

Epileptic Seizure Detection on TUSZ: Statistics and Channel-wise Approach

V. Stragier¹, P. Vanabelle² and R. El Tahry^{3,4}

1. Information, Signal and Artificial Intelligence (ISIA) Lab, University of Mons (UMONS), Boulevard Dolez 31, 7000 Mons, Belgique
2. Data Science Department, Centre of Excellence in Information and Communication Technologies (CETIC), Avenue Jean Mermoz 28, 6041 Charleroi, Belgique
3. Refractory Epilepsy Centre, University Hospital of Saint-Luc, 1200 Bruxelles, Belgique
4. Institute of Neuroscience, Catholic University of Louvain, 1200 Bruxelles, Belgique
vincent.stragier@umons.ac.be, paul.vanabelle@cetic.be, riem.eltahry@uclouvain.be

The detection of epileptic seizures on electroencephalogram (EEG) signals presents a considerable challenge, whether for the diagnosis of epileptic seizures or for outpatient follow-up. Developing an automated system to assist neurologists in this task is essential [1]. Indeed, the task of analyzing an EEG is time-consuming, as well as costly, since it requires the expertise of neurologists specialized in interpreting epilepsy EEG signals.

Our epilepsy seizure detection pipeline [2] is built on feature extraction with an extreme gradient boosting (XGBoost) classifier. The particular interest of decision tree-based algorithms – random forest, XGBoost, etc. – is that the selection of features is not necessary and that models are interpretable by themselves. The dataset we used is the Temple University Hospital EEG Seizure Corpus (TUSZ) [3]. Recent changes to the pipeline were done to pre-processing, feature extraction and architecture. Version 1.5.2. of the TUSZ was considered as input. For pre-processing, the extraction of overlapping windows of EEG signals was added, i.e., a 4-second-wide window extracted from each channel every second. Concerning the features, some PyEEG [4] functions were corrected, and others, like the line length [5], were added to complete our set of features. Improvements to the pipeline can accelerate the exploration of new techniques and decrease the likeliness of implementation errors.

Moreover, to better understand past results, a detailed investigation of the onset and propagation of TUSZ seizures in relation to channels was carried out. This descriptive analysis is an important step in the methodology of data science. Besides some issues revealed in the corpus, the results demonstrate that our system, built on a global approach, is inadequate for detecting whether a seizure is occurring somewhere on the channels. Rather, the results suggest the utility of a channel-wise approach, due to the broad variety of seizure dynamotypes observed. Dynamotypes define the dynamic composition of the seizure, i.e., how they begin, evolve and end [6]. This method could answer whether seizure modeling is possible at the channel level. If successful, the advantages would be a better localization and segmentation of seizures. Hence, this is very interesting for research purposes and could help neurologists locate the origin of seizures within the brain, diagnose or decide on possible surgery more quickly. Furthermore, it is assumed that, if the seizure state of nearby channels is considered in a post-processing unit, seizure detection could be enforced with a reduced number of false alarms.

The rebuilt pipeline was tested for validation, giving us updated results for the v1.5.2. dataset with the global approach. Still, the highest F1 scores were low – at most 48 % – and the number of false detections was relatively elevated. After inspecting the results, artifacts were found to be the root of some false detections. This then advocates for proceeding with artifact removal during the pre-processing of the signals, which has not yet been implemented. On a few channels in the temporal area – F7-T3, T3-C3, and T3-T5 – (a topic of interest for temporal lobe epilepsy research for the neurologists consulted), the channel-wise approach was applied, and the combination of the results was tested in pairs. While the combination rules worked on the training set, performance on unseen data (dev set) did not improve. This observation is important, as it could imply that the extensive use of features as input of a classifier is inefficient to solve

the detection task at the channel level. There are numerous explanations for this, from the artifact issues to the different distribution of seizure types between the training and dev set. However, this approach must be further investigated with a broader implementation and a new set of data.

ACKNOWLEDGMENTS

This work was carried out as part of the ARIAC project and supported by the Walloon Region.

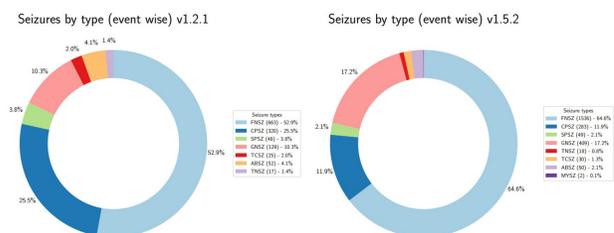
REFERENCES

- [1] S. Roy, I. Kiral, M. Mirmomeni, T. Mummert, A. Braz, J. Tsay, J. Tang, U. Asif, T. Schaffter, M. E. Ahsen, T. Iwamori, H. Yanagisawa, H. Poonawala, P. Madan, Y. Qin, J. Picone, I. Obeid, B. D. A. Marques, S. Maetschke, R. Khalaf, M. Rosen-Zvi, G. Stolovitzky, and S. Harrer, "Evaluation of artificial intelligence systems for assisting neurologists with fast and accurate annotations of scalp electroencephalography data," *EBioMedicine*, vol. 66, p. 103275, Apr. 2021. [Online]. Available: <https://linkinghub.elsevier.com/retrieve/pii/S2352396421000682>.
- [2] P. Vanabelle, P. De Handschutter, R. El Tahry, M. Benjelloun, and M. Boukhebouze, "Epileptic seizure detection using eeg signals and extreme gradient boosting," *Journal of biomedical research*, vol. 34, no. 3, p. 228, 2020.
- [3] V. Shah, E. von Weltin, S. Lopez, J. McHugh, L. Veloso, M. Golmohammadi, I. Obeid, and J. Picone, "The Temple University Hospital Seizure Detection Corpus," *Frontiers in Neuroinformatics*, vol. 12, Jan. 2018.
- [4] F. Bao, X. Liu, and C. Zhang, "PyEEG: An open source python module for EEG/MEG feature extraction," *Computational intelligence and neuroscience*, vol. 2011, p. 406391, Mar. 2011.
- [5] J. G. Bogaarts, D. M. W. Hilkman, E. D. Gommer, V. H. J. M. van Kranen-Mastenbroek, and J. P. H. Reulen, "Improved epileptic seizure detection combining dynamic feature normalization with EEG novelty detection," *Med Biol Eng Comput*, vol. 54, no. 12, pp. 1883–1892, Dec. 2016.
- [6] M. L. Saggio, D. Crisp, J. M. Scott, P. Karoly, L. Kuhlmann, M. Nakatani, T. Murai, M. Dumpelmann, A. Schulze-Bonhage, A. Ikeda, M. Cook, S. V. Gliske, J. Lin, C. Bernard, V. Jirsa, and W. C. Stacey, "A taxonomy of seizure dynamotypes," *eLife*, vol. 9, p. e55632, Jul. 2020. [Online]. Available: <https://doi.org/10.7554/eLife.55632>.

Abstract

- Electroencephalograms (EEG) are widely used to diagnose epilepsy. However, specialized neurologists must be available to analyze these signals, which is time-consuming and expensive.
- With the current progress in artificial intelligence (AI), more automatic or semi-automatic seizure detection tools are developed for research purposes.
- For all AI technics, it is mandatory to use a well-built dataset for the use case. We can see that this is indeed the situation at Temple University Hospital Seizure Corpus (TUSZ) since it has the largest publicly available dataset that we know. The quantity of data, as well as the quality, plays a significant role in training an AI model. Thus, the dataset must be analyzed to better understand the epileptic seizure detection problem.
- Later, efforts were focalized on using an extreme gradient boosting classifier (XGBoost classifier). XGBoost is effective with projects related to time series.
- New training technics were tested to see if an approach per channel was more effective for this problem.
- Results showed that the newly implemented technics were less effective and that artifacts have a strong effect on the models' performances. Still, a per-channel approach brings an insight to the neurologist.

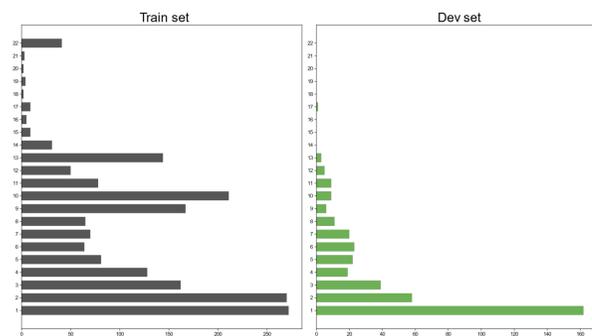
Overview of the TUSZ dataset



- The exploration of the dataset showed that the distribution of seizure events by type is uneven, with more than 75% of focalized seizures.
- Regarding the evolution of the dataset from Version 1.2.1 to Version 1.5.2 seems to be mostly stable.
- Some seizure events were reclassified as background or as other types of seizure.

Seizure Onset Analysis

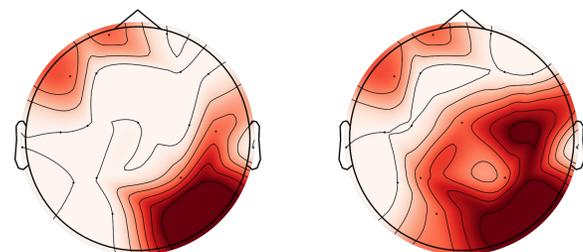
- A special interest in the focalized seizures existed during the project. Therefore, an analysis of the focalized seizure onset was realized by isolating every focal seizure onset and by counting the number of channels on which these onsets began.
- The previous analysis showed a few focalized seizures starting on more than half of the channels, and even some starting on all channels.



- The analysis helped to detect annotation errors in the dataset (improving its quality).
- The above graphs (train set on the left, dev set on the right) show the dev set as being cleaner.

Spatial Analysis

- Going through the files, seizures with "sparse" onsets were detected (see figure hereunder)



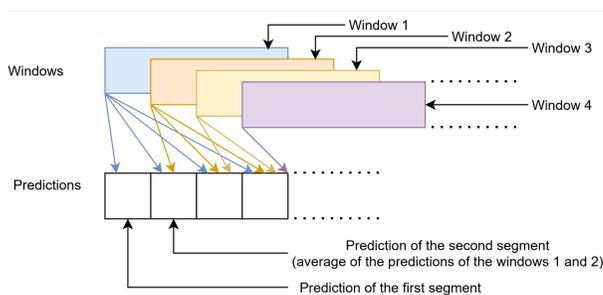
- The above figures were generated using the record 00009866_s003_t_005 at 40.5117 (left) and 457.082 seconds (right) from the TUSZ Version 1.5.2.
- The latter showed that an annotator was using the wrong montage file to annotate files.

Duration of Focal Seizures

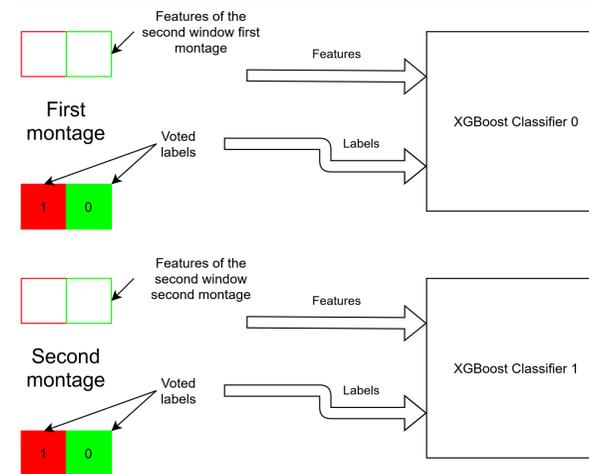
- Analyzing the duration of the focal seizures showed that the range of duration was quite wide from less than 2 seconds to more than half an hour.
- Moreover, all types of seizures have that kind of wide distribution, making any temporal analysis ineffective when trying to classify various types of seizures.

Pipeline Evolution

- Overlapping windows are implemented in the latest version of our pipeline. To each overlapping window is attributed a new label.



- New features were added, e.g., line length.
- Training and prediction with XGBoost are now realized on each channel.



Results

- Using virtually the same pipeline as before (training and prediction using all channels at once), we were able to reproduce the results of our older pipeline. Then, the results were used as a baseline for the subsequent experiments.

	(old) model 1	Model 11
Sensitivity	36.8 %	52.67 %
Specificity	92.46 %	89.06 %
F1 score	33.6 %	40.34 %

- Our experiment was focalized on montages F7-T3 (model 7), T3-C3 (model 8), and T3-T5 (model 9).

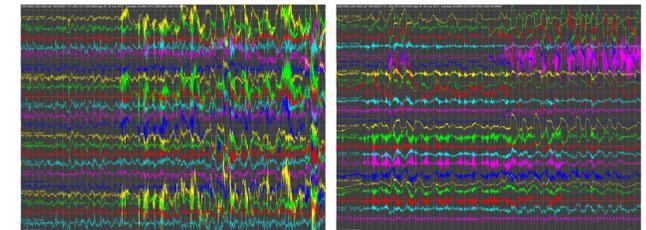
Models	Train set		
	Sensitivity (%)	Specificity (%)	F1 Score (%)
7	56.342	91.137	37.581
8	61.617	87.559	33.944
9	61.376	89.609	37.241
7-8	49.75	95.267	43.954
8-9	52.854	94.624	44.064
7-9	49.661	94.896	42.755
7-8-9	46.701	96.68	46.595

Models	Dev set		
	Sensitivity (%)	Specificity (%)	F1 Score (%)
7	41.515	86.972	31.401
8	50.649	77.936	28.234
9	50.77	84.12	33.782
7-8	30.494	92.441	30.223
8-9	36.36	91.139	33.063
7-9	33.965	92.204	32.736
7-8-9	26.859	94.571	30.167

- When combining predictions of the models, improvements are only on the training set. Therefore, this pipeline implementation seems ineffective on new data.

Artifacts

- Analyzing the results highlighted the necessity of artifact removal. Indeed, most of the models' false detections lead to artifacts.



- The above EEG shows a seizure on the left and artifacts on the right (probably induced by muscles movements), hence, making it difficult to differentiate artifacts and seizures.

Conclusions

- Analysis of the dataset sheds light on the nature of the seizures and detected annotation errors in the dataset.
- The duration of a seizure is not sufficient information to classify any kind of seizure.
- Using a combination of channel-wise models seems to be ineffective on new data but could improve the diagnosis of patients.
- Artifacts are a major issue for seizure detection since EEG signals are close to stochastic and have a small (around 1 μV) amplitude with respect to other biological signals.
- Our code is hosted on GitHub: https://github.com/cetic/TUH_EEG_Seizure_Detection

Summary

- The first goal of this work was to analyze the TUSZ to extract pieces of information relative to the distribution of seizure events, to the dynamic of the seizure onsets, and to the duration of the seizures.
- The second goal was to rebuild and improve an existing seizure detection pipeline developed by the CETIC.
- The third goal was to test a new technique that could improve the diagnosis of epileptic seizures.
- The analysis of the dataset helped to understand the topic of seizure detection and to detect errors.
- The new implementation of our pipeline helped to validate the previous experiment and to test new technics.
- Obtained results helped to find the main problem of our pipeline, which is the lack of artifact removal technics.

Acknowledgments

- This work was carried out as part of the ARIAC project and supported by the Walloon Region.