Automated Detection of Electrical Onset of Epileptic Seizures in EEG Data

*Projectseminar: Wettbewerb in der Künstlichen Intelligenz

Magnus Senfter
Tu Darmstadt
Darmstadt, Germany
magnus.senfter@stud.tu-darmstadt.de

[redacted]
Tu Darmstadt
Darmstadt, Germany
[redacted]

[redacted]
Tu Darmstadt
Darmstadt, Germany
[redacted]

***Abstract*—Epileptic seizures are sudden and unpredictable events that can significantly impact the quality of life of individuals with epilepsy. Early detection of seizure onset plays a crucial role in providing timely interventions and improving patient outcomes. In this project, we aimed to develop a machine-learning model capable of detecting the onset of epileptic seizures using electroencephalogram (EEG) data. The proposed approach involves a combination of ensemble learning and regression techniques. For the classification task, a voting ensemble model was utilized, and for the Onset Detection, a CatBoost Regressor. The main challenge encountered during this project was to achieve accurate seizure predictions. Epileptic seizures are highly diverse and complex. It is shown that the length of the segment used for Seizure Detection has a decisive influence. Furthermore, when using an ensemble method, great importance should be attached to maximizing the diversity of the base estimators.**

***Index Terms*—machine learning, epileptic seizure, onset detection, voting ensemble, CatBoost regressor, EEG data.**

I. Introduction

Epilepsy is a prevalent neurological condition that can impact individuals of diverse demographics, including all ages and genders. It is characterized by a chronic disposition towards recurrent, unprovoked seizures, leading to various neuro-biological, cognitive, psychological, and social consequences.[1] Epileptic seizures arise due to abnormal and excessive neural activity in the brain, leading to sudden disturbances in brain function. These seizures are often associated with hyper-synchronous activity of neurons in the cerebral cortex.[2] Electroencephalography (EEG) plays a vital role in epilepsy diagnosis and management by providing direct measurements of the electrical activity in the brain. As a reliable tool for monitoring brain activity, EEG can assist in the accurate detection of seizure onset and duration.[3]

In the context of the project seminar "Artificial Intelligence in Medicine Challenge", the problem of developing a machine learning model from an EEG data set that can reliably detect the electrical onset in EEG data was addressed. Such automation can not only streamline monitoring procedures but also can support in the treatment of epileptic seizures.[3]

II. Dataset

The data set was a part of the project seminar and contains 6265 EEG samples and 2473 of them are seizures. For the EEG recordings, the EEG electrodes were placed according to the 10-20 system. Most of the data uses 17 electrodes. But there are also EEG recordings where the electrodes 'Pz' and 'Cz' are used additionally. Two different reference systems were used 'LE' and 'AR'. Also, the samples vary in their sampling frequency and duration. The duration of the EEG recordings varies greatly - the shortest recordings last a few seconds and the longest several hours. To develop the model, 3598 EEG records of 452 patients were used as the training set, 899 EEG records of 261 patients were used as the validation set, and the remaining 1768 EEG records of 326 patients were used as the test set.

III. Preprocessing

Data preprocessing is a crucial step in the data analysis pipeline that involves preparing and transforming raw data into a clean, organized, and suitable format for further analysis.[4] In the following sections we will discuss the steps for data preprocessing:

A. *Segmentation*

Segmentation refers in this context to the process of dividing a data sample into distinct and homogeneous segments. The division can be based on specific characteristics or patterns to gain a deeper understanding of the data. [5]

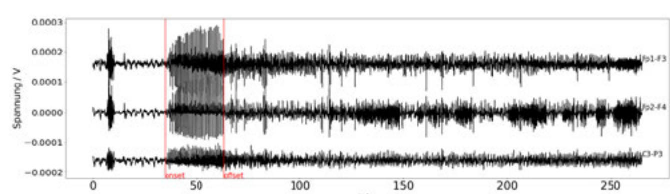

Fig. 1. Three montages of an EEG Sample with a long duration

In the case of the project, a segmentation into shorter EEG samples is necessary. As exemplified in Figure 1, the dataset contains very long EEG samples. The combination of long EEG recordings and short seizure times leads to the fact that the ictal phase has only little influence on an extracted feature. Thus, a feature has no real information value. For example, if we use the frequency of short peaks (spikes) as a feature. These short occur more frequently in the ictal phase.[6] If now the complete EEG recording is used where the largest time range is non-ictal, this distorts the result very much. Segmenting the data leads to an even more imbalanced data set, so random oversampling was performed.[7] This meant that the labels were equally represented.

B. Filtering

The scheme for filtering can be seen in Figure 2. First, the supply line noise is filtered from the data. Since the countries of origin of the recording are not known and the frequencies in many countries are 50 or 60 Hz. These and their harmonics are removed with the help of Notch Filters.[8] The next step is a bandpass filter to set the frequency range to 0.1-80 Hz. The lower limit was chosen to reduce the influence of slow drifts. The upper limit is used to attenuate sensor noise and reduce myogenic artifacts. An attempt was made to keep the loss of information in the data to a minimum, because the limits were chosen to be very wide.[9] Since the samples were recorded with different sampling rates, a re-sampling to 256 Hz was performed.

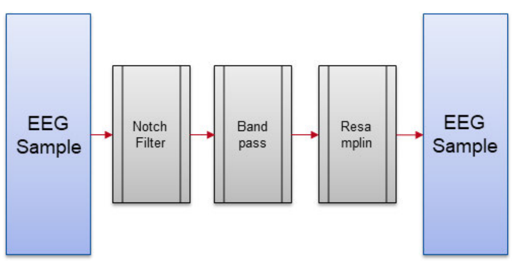

Fig. 2. Process scheme of EEG data filtering

C. Feature Extraction & Selection

One challenge in machine learning is choosing the right features to train a model. The goal is to identify the most informative and discriminative features that contribute the most to the predictive performance of a model.[10] In this approach, from the raw EEG data the three Bipolar Montages "Fp1-F3"(M1), "Fp2-F4"(M2), and "C3-P3"(M3) are generated. For each channel 25 features were extracted. The features were extracted using different types of signal analysis. [11][12][13]

A large part of the features belongs to the time domain analysis. Time domain features are extracted from the raw EEG signal within the time domain. These features provide valuable information about the temporal dynamics and patterns of brain activity.[14] [10] Examples include Hjorth mobility and complexity.[15] Frequency analysis features contain important information. The features are characteristics extracted from the EEG signal after it has been transformed from the time domain to the frequency domain using techniques like Fourier transformation or wavelet transformation. An example for this would be the spectral flatness. [16] [10]

Features without relevant information content or a strong correlation of the features should be possible to avoid. In the next step a feature selection was performed. Thereby five different tests were performed. The tests were Pearson Correlation, Chi-2, Recursive Feature Elimination and the Feature Importance of a Random Forest and LightGBM.[17] Each of them is used to determine the 20 most important features. The results of the individual tests were summed. The selected features are those that are rated as relevant by at least three of the methods (Figure 3). In total 14 features are used for model training.

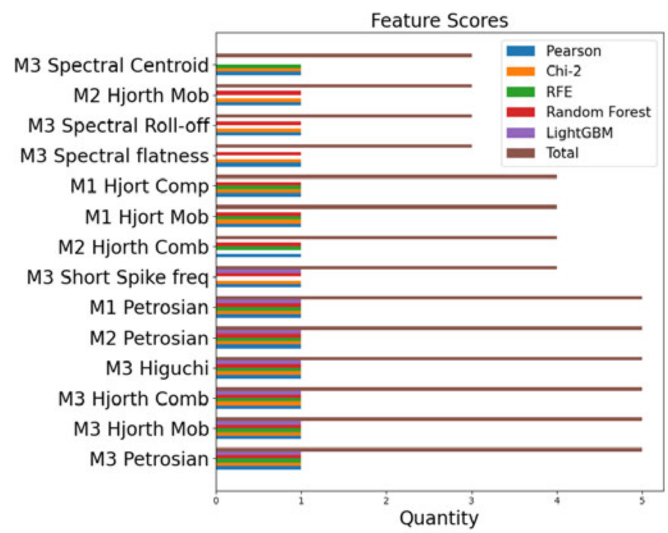

Fig. 3. Selected Features: It is shown from which test the features were selected as a relevant feature and from how many tests in total.

IV. Seizure Detection

The sequence for the detection of a seizure is shown in Figure 4. First the EEG-Sample is segmented into three non overlapping parts. Each segment is filtered and the features are extracted. A Voting Ensemble with four Base Estimators classifies if a segment contains a Seizure or Not.[18][19] The Base Estimators are a Support Vector Machine, Gradient Boosting, Extra Tree Classifier and a CatBoost Classifier. In the last step, the predictions for the different segments are combined to a final prediction. If in one of the segments a seizure is detected the whole EEG sample gets this prediction.

V. OnSet Detection

If an EEG Sample is predicted as a seizure, a separately trained model is used for used to detect the Onset time.[20] The onset detection model is based on a Catboost regressor model. The model uses six features: spectral entropy, total

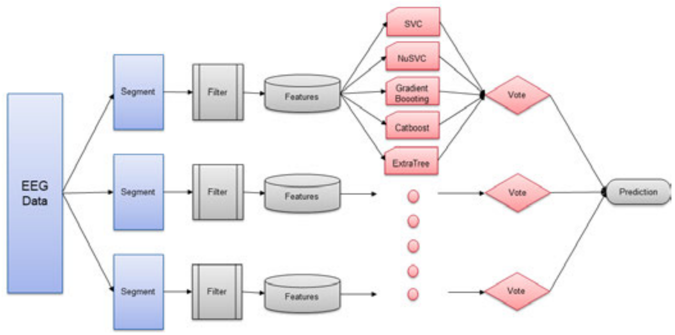

Fig. 4. Process diagram of the seizure detection of an EEG sample

power, median frequency, peak frequency, Hjorth mobility, and complexity. These features are calculated for all three montage channels. So for Onset detection in total 18 Features are used. Based on these features, a Catboost regressor model predicts the onset of the seizure. To analyze the OnSet Detection the average latency is used. The model was trained on 2783 seizure data points and then the model's latency is tested on 1009 hold-out data points. The latency on the test data was 38 seconds. The model achieved a latency of 38.38 seconds on the seminar's unseen validation dataset and 39.01 on the test dataset,

VI. Results

The analysis of the results is limited to the selection of the features, the influence of the segment length, and the diversity of the base classifiers. Other aspects, such as the effectiveness of the filters, are not considered in detail. However, it can be said that the used filtering only slightly improves the classification performance.

A. Feature Importance

The selection of features for the classifier shows that features from the field of spectral analysis like the Petrosian fractal entropy have a very high information value.[21] Furthermore, the Hjorth parameters, which represent the statistical behavior in the time domain, also play an important role in the detection of seizures. A further remark is that eight out of 14 features evaluate the behavior of the "C3-P3" montage. It can therefore be assumed that the seizures occurring in this data set have a particularly strong influence on this montage.

B. Segmentation

In this part, the previously mentioned influence of the segment length is studied in more detail. Table 1 shows the test score for a voting ensemble that extracts the features from the whole EEG sample. It can be seen that the recall for predicting non-seizures is relatively high (92%). Albeit the classifier has a way worse precision (63%). It can be deduced that the classifier tends to classify all data as non-seizures. This hypothesis is supported by the low recall (29%) in the classification of the sample with seizure.This indicates that only a small proportion of seizures were correctly identified. The designed model has therefore no suitable capability for seizure detection.

TABLE I
Test Results 1 Segment

	Precision	Recall	F1 Score
No Seizure	0,63	0,92	0,75
Seizure	0,75	0,29	0,52
Average	0,69	0,61	0,59

If the model uses shorter EEG samples in which the seizure duration has a larger share of the total duration, a significantly better classifier can be achieved. For this purpose, the training data were first adjusted. Each EEG sample was divided into three segments of equal length. The labels were then adjusted using the OnSet and Offset times to assign the appropriate labels to the segments. If the sample is between onset and offset, the label is "1" if not then the label is "0". The test score can be seen in Table 3 Although the model makes more false positive predictions, since the recall for predicting no seizures has decreased by 33 percent, it can predict seizures much better. The recall has increased by 32 percent for seizures. In summary, the model is now better at detecting a seizure, but it also detects more false positives. The strong difference in the models can mainly be attributed to the information content of the features. Because the proportion of seizure time is now significantly higher in the positive samples, the seizure time has thereby a significantly greater influence on the features. Especially samples with long duration and short seizure times are otherwise most likely hardly distinguishable from non-seizures data. By comparing the two models, the one with shorter EEG batches is preferable because a model that detects seizures but also detects more false positives has more added values than a model that detects almost all samples as negative.

TABLE II
Test Score: Voting Classifier using 3 Segments

	Precision	Recall	F1 Score
No Seizure	0.58	0.59	0.58
Seizure	0.63	0.61	0.62
Average	0.60	0.60	0.60

C. Diversity of Base Classifiers

The goal was to create a strong ensemble from a set of classifiers. As can be seen in Table 4, this goal could not be achieved. Compared to the strongest base estimator, an extra tree classifier, the two ensemble methods did not produce a significant improvement.

One possible reason for the failure of the combination is a lack of diversity in the base classifiers. In the context of ensemble learning, diversity means that the classifiers classify different samples incorrectly.[22] One way to examine diversity is to use the disagreement measure, a pairwise comparison of classifiers.[22] As shown in Table 5, the predictions of two classifiers in the context of diversity have 4 states. Either both are correct, only one is correct or neither is correct.

TABLE III
COMPARISATION OF TEST SCORES

		Extra Tree	Voting	Blending
No Seizure	Precision	0,63	0,58	0,59
	Recall	0,49	0,59	0,54
	F1 Score	0,55	0,58	0,56
Seizure	Precision	0,62	0,63	0,62
	Recall	0,74	0,61	0,66
	F1 Score	0,68	0,62	0,64

TABLE IV
RELATIONSHIP OF THE PREDICTIONS BETWEEN A PAIR OF CLASSIFIER

	$C_1(correct)$	C_0(wrong)
$D_1(correct)$	N_{11}	N_{10}
D_0(wrong)	N_{01}	N_{00}
	Total $N = N_{11} + N_{10} + N_{01} + N_{00}$	

In this case the samples that are correctly classified by both classifiers are not relevant. Therefore, the disagreement measure was adjusted that they no longer have any influence. From the remaining states, the disagreement measure can be calculated as follows:

$$Dis = \frac{N_{10} + N_{01}}{N_{10} + N_{01} + N_{11}} \tag{1}$$

To get the following diversity values, we calculate the class indicators using the formula **??**. A score near "1" means high pairwise diversity and near '0' means a low diversity. It can be seen that the base classifiers have a low pairwise diversity. Partly the classifiers make different mistakes but it can be assumed that the diversity of the entire ensemble is even lower. For this reason, in this case an ensemble can not really improve the classification result.

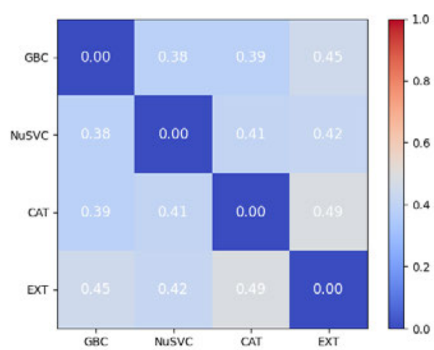

Fig. 5. Heatmap of disagreement measures

VII. DISCUSSION

Training a machine learning application that can reliably and robustly detect the onset of an epileptic seizure proves to be extremely complicated. And a critical look shows that the methods attempted in the paper need to be significantly improved to have a positive impact on the lifes of people with epilepsy. It has been shown that the segmentation of the EEG data in batches has an enormous influence on the classification result. However, a naïve split of the data as chosen here proves to be insufficient. Especially the handling of long EEG recordings was insufficient. It is probably more useful to create segments of a fixed length. It must be examined which window length proves to be optimal. Also the approach with overlapping moving windows could prove to be more useful because the transition from non-ictal to ictal phase can be captured better. In the approach carried out here, however, the lack of diversity in the basic classifiers emerges as the biggest problem. In order to a combination of the predictions, a more diverse set is needed to be useful. For this purpose, the classifiers would have to be trained with different feature subspace or training data composition. [23] Another critical point is the chosen filtering. Better results could be achieved compared to without filtering. However, the effectiveness of the individual filters was not investigated further. Especially with the bandpass filter, it should be investigated for which band range the result is optimal. Also, a stronger focus should be placed on the removal of artifacts or the removal of bad channels. Fundamentally, there is also the question of whether it is a reasonable approach to create models that should recognize all types of seizures. Since there are many different types of seizures, which are also reflected in different EGG characteristics. Rather, it might be more sustainable to train classifiers that are designed to recognize specific types of seizures. [24] Alternatively, clustering of the data could also be effective, if appropriate with the help of machine learning. Thus, the EEG recordings can be classified according to characteristics, which may simplify an analysis. EEG characteristics change, for example, with increasing age.[25]

VIII. CONCLUSION

In this project, we aimed to develop a machine learning model capable of detecting the onset of epileptic seizures using EEG data. Our approach involved a combination of ensemble learning and regression techniques to improve the accuracy of seizure predictions. Despite the challenges encountered, we made important strides towards achieving reliable seizure detection. Segmentation of the EEG data proved crucial in improving the classifier's performance. Shorter EEG segments with a higher proportion of seizure time demonstrated better results compared to using the entire EEG sample for prediction. Additionally, feature selection was performed to identify the most informative and discriminative features for training the model. Here, features from the field of spectral analysis proved to be very informative While ensemble learning methods are known to enhance classifier performance, we encountered a lack of diversity in the base classifiers, leading to limited improvement with the ensemble approach. To overcome this, diverse classifiers should be considered for future enhancements. The project highlights the importance of data preprocessing, segmentation, and feature selection in developing a reliable seizure classifier. While significant progress has been made, more work is needed to achieve a truly robust and accurate classifier.

References

[1] Ettore Beghi. “The Epidemiology of Epilepsy”. In: *Neuroepidemiology* 54.2 (2020), pp. 185–191. DOI: 10.1159/000503831.

[2] Lasitha S. Vidyaratne and Khan M. Iftekharuddin. “Real-Time Epileptic Seizure Detection Using EEG”. In: *IEEE transactions on neural systems and rehabilitation engineering : a publication of the IEEE Engineering in Medicine and Biology Society* 25.11 (2017), pp. 2146–2156. DOI: 10.1109/TNSRE.2017.2697920.

[3] M. E. Saab and J. Gotman. “A system to detect the onset of epileptic seizures in scalp EEG”. In: *Clinical Neurophysiology* 116.2 (2005), pp. 427–442. ISSN: 1388-2457. DOI: 10.1016/j.clinph.2004.08.004. URL: https://www.sciencedirect.com/science/article/pii/S1388245704003098.

[4] Carlos Vladimiro Gonzalez Zelaya. “Towards Explaining the Effects of Data Preprocessing on Machine Learning”. In: *2019 IEEE 35th International Conference on Data Engineering (ICDE)*. IEEE, 2019, pp. 2086–2090. ISBN: 978-1-5386-7474-1. DOI: 10.1109/ICDE.2019.00245.

[5] Zeba Karin Ahmad, Vikram Singh, and Yusuf Uzzaman Khan. “Sequential Segmentation of EEG Signals for Epileptic Seizure Detection using Machine Learning”. In: *2019 2nd International Conference on Signal Processing and Communication (ICSPC)*. 2019, pp. 258–262. DOI: 10.1109/ICSPC46172.2019.8976487.

[6] Itaf Ben Slimen, Larbi Boubchir, and Hassene Seddik. “Epileptic seizure prediction based on EEG spikes detection of ictal-preictal states”. In: *Journal of Biomedical Research* 34.3 (2020), pp. 162–169. DOI: 10.7555/JBR.34.20190097. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7324272/.

[7] Giulia Varotto et al. “Comparison of Resampling Techniques for Imbalanced Datasets in Machine Learning: Application to Epileptogenic Zone Localization From Interictal Intracranial EEG Recordings in Patients With Focal Epilepsy”. In: *Frontiers in Neuroinformatics* 15 (2021), p. 715421. DOI: 10.3389/fninf.2021.715421. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8641296/.

[8] Manish N. Tibdewal et al. “Power line and ocular artifact denoising from EEG using notch filter and wavelet transform”. In: *2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom)*. 2016, pp. 1654–1659.

[9] Alain de Cheveigné and Dorothée Arzounian. “Robust detrending, rereferencing, outlier detection, and inpainting for multichannel data”. In: *Neuroimage* 172 (2018), pp. 903–912. DOI: 10.1016/j.neuroimage.2018.01.035. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5915520/.

[10] L. Boubchir, B. Daachi, and V. Pangracious. “A review of feature extraction for EEG epileptic seizure detection and classification”. In: *2017 40th International Conference on Telecommunications and Signal Processing (TSP)*. 2017, pp. 456–460. DOI: 10.1109/TSP.2017.8076027.

[11] Moctezuma LA, Molinas M. *”Classification of low-density EEG for epileptic seizures by energy and fractal features based on EMD”*. J Biomed Res. 2019. 2019.

[12] Igor Stancin, Mario Cifrek and Alan Jovic. *”A Review of EEG Signal Features and Their Application in Driver Drowsiness Detection Systems”*. Faculty of Electrical Engineering and Computing, University of Zagreb. 2014.

[13] Lorena Orosco. “Review: A Survey of performance and techniques for automatic epilepsy detection”. In: *Journal of Medical and Biological Engineering* 33.6 (2013), p. 526. ISSN: 1609-0985. DOI: 10.5405/jmbe.1463. URL: https://ri.conicet.gov.ar/handle/11336/26634.

[14] *A Novel EEG Feature Extraction Method using Hjorth Parameter*. 2014.

[15] Seung-Hyeon Oh, Yu-Ri Lee and Hyoung-Nam Kim. *”A Novel EEG Feature Extraction Method using Hjorth Parameter”*. Pusan National University,Republic of Korea. 2014.

[16] Fingelkurts AA, Fingelkurts AA. *”Short-term EEG spectral pattern as a single event in EEG phenomenology”*. Open Neuroimag J. 2010. 2010.

[17] Rahul Agarwal. “The 5 Feature Selection Algorithms every Data Scientist should know”. In: *Towards Data Science* (27.07.2019). URL: https://towardsdatascience.com/the-5-feature-selection-algorithms-every-data-scientist-need-to-know-3a6b566efd2.

[18] Mussabayeva A, Jamwal PK, Akhtar MT. *”Ensemble Voting-Based Multichannel EEG Classification in a Subject-Independent P300 Speller”*. Applied Sciences. 2021. 2021.

[19] Gu, Xiaotong Cao, Zehong. *”An EEG Majority Vote Based BCI Classification System for Discrimination of Hand Motor Attempts in Stroke Patients”*. 2020.

[20] Ali Shoeb, Herman Edwards, Jack Connolly, Blaise Bourgeois, Ted Treves, and John Guttag. *”Patient-Specific Seizure Onset Detection”*. Cambridge, MA. 2004.

[21] Navid Ghassemi et al. “Epileptic seizures detection in EEG signals using TQWT and ensemble learning”. In: *2019 9th International Conference on Computer and Knowledge Engineering (ICCKE)*. 2019, pp. 403–408. DOI: 10.1109/ICCKE48569.2019.8964826.

[22] *IEEE Xplore Full-Text PDF*. 4.05.2023. URL: https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7911296.

[23] Ludmila I. Kuncheva and Christopher J. Whitaker. “Measures of Diversity in Classifier Ensembles and Their Relationship with the Ensemble Accuracy”. In: *Machine Learning* 51.2 (2003), pp. 181–207. ISSN: 1573-0565. DOI: 10.1023/A:1022859003006. URL:

https : / / link . springer . com / article / 10 . 1023 / A : 1022859003006#citeas.

[24] Shah T. Sarmast, Abba Musa Abdullahi, and Nusrat Jahan. "Current Classification of Seizures and Epilepsies: Scope, Limitations and Recommendations for Future Action". In: *Cureus* 12.9 (2020), e10549. DOI: 10.7759/cureus.10549. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7575300/.

[25] Melinda L. Morgan et al. "Influence of age, gender, health status, and depression on quantitative EEG". In: *Neuropsychobiology* 52.2 (2005), pp. 71–76. ISSN: 0302-282X. DOI: 10.1159/000086608. URL: https://karger.com/nps/article/52/2/71/233052/Influence-of-Age-Gender-Health-Status-and.