Cascade CNN-based model for epileptic seizure diagnostics

Sergei Nazarikov
Baltic Center for Neurotechnology and Artificial Intelligence Immanuel Kant Baltic Federal University
Kaliningrad, Russia
snazarikov@gmail.com

Olga Vertinskaya
Institute of Education and Human Sciences Immanuel Kant Baltic Federal University
Kaliningrad, Russia

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

***Abstract*—Automated seizure detection is a complex task in epilepsy diagnosis, mainly due to the rarity of seizures. Automatic solutions often lead to a high number of false positives. This study proposes a two-stage approach that combines iterative refinement algorithms and convolutional neural networks to minimize false positives in seizure detection task. The method was tested on a real-world EEG dataset using epilepsy-specific metrics and demonstrated a significant reduction in false positives, demonstrating the potential of the method for use in clinical decision-support systems.**

***Index Terms*—Epileptic seizure detection, cascade approach, convolutional neural network, continuous wavelet transform, EEG**

I. Introduction

Epilepsy is a neurological disorder that manifests itself in the form of repeated seizures. These manifestations significantly affect the quality of life of any subject and require accurate and early diagnosis for effective treatment. Also, through the early detection and appropriate management of this disease, the majority of patients can attain a state of remission [1].

Electroencephalography (EEG) is a primary tool for epilepsy diagnosis which captures brain electrical activity. Although with the help of an experienced doctor, seizures can be precisely identified it is still a very time-consuming process due to the fact that EEG recording can last for several days during monitoring.

In recent years, artificial neural networks (ANNs) and, especially, convolutional neural networks (CNNs) have attracted the attention of researchers around the world and have been applied to a wide range of tasks, demonstrating impressive results [2]–[4]. Obviously, the seizure detection field is not an exception [5], [6]. Nevertheless, these approaches still struggle with high false positive rates due to data imbalance. In the field of classical ML, cascade algorithms occupy a dominant position [7], [8], the idea of which gradually penetrates into the field of deep learning (DL) [9], [10].

This research was funded by Academic Leadership Program PRIORITY'2030 of Immanuel Kant Baltic Federal University of the Ministry of Science and Education of Russian Federation.

This study, inspired by the cascade algorithms widely used in classical ML, proposes a CNN-based cascade algorithm to improve seizure detection targeting the FP problem specifically. The proposed algorithm consists of two stages: first training a CNN model on the original dataset and then using its errors to train a second, more precise CNN model. To demonstrate the potential of this approach we compared its performance with the standard CNN-based method.

II. Materials and methods

A. Dataset

The real EEG dataset used in this study contains recordings from patients with focal epilepsy, provided by the National Medical and Surgical Center named after N. I. Pirogov of the Ministry of Health of the Russian Federation (Moscow, Russia). The recordings were carried out in the period from 2017 to 2019. The dataset in total includes recordings from 83 patients, however, after excluding recordings with an excessive amount of artifacts, the final dataset consists of 67 recordings. Each recording, whose duration ranged from 8 to 84 hours, was manually reviewed and labeled by an epileptologist. The ratio of the normal activity to the epileptic activity is more than 200:1 which indicates a strong data imbalance, which, however, aligns with the rare nature of seizures. The data was recorded at a 128 Hz sampling rate with $N = 25$ channels according to the "10—20" montage [11].

B. Seizure detetion with CNN

EEG signals $\{x_n\}_{n=1}^{N}$ were analyzed using continuous wavelet transform (CWT) to construct a feature space for CNN models [12]–[15]. The wavelet power (WP) in the 1–40 Hz range was used as the main characteristic [16], [17]. Since CNNs can't handle arbitrary long recordings directly, all EEG recordings were segmented into 10-second long intervals for further classification with CNN. Moreover, classification networks usually work with normalized images, therefore, mimicking such behavior the WP data was normalized in the following way:

$$w_n^{\ln} = \frac{\ln(w_n) - \mu(\ln(w_n))}{\sigma(\ln(w_n))}, \tag{1}$$

where $n = 1, 2, \ldots N$ – number of the channel, w_n – WP of a signal x_n, $\mu(\cdot)$ – mean value, $\sigma(\cdot)$ – standard deviation.

For the classification of 10-second normalized segments, we used a CNN-based classifier of ResNet-18 architecture [18], whose main feature lies in residual layers which make smoother loss surface. We modified ResNet-18 architecture, originally designed not for binary classification. First of all, we replaced the first convolutional layer designed for 3 channels RGB images with a convolutional layer that accepts 25-channel input (representing the WP spectrum). Second, we modified the final fully connected layer to produce a single output. The total architecture consists of 18 layers with approximately 11.3 million parameters.

To train CNN models we can't use all 10-second segments because of the strong data imbalance mentioned earlier. As a workaround, we sampled 100 segments per patient each epoch with approximately 50% of them containing seizure activity. This approach can be seen as a form of oversampling and undersampling [19]. Except sampling process, there were no other specific tricks and the model was trained using the following hyperparameters:

- augmentations: random flip, SpecAugment [20],
- loss function: binary cross-entropy (BCE),
- number of epochs: 10,
- learning rate: 0.001,
- batch size: 4,
- optimizer: Adam.

III. False positives suppression

A. Error-aware CNN

The proposed model to reduce the number of false positive predictions is called error-aware CNN. For this model we modified the training procedure with an error-aware approach, inspired by cascade algorithms, which involves two steps. In the first step, we train the baseline CNN model according to the procedure described in the previous section. This baseline model allows us to identify examples that are difficult for CNN to classify correctly, namely those on which it made an error. In the second step, we train another model of the same architecture from scratch, focusing on these difficult examples, thereby improving the model's overall performance. This second model is called an error-aware CNN.

Technically speaking, the major difference between the error-aware and baseline model lies in the sampling of training examples. For the error-aware CNN, half of the samples were selected similarly to the baseline but the other half were chosen from the examples where the baseline model made a mistake.

B. Postprocessing

It's worth mentioning that FP can be reduced not only with the modification of CNN architecture and training procedure. We also applied postprocessing for the same goal. During an exploratory data analysis (EDA) we noticed a significant number of short predictions from the baseline CNN model. After the analysis of the lengths of true seizures, we found out that the average seizure duration is about 100 seconds, while many CNN predictions were much shorter, which is a signal of false positives. Another insight from an EDA - predictions of the class often were grouped together or "almost together" (separated by one or more segments of the other class).

The results of an EDA led us to two types of post-processing techniques:

1) We use a median filter with a kernel size of $K = 7$ to smooth CNN outputs and reduce the number of lonely short predictions
2) To handle grouped predictions we merged neighboring segments of the same predicted class into a single longer segment. In addition to it, positive predictions separated by a single negative prediction were merged into a single positive segment

C. Evaluation procedure

Model evaluation typically employs standard metrics such as $recall$, $precision$, and F_1. However, these traditional metrics may be unsuitable for CDSS due to the rarity of seizures in EEG data. To take into account the specific requirements of the task, we have adjusted the way TP, FP, and FN are calculated. More specifically, if one or more predictions occur within a T-second neighborhood of a true seizure, they are all considered as a single TP prediction. Predictions outside a T-second neighborhood of any true seizure are treated as FP predictions. Finally, if there are no predictions within a T-second neighborhood of a true seizure then we have a FN prediction. In this study, the parameter $T = 60\,\mathrm{s}$ acknowledges possible imperfections of expert labels.

Important to note that, before metrics computations, raw predictions are transformed into binary predictions using thresholding. We choose a threshold automatically by maximization of $precision$ while maintaining decent $recall > 0.8$ on a validation set. Such an approach aligns with the goals of the CDSS to detect the most possible seizures, despite higher false positives.

IV. Results

In this section, we evaluated the proposed CNN model for epileptic seizure detection comparing the Baseline CNN and the Error-aware CNN performance. The results, at different processing steps, are provided in Table I. Each model's performance is evaluated at three different stages: *raw predictions* with 10-second segments, after applying *filtration* to 10-second segments, and after *merging* with arbitrary long segments.

It can be seen, that for the Baseline model, the $precision$ improves from 0.0638 at the raw prediction stage to 0.1462 after filtration, and slightly decreases to 0.1273 after merging. The $recall$ starts at 0.7169, increases to 0.7339 post-filtration, and dramatically rises to 0.9608 after merging. Consequently, the F_1 score follows a similar trend, starting at 0.1171, improving to 0.2438, and then slightly dropping to 0.2248. The Error-aware model shows significant improvements across all metrics at each stage in comparison with a baseline. The $precision$ improves from 0.2763 to 0.4567 after filtration and finally to 0.5366 after merging. The $recall$ shows an

TABLE I
RESULTS OF THE PERFORMANCE EVALUATION OF THE PROPOSED CNN MODEL FOR EPILEPTIC SEIZURE DETECTION: COMPARISION OF THE BASELINE CNN AND THE ERROR-AWARE CNN

Model	Step	$precision$	$recall$	F_1	FN	FP	TP
Baseline	raw preds	0.0638	0.7169	0.1171	167	6212	423
	filtration	0.1462	0.7339	0.2438	157	2529	433
	merging	0.1273	0.9608	0.2248	2	336	49
Error-aware	raw preds	0.2763	0.5559	0.3692	262	859	328
	filtration	0.4567	0.5271	0.4894	279	370	311
	merging	0.5366	0.8627	0.6617	7	38	44

increase from 0.5559 to 0.5271 and peaks at 0.8627. These results indicate substantial overall performance enhancement with each successive step.

V. CONCLUSION

In this paper, we proposed a cascade model for epileptic seizure detection, based on ResNet-18 architecture that utilizes the information about the errors made by the baseline model. The proposed approach effectively reduces false positives in seizure detection. The error-aware CNN consistently outperforms the baseline CNN, improving $precision$ and F_1-score. Techniques like median filtering and segment merging further enhance performance by reducing the number of false positive predictions and simplifying evaluation. Despite the promising results, future research should focus on optimizing CNN architecture, reducing computational demands, and improving model interpretability for real-world applications.

REFERENCES

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

[2] V. V. Grubov, S. I. Nazarikov, S. A. Kurkin, N. P. Utyashev, D. A. Andrikov, O. E. Karpov, and A. E. Hramov, "Two-stage approach with combination of outlier detection method and deep learning enhances automatic epileptic seizure detection," *IEEE Access*, vol. 12, pp. 122 168–122 182, 2024.

[3] 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.

[4] A. V. Gusev, A. V. Vladzymyrskyy, D. E. Sharova, K. M. Arzamasov, and A. E. Khramov, "Evolution of research and development in the field of artificial intelligence technologies for healthcare in the russian federation: results of 2021," *Digital Diagnostics*, vol. 3, no. 3, pp. 178–194, 2022.

[5] Y. Roy, H. Banville, I. Albuquerque, A. Gramfort, T. H. Falk, and J. Faubert, "Deep learning-based electroencephalography analysis: a systematic review," *Journal of neural engineering*, vol. 16, no. 5, p. 051001, 2019.

[6] W. Zhao, W. Zhao, W. Wang, X. Jiang, X. Zhang, Y. Peng, B. Zhang, G. Zhang *et al.*, "A novel deep neural network for robust detection of seizures using eeg signals," *Computational and mathematical methods in medicine*, vol. 2020, 2020.

[7] J. H. Friedman, "Greedy function approximation: a gradient boosting machine," *Annals of statistics*, pp. 1189–1232, 2001.

[8] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, and T.-Y. Liu, "Lightgbm: A highly efficient gradient boosting decision tree," *Advances in neural information processing systems*, vol. 30, 2017.

[9] L. Jing, Y. Chen, and Y. Tian, "Coarse-to-fine semantic segmentation from image-level labels," *IEEE transactions on image processing*, vol. 29, pp. 225–236, 2019.

[10] N. Xu, B. Price, S. Cohen, and T. Huang, "Deep image matting," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 2970–2979.

[11] R. W. Homan, "The 10-20 electrode system and cerebral location," *American Journal of EEG Technology*, vol. 28, no. 4, pp. 269–279, 1988.

[12] A. N. Pavlov, A. E. Hramov, A. A. Koronovskii, E. Y. Sitnikova, V. A. Makarov, and A. A. Ovchinnikov, "Wavelet analysis in neurodynamics," *Physics-Uspekhi*, vol. 55, no. 9, p. 845, 2012.

[13] A. Kuc, S. Korchagin, V. A. Maksimenko, N. Shusharina, and A. E. Hramov, "Combining statistical analysis and machine learning for eeg scalp topograms classification," *Frontiers in Systems Neuroscience*, vol. 15, p. 716897, 2021.

[14] A. Batmanova, A. Kuc, V. Maksimenko, A. Savosenkov, N. Grigorev, S. Gordleeva, V. Kazantsev, S. Korchagin, and A. E. Hramov, "Predicting perceptual decision-making errors using eeg and machine learning," *Mathematics*, vol. 10, no. 17, p. 3153, 2022.

[15] 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.

[16] 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.

[17] 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.

[18] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[19] J. Hernandez, J. A. Carrasco-Ochoa, and J. F. Martínez-Trinidad, "An empirical study of oversampling and undersampling for instance selection methods on imbalance datasets," in *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications: 18th Iberoamerican Congress, CIARP 2013, Havana, Cuba, November 20-23, 2013, Proceedings, Part I 18*. Springer, 2013, pp. 262–269.

[20] D. S. Park, W. Chan, Y. Zhang, C.-C. Chiu, B. Zoph, E. D. Cubuk, and Q. V. Le, "Specaugment: A simple data augmentation method for automatic speech recognition," *arXiv preprint arXiv:1904.08779*, 2019.