution.

The present work is not without limitations; namely, it relies on a single dataset (CHB-MIT) and is affected by the prominent class imbalance within that dataset [20–22], which significantly affected the classification results for k-NN and Decision Tree, and therefore limits the generalizability of our findings.

For further studies, we recommend using additional datasets mentioned in Wong et al. [27], such as University of Bonn dataset, NeuroVista Ictal, Helsinki University Hospital EEG, Siena Scalp EEG, Neurology and Sleep Centre Hauz Kha datasets, considering the use of a visualization of the EEG signal with Garmian Angular Fields proposed in [28], use of more classifiers, and minimizing the clas imbalance.

Author Statement

Research funding: Financed by a subvention from the Ministry of Science and Higher Education, Poland to the Academy of Silesia in Katowice. Conflict of interest: Authors state no conflict of interest. Informed consent: Informed consent has been obtained from all individuals included in this study. Ethical approval: Not applicable—The research was based on publicly available data sets.

References

[1] Thijs RD, Surges R, O'Brien TJ, Sander JW. Epilepsy in adults. The Lancet. 2019 Feb;393(10172):689–701. Available from: http://dx.doi.org/10.1016/s0140-6736(18)32596-0.

[2] Benbadis SR, Beniczky S, Bertram E, MacIver S, Moshé SL. The role of EEG in patients with suspected epilepsy. Epileptic Disorders. 2020 Apr;22(2):143–155. Available from: http://dx.doi.org/10.1684/epd.2020.1151.

[3] Mormann F, Andrzejak RG, Elger CE, Lehnertz K. Seizure prediction: the long and winding road. Brain. 2007 Feb;130(2):314–333. Available from: http://dx.doi.org/10.1093/brain/awl241.

[4] LeCun Y, Bengio Y, Hinton G. Deep learning. Nature. 2015 May;521(7553):436–444. Available from: http://dx.doi.org/10.1038/nature14539.

[5] Bou Assi E, Nguyen DK, Rihana S, Sawan M. Towards accurate prediction of epileptic seizures: A review. Biomedical Signal Processing and Control. 2017 Apr;34:144–157. Available from: http://dx.doi.org/10.1016/j.bspc.2017.02.001.

[6] Rasheed K, Qayyum A, Qadir J, Sivathamboo S, Kwan P, Kuhlmann L, et al. Machine Learning for Predicting Epileptic Seizures Using EEG Signals: A Review. IEEE Reviews in Biomedical Engineering. 2021;14:139–155. Available from: http://dx.doi.org/10.1109/rbme.2020.3008792.

[7] Shoeibi A, Moridian P, Khodatars M, Ghassemi N, Jafari M, Alizadehsani R, et al. An overview of deep learning techniques for epileptic seizures detection and prediction based on neuroimaging modalities: Methods, challenges, and future works. Computers in Biology and Medicine. 2022 Oct;149:106053. Available from: http://dx.doi.org/10.1016/j.compbiomed.2022.106053.

[8] Bandarabadi M, Teixeira CA, Rasekhi J, Dourado A. Epileptic seizure prediction using relative spectral power features. Clinical Neurophysiology. 2015 Feb;126(2):237–248. Available from: http://dx.doi.org/10.1016/j.clinph.2014.05.022.

[9] Cho D, Min B, Kim J, Lee B. EEG-Based Prediction of Epileptic Seizures Using Phase Synchronization Elicited from Noise-Assisted Multivariate Empirical Mode Decomposition. IEEE Transactions on Neural Systems and Rehabilitation Engineering. 2017 Aug;25(8):1309–1318. Available from: http://dx.doi.org/10.1109/tnsre.2016.2618937.

[10] Myers MH, Padmanabha A, Hossain G, de Jongh Curry AL, Blaha CD. Seizure Prediction and Detection via Phase and Amplitude Lock Values. Frontiers in Human Neuroscience. 2016 Mar;10. Available from: http://dx.doi.org/10.3389/fnhum.2016.00080.

[11] Cui S, Duan L, Qiao Y, Xiao Y. Learning EEG synchronization patterns for epileptic seizure prediction using bag-of-wave features. Journal of Ambient Intelligence and Humanized Computing. 2018 Sep;14(11):15557–15572. Available from: http://dx.doi.org/10.1007/s12652-018-1000-3.

[12] Shahidi Zandi A, Tafreshi R, Javidan M, Dumont GA. Predicting Epileptic Seizures in Scalp EEG Based on a Variational Bayesian Gaussian Mixture Model of Zero-Crossing Intervals. IEEE Transactions on Biomedical Engineering. 2013 May;60(5):1401-13.

[13] Elgohary S, Eldawlatly S, Khalil MI. Epileptic seizure prediction using zero-crossings analysis of EEG wavelet detail coefficients. In: 2016 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB); 2016. p. 1-6.

[14] Tsiouris KM, Pezoulas VC, Koutsouris DD, Zervakis M, Fotiadis DI. Discrimination of Preictal and Interictal Brain States from Long-Term EEG Data. In: 2017 IEEE 30th International Symposium on Computer-Based Medical Systems (CBMS); 2017. p. 318-23.

[15] Rizal A, Hadiyoso S, Aulia S, Wijayanto I, Triwiyanto, Said Z. Multiscale Hjorth Descriptor on Epileptic EEG Classification. Journal of Electrical and Computer Engineering. 2023 Dec;2023:1–11. Available from: http://dx.doi.org/10.1155/2023/4961637.

[16] Sayeed MA, Mohanty SP, Kougianos E, Zaveri H. A Fast and Accurate Approach for Real-Time Seizure Detection in the IoMT. In: 2018 IEEE International Smart Cities Conference (ISC2); 2018. p. 1-5.

[17] Hjorth B. EEG analysis based on time domain properties. Electroencephalography and Clinical Neurophysiology. 1970 Sep;29(3):306–310. Available from: http://dx.doi.org/10.1016/0013-4694(70)90143-4.

[18] Hjorth B. The physical significance of time domain descriptors in EEG analysis. Electroencephalography and Clinical Neurophysiology. 1973 Mar;34(3):321–325. Available from: http://dx.doi.org/10.1016/0013-4694(73)90260-5.

[19] Abbaszadeh B, Yagoub MCE. Optimum Window Size and Overlap for Robust Probabilistic Prediction of Seizures with iEEG. In: 2019 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB). IEEE; 2019. p. 1–5. Available from: http://dx.doi.org/10.1109/cibcb.2019.8791450.

[20] Shoeb A. CHB-MIT Scalp EEG Database. physionet.org; 2010. Available from: https://physionet.org/content/chbmit/.

[21] Shoeb AH. Application of machine learning to epileptic seizure onset detection and treatment [Ph.D. Thesis]. Cambridge, MA, USA: Harvard University–MIT Division of Health Sciences and Technology; 2009. Available from: https://dspace.mit.edu/bitstream/handle/1721.1/54669/607331139-MIT.pdf?sequence=2&isAllowed=y.

[22] Goldberger AL, Amaral LAN, Glass L, Hausdorff JM, Ivanov PC, Mark RG, et al. PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals. Circulation. 2000 Jun;101(23). Available from: http://dx.doi.org/10.1161/01.cir.101.23.e215.

[23] Liu CL, Xiao B, Hsaio WH, Tseng VS. Epileptic Seizure Prediction With Multi-View Convolutional Neural Networks. IEEE Access. 2019;7:170352-61.

[24] Daoud H, Bayoumi MA. Efficient Epileptic Seizure Prediction Based on Deep Learning. IEEE Transactions on Biomedical Circuits and Systems. 2019 Oct;13(5):804-13.

[25] Ji D, He L, Dong X, Li H, Zhong X, Liu G, et al. Epileptic Seizure Prediction Using Spatiotemporal Feature Fusion on EEG. International Journal of Neural Systems. 2024 May;34(08). Available from: http://dx.doi.org/10.1142/s0129065724500412.

[26] Guhdar M, Mstafa RJ, Mohammed AO. Hybrid Deep Learning Model for epileptic seizure classification by using 1D-CNN with multi-head attention mechanism; 2025. Available from: https://arxiv.org/abs/2501.10342.

[27] Wong S, Simmons A, Rivera-Villicana J, Barnett S, Sivathamboo S, Perucca P, et al. EEG datasets for seizure detection and prediction— A review. Epilepsia Open. 2023 Feb;8(2):252–267. Available from: http://dx.doi.org/10.1002/epi4.12704.

[28] Damasevicius R, Maskeliunas R, Wozniak M, Polap D. Visualization of physiologic signals based on Hjorth parameters and Gramian Angular Fields. In: 2018 IEEE 16th World Symposium on Applied Machine Intelligence and Informatics (SAMI). IEEE; 2018. p. 000091–000096. Available from: http://dx.doi.org/10.1109/sami.2018.8323992.