Explainability Of CNNs In EEG-Based Epileptic Seizure Detection Task

Vadim Grubov
Research Institute of Applied Artificial Intelligence and Digital Solutions,
Plekhanov Russian University of Economics,
Moscow, Russia
vvgrubov@gmail.com

2026 10th Scientific School Dynamics of Complex Networks and their Applications (DCNA) | 979-8-3195-2599-4/26/$31.00 ©2026 IEEE | DOI: 10.1109/DCNA72004.2026.11733378

***Abstract*—CNNs achieve strong performance in EEG-based epileptic seizure detection, but their "black-box" nature limits interpretability. Existing interpretability methods for this task suffer from a number of recurring methodological gaps. To address them, we propose a CNN interpretability method that combines Grad-CAM approach with a data-driven occlusion technique. The method yields both global and local frequency-spatial importance maps over a clinically meaningful coordinate system. We validate our method using a clinical-grade EEG dataset from the Pirogov National Medical and Surgical Center.**

***Index Terms*—convolutional neural networks, explainability, EEG, epileptic seizure detection.**

I. Introduction

Epilepsy is a chronic neurological disorder characterized by sudden and recurrent seizures that greatly impact quality of life, threaten health, and increase the risk of cognitive deficits [1]. A diagnosis of epilepsy includes an electroencephalogram (EEG) analysis with seizure detection, which establishes a foundation for future treatment [2]. Despite advances in this area [3], such diagnosis remains a significant clinical challenge. Epileptic seizures are rare and highly variable events [4], which complicates their automated detection. Convolutional neural networks (CNNs) have shown great promise in this area [5], as they reliably outperform classical machine learning methods in image classification and object recognition tasks. The human brain-like structure of CNNs is especially relevant for the seizure detection task, in which clinicians commonly recognize seizures as characteristic patterns in neuroimaging data [6]. However, the "black-box" nature of CNNs hinders clinical adoption — clinicians require understandable predictions to trust and act upon them [7]. This has led to an increased reliance on interpretability methods that can enhance trust in AI-based predictions [8], [9] and serve exploratory purposes, highlighting possibly unknown biomarkers. The demand for such methods is further driven by the rapid growth of AI applications in medicine, where neurology and EEG analysis form one of the most active research directions [10]. Nevertheless, interpretability methods for CNNs face many challenges.

In this work, we analyze key methodological gaps among the methods for interpreting CNN predictions in the EEG-based seizure detection task. Considering these gaps, we propose a new interpretability method that extracts neurophysiologically important features from CNN models in the frequency-spatial domain.

II. Key methodological gaps

In recent years, methods for interpreting CNN predictions in EEG analysis, particularly seizure detection, have received increasing attention. We reviewed several key approaches and identified aspects important for effective interpretability analysis.

A. Feature importance domain

Domain knowledge can greatly benefit interpretability methods in the seizure detection task [11]. Researchers have reported several oscillatory patterns within particular EEG frequency bands that are associated with seizure onset in focal epilepsy [12]. Localizing such patterns (biomarkers) in certain EEG channels strongly indicates the presence of an epileptic focus. It is reasonable to assume that, with information about important frequency ranges and EEG channels, a clinician can establish connections between their domain-based decisions and the CNN's predictions. Thus, for a CNN interpretability method tailored to the seizure detection task, both *frequency-domain* and *spatial-domain interpretations* are crucial. However, many studies neglect the former [13] or the latter [14], so our method aimed to combine both.

B. Multi-scale interpretation

Model-level (*global*) interpretation provides insight into features that are important for a class as a whole. Epileptic seizures have some robust frequency-spatial features that establish them as a distinct class [15]. However, generic seizure patterns are not always viable, as seizures are nonstationary and possess both inter- and intra-patient variability [16]. Sample-level (*local*) interpretation can be more informative, offering insight into how the model makes decisions for a particular seizure. Thus, global and local interpretations can benefit CNN interpretability methods tailored to seizure detection. However, few studies combine both levels of interpretability [13], [17], so our method aimed to do so.

Authorized licensed use limited to: Kant Baltic Federal Univ. Downloaded on October 23,2026 at 15:23:08 UTC from IEEE Xplore. Restrictions apply.

C. Neurophysiological grounding

Typically, validating an interpretability method requires comparing its results to the ground truth (GT). For a seizure detection task, the GT includes frequency ranges and EEG channels crucial for detection. However, most existing datasets lack such annotation as it is not required for clinical diagnosis.

In the absence of GT, many studies rely on *neurophysiological grounding* of important CNN features — the features are compared to known concepts from domain knowledge. In most studies neurophysiological grounding is *post hoc* — important features are identified first, and then domain knowledge is sought to support them [14], [17]. While this approach can help account for features of a particular dataset, it can also introduce bias. We introduced another approach (*ex ante*) — domain knowledge is collected and organized across multiple impactful studies to create "literature evidence"; CNN's important features are then correlated with this evidence. We believe this approach better reflects the method's ability to detect meaningful neurophysiological features.

D. Multi-model validation

In the absence of GT we have no verifiable common-sense expectations about the model's behavior. There are certain pitfalls to accepting interpretability analysis results at face value [18], [19]. To add some common-sense expectations, we introduced *multi-model validation* — we considered two CNN models with different architectural complexities, which were reflected in their performance. We compared important features between the models and correlated the differences in features with the differences in complexity and performance.

III. Materials and methods

A. Dataset

Most seizure detection methods are tested on open EEG datasets [20], which has its pitfalls. Many datasets undergo extensive preprocessing, often manual. For instance, to address class imbalance, many datasets include pre-selected epochs rather than the entire recording. Such pre-selection can introduce bias, and involving humans in data preprocessing contradicts the purpose of automated seizure detection.

For our study, we selected a dataset collected at the Pirogov National Medical and Surgical Center (Moscow, Russia). This dataset was designed to mimic real clinical EEG data obtained during a prolonged monitoring of epileptic patients [21], [22].

Characteristics of the dataset:

- 67 drug-resistant focal epilepsy patients.
- Recording duration: 8–84 hours per patient (955 total).
- Seizures: 1–5 per patient (132 total).
- Features of EEG: 25 channels, 128 Hz sampling rate, "10–20" EEG system, "Micromed" recorder.
- Automated data preprocessing: filtering and artifact removal with Independent Component Analysis [23].
- Natural class imbalance: 4.10 hours of epileptic activity (0.43% of the whole dataset).

B. CNN models

To test our method, we considered the CNN model introduced in our papers [24], [25]. This model employs the reliable ResNet-18 architecture [26] to classify segmented EEG data and detect seizures. The class imbalance problem is addressed through sample selection and data augmentation. Continuous Wavelet Transform (CWT) [27], a versatile tool for multiscale characterization of biomedical signals [28], is used to represent the EEG data in the time-frequency domain. The CNN input for each fragment is a 2D spectrum of wavelet power (WP), with dimensions of "time" and "frequency". The first convolutional layer of ResNet-18 is modified to accept all 25 inputs (EEG channels).

We considered two variants of CNN model:

- *Baseline* — a straightforward CNN-based classifier;
- *Hybrid* — an approach, that combines "filtration" of EEG data with One-Class Support Vector Machine (OCSVM) and further classification with *Baseline* CNN.

The *Hybrid* model was more efficient in seizure detection — it had somewhat decreased *recall* (0.84 vs 0.96) but massively increased *precision* (0.57 vs 0.13) leading to a noticeable increase in *F1-score* (0.68 vs 0.22). This model's high performance is explained by its domain knowledge-based design. Seizures are extreme events — outliers in EEG data — seen as sharp increases in EEG WP in the 2–5 Hz frequency range [22]. This is also consistent with existing domain knowledge: elevated activity in the δ-band (1–4 Hz) is one of the most prominent seizure biomarkers [29]. However, the plain *OCSVM*, a method designed to detect outliers, operating in the δ-band, produced a large number of False Positives (FPs). It can be theorized that seizure is a combination of an outlier in the δ-band and other fine EEG features, probably in other frequency bands. The *Baseline* model could not account for all these features at once, likely due to a limited capacity. The *Hybrid* model used the OCSVM part to "filter out" most of the irrelevant EEG data, leaving only δ-band outliers. Then, the CNN part did not waste its capacity on δ-band features, learning only the fine EEG features to distinguish seizures from other patterns. These results provide common-sense expectations about each model's behavior.

C. Interpretability method

In this work, we propose interpretability method for CNN models processing multi-channel EEG data. Such models can be represented as follows. Let $x = \{x_i\}_{i=1}^{N}$ be raw EEG signals from N electrodes that are segmented into the fragments of length T. These N signals are transformed into 2D image-like spectra, frequency F × time T. Each CNN input sample can be represented as follows:

$$G(x) = \{G(x_i)\}_{i=1}^{N} \in \mathbb{R}^{N \times F \times T}, \tag{1}$$

where G stands for the CWT processing. CNN model $m : \mathbb{R}^{N \times F \times T} \to \mathbb{R}$ maps these inputs to seizure probability scores.

The method provides interpretations through two complementary components: *spatial-domain interpretation* and *frequency-domain interpretation*.

Authorized licensed use limited to: Kant Baltic Federal Univ. Downloaded on October 23,2026 at 15:23:08 UTC from IEEE Xplore. Restrictions apply.

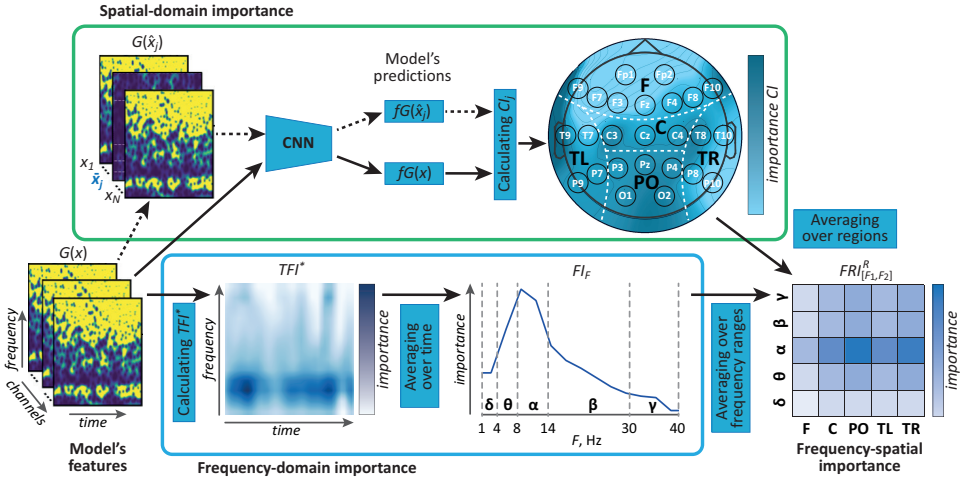


Fig. 1. General pipeline of interpretability method. Two major components are spatial-domain importance assessed with occlusion-based technique, and frequency-domain importance assessed with Grad-CAM approach. The final result is shown as 5×5 heatmap matrix of frequency-spatial importance.

1) Spatial-domain interpretation: Spatial-domain interpretation reveals which brain regions/EEG channels are crucial for prediction. The approach for assessing channel importance (CI) was inspired by the occlusion-based technique for RGB images [30]. A part of the image is occluded with a patch, and significant changes in the model's prediction reflect high importance of the occluded part. However, occlusion suggests "zeroing" the data, and the CNN model does not encounter a 0-signal during training. This could lead to a "domain shift" [31] and render the interpretations unreliable.

We proposed a more data-driven approach. For EEG sample $x = \{x_i\}_{i=1}^{N}$ containing epileptic activity, we introduce a "reference" sample $\bar{x}$ containing non-epileptic activity. $\bar{x}$ is a sample with the minimum average WP chosen from the samples in the 60-second interval preceding the seizure.

The importance of the j-th channel is defined as the change in the model's prediction after replacing the j in the original sample x with its counterpart from the reference sample $\bar{x}$:

$$CI_j = |m(G(x)) - m(G(\hat{x}_j))|,\ \hat{x}_j = \{x_1, \ldots, \bar{x}_j, \ldots, x_N\}. \tag{2}$$

For a brain region R of channels j_r, region importance (RI) can be obtained by averaging CI_j for all channels j_r:

$$RI^R = \frac{1}{|R|} \sum_{j_r \in R} CI_{j_r}. \tag{3}$$

2) Frequency-domain interpretation: Frequency-domain interpretation reveals which frequency bands are crucial for prediction. To assess frequency band importance, we implemented Grad-CAM [32] — a reliable approach for RGB images, that also works well for EEG analysis tasks [33].

In most studies, Grad-CAM is used to highlight important segments of raw EEG signals and interpretability analysis requires additional time-frequency representation of those segments [14]. This approach seems natural — clinicians visually analyze raw EEG signals to detect seizures. Our work is novel in its application of Grad-CAM to spectra obtained from EEG data. We believe our approach is more reliable and controllable for identifying important features in the frequency domain because the frequencies revealed by the subsequent analysis of Grad-CAM-highlighted EEG segments may not necessarily align with what the CNN has truly learned.

The Grad-CAM saliency map is used to introduce the frequency importance (FI) of band $[F_1, F_2]$ as follows:

$$FI_{[F_1,F_2]} = \frac{1}{(F_2 - F_1)T} \int_{F_1}^{F_2} \int_{0}^{T} TFI^*(f,t)\, dtdf, \tag{4}$$

where TFI^* is the upscaled saliency map.

3) Frequency-spatial interpretation: Joint frequency band and brain region importance (FRI) is defined as:

$$FRI^R_{[F_1,F_2]} = FI_{[F_1,F_2]} \cdot RI^R. \tag{5}$$

Our interpretability method is inspired by human pattern detection rules: clinicians often search for elevated activity in specific frequency bands and brain regions to identify seizure biomarkers. We presented the results of interpretation using the same "coordinate system": $Bands \times Regions$. We considered five classic frequency bands (δ, θ, α, β, γ) and divided 25 EEG electrodes into five brain regions (frontal F, central C, parietal-occipital PO, temporal left TL, temporal right TR). $FRI^R_{[F_1,F_2]}$ can be visualized as 5×5 importance matrix and represents the local interpretation. The global interpretation of the model can be obtained by averaging $FRI^R_{[F_1,F_2]}$ over all samples identified as epileptic activity. Important features in the matrices were identified through statistical testing. Fig. 1 illustrates the pipeline of the interpretability method.

D. Literature evidence analysis

We performed a systematic analysis of the appearance of different frequency-spatial biomarkers in the literature according to our introduced $Bands \times Regions$ "coordinate system". We considered a total of 38 related works and assigned an "evidence" score E to each biomarker based on how many

Authorized licensed use limited to: Kant Baltic Federal Univ. Downloaded on October 23,2026 at 15:23:08 UTC from IEEE Xplore. Restrictions apply.

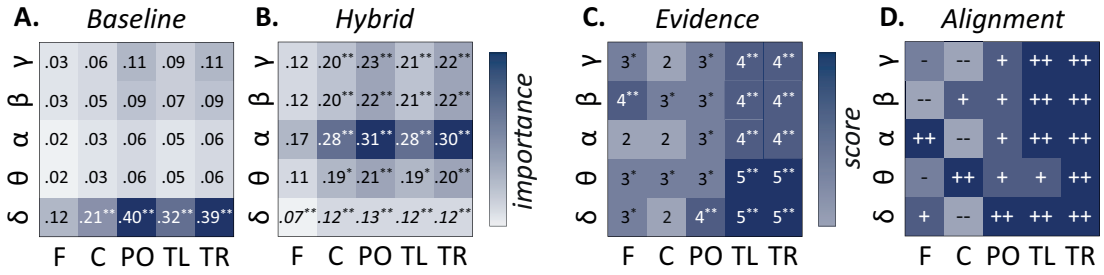


Fig. 2. Results of global interpretation: frequency-spatial importance of epileptic EEG features for *Baseline* (A) and *Hybrid* (B) models. Significant frequency-spatial features are shown with "**", while "*" denotes "borderline" cases. Note that in the *Hybrid* model all δ-band features are considered significant due to their involvement in the OCSVM part. Evidence matrix of scores E (C). "**" denotes "strong" evidence and "*" denotes "moderate" evidence. Alignment matrix (D). Legend: "++" — "hit", "+" — "weak hit", "-" — "weak miss", "--" — "miss".

times it was mentioned in these works. The results were presented in the form of an evidence matrix, which is similar in structure to an importance matrix. The intersections between these two matrices show which important domain knowledge biomarkers are detected by our interpretability method. This provides neurophysiological grounding *ex ante*.

IV. Results and Discussion

A. Results of interpretation

Fig. 2A,B illustrates the results of global interpretation for two CNN models. Significant frequency-spatial features are shown with "**", while "*" denotes "borderline" cases.

Fig. 2A shows that the *Baseline* model focuses on the δ-band, with high importance values across all brain regions. The contribution of other bands (e.g., γ) is non-zero, but remains low. These findings align with existing domain knowledge: epileptic seizures represent a hypersynchronous state of the cerebral cortex, that commonly manifests itself on EEG as an elevated activity in the δ-band [29]. However, non-zero importance values in other bands suggest that the model still "attempts" to utilize them, which aligns with the hypothesis about important biomarkers in other frequency bands [25].

Fig. 2B shows that the CNN part of the *Hybrid* model does not rely on the δ-band. Instead, high importance values are found in all the other bands, with the α-band exhibiting the highest. These results align with the model's design. In the first stage, OCSVM "filters out" trivial examples characterized by pronounced activity in the δ-band. In the second stage, the CNN can dedicate its full capacity to discerning subtle differences within the other frequency bands. The *Hybrid* model's higher *precision* and *F1-score* suggest that the θ-, α-, β-, and γ-bands also play important roles in seizure detection, particularly in distinguishing true seizures from other patterns with high δ-band activity. These results contribute to the validation of our method since the results of interpretability analysis meet the initial common-sense expectations.

The results of local interpretation reveal the temporal evolution of feature importance and suggest a potential temporal scenario for seizure evolution. While this pattern probably varies between patients, epilepsy types, or even seizures within a patient, it reveals fine-grained temporal dynamics in EEG signals that is considered by the CNN model. Such dynamics could contribute to seizure detection.

B. Neurophysiological grounding

1) Neurophysiological grounding post hoc: First, we provided neurophysiological grounding *post hoc* by finding literature evidence that supports the identified features.

The analysis of important brain regions revealed low importances assigned to the frontal region. According to domain knowledge, spatial features are typically localized in areas of epileptic foci [34]. Patients from the present dataset had epileptic foci primarily in the temporal and occipital regions, which aligns with highest importance values in TR and PO regions for both models. Thus, the results of interpretability analysis in the spatial domain meet preliminary expectations.

High feature importances in frequency domain can be tied to specific neurophysiological processes. Activity in the δ-band (1–4 Hz) is often associated with a hypersynchronous state of the cerebral cortex [29]. The θ-rhythm (4–8 Hz) in the occipital region of the cortex is often linked to the activity of thalamocortical neural network. In the case of epilepsy, important features in the θ-band may reflect the underlying mechanisms of seizure generation and propagation, which involve not only neuronal but also glial processes, e.g., reactive astrocytes shaping the transitions between normal and ictal states [35]. The combined α and β bands (8–30 Hz) in the occipital and temporal regions are often associated with the default mode network (DMN) activity. DMN suppression typically occurs in altered states or loss of consciousness such as seizures [36]. Note that such large-scale networks are commonly characterized with functional connectivity measures, ranging from classical pairwise coupling to higher-order interactions [37], which suggests that the importance of individual regions may partially reflect their coordinated network activity. The γ-band (30–40 Hz and above) is typically associated with motor activity that emerges during the convulsive phase of a seizure and is considered an artifact. However, some studies show [38] that γ-band activity can also be a seizure biomarker.

2) Neurophysiological grounding ex ante: Additional neurophysiological grounding was performed using pre-collected literature evidence. Fig. 2C shows the evidence matrix of scores E. We divided the evidence into "strong" ($E = 4, 5$,

Authorized licensed use limited to: Kant Baltic Federal Univ. Downloaded on October 23,2026 at 15:23:08 UTC from IEEE Xplore. Restrictions apply.

"**" in Fig. 2C), "moderate" ($E = 3$, "*" in Fig. 2C), and "weak" ($E < 3$). Then we investigated alignment between the importance matrix of the *Hybrid* model (Fig. 2B) and the evidence matrix (Fig. 2C). The resulting alignment matrix is shown in Fig. 2D. We considered four possible cases:

- "hit" ("++"): $FRI^{**} \cap E^{**}$ or $FRI^{*} \cap E^{*}$ or $FRI \cap E$;
- "weak hit" ("+"): $FRI^{**} \cap E^{*}$ or $FRI^{*} \cap E^{**}$;
- "weak miss" ("-"): $FRI^{*} \cap E$ or $FRI \cap E^{*}$;
- "miss" ("- -"): $FRI^{**} \cap E$ or $FRI \cap E^{**}$.

Fig. 2D reveals a significant overlap between literature evidence and the results of interpretability analysis. We believe these results complement those obtained in Section IV-B1.

V. Conclusion

We presented a new interpretability method for CNNs operating on multi-channel EEG in the seizure detection task. The method combines a Grad-CAM-based frequency-domain component with an occlusion-based spatial-domain component built on data-driven "reference" samples rather than zero or surrogate signals. Importance is reported jointly in a $Bands \times Regions$ coordinate system that mirrors how clinicians reason about epileptic biomarkers, and the method supports both global (model-level) and local (sample-level) interpretation without architectural changes or retraining.

By combining frequency-spatial coverage, multi-scale interpretation, *ex ante* grounding, and multi-model validation in a single pipeline, the method narrows the gap between CNN-based seizure detection and clinical EEG analysis, and contributes a practical XAI tool. A natural extension of the method is to move from the importance of individual regions to the importance of interactions between them, using wavelet-based functional connectivity [28] and higher-order network approaches [37].

References

[1] 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.

[2] B. C. Jobst and G. D. Cascino, "Resective epilepsy surgery for drug-resistant focal epilepsy: a review," *Jama*, vol. 313, no. 3, pp. 285–293, 2015.

[3] L. Harris and H. Angus-Leppan, "Epilepsy: diagnosis, classification and management," *Medicine*, vol. 48, no. 8, pp. 522–528, 2020.

[4] H. Anwar, Q. U. Khan, N. Nadeem, I. Pervaiz, M. Ali, and F. F. Cheema, "Epileptic seizures," *Discoveries*, vol. 8, no. 2, p. e110, 2020.

[5] U. R. Acharya, S. L. Oh, Y. Hagiwara, J. H. Tan, and H. Adeli, "Deep convolutional neural network for the automated detection and diagnosis of seizure using EEG signals," *Computers in biology and medicine*, vol. 100, pp. 270–278, 2018.

[6] S.-A. Lee, D. D. Spencer, and S. S. Spencer, "Intracranial eeg seizure-onset patterns in neocortical epilepsy," *Epilepsia*, vol. 41, no. 3, pp. 297–307, 2000.

[7] S. Reddy, "Explainability and artificial intelligence in medicine," *The Lancet Digital Health*, vol. 4, no. 4, pp. e214–e215, 2022.

[8] R. Ibrahim and M. O. Shafiq, "Explainable convolutional neural networks: A taxonomy, review, and future directions," *ACM Computing Surveys*, vol. 55, no. 10, pp. 1–37, 2023.

[9] A. V. Andreev, S. A. Kurkin, D. Stoyanov, A. A. Badarin, R. Paunova, and A. E. Hramov, "Toward interpretability of machine learning methods for the classification of patients with major depressive disorder based on functional network measures," *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 33, no. 6, 2023.

[10] 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.

[11] C. E. Elger and C. Hoppe, "Diagnostic challenges in epilepsy: seizure under-reporting and seizure detection," *The Lancet Neurology*, vol. 17, no. 3, pp. 279–288, 2018.

[12] S. Lagarde, S. Buzori, A. Trebuchon, R. Carron, D. Scavarda, M. Milh, A. McGonigal, and F. Bartolomei, "The repertoire of seizure onset patterns in human focal epilepsies: determinants and prognostic values," *Epilepsia*, vol. 60, no. 1, pp. 85–95, 2019.

[13] T. Tuncer and S. Dogan, "An explainable EEG epilepsy detection model using friend pattern," *Scientific Reports*, vol. 15, no. 1, p. 16951, 2025.

[14] X. Zhao, N. Yoshida, T. Ueda, H. Sugano, and T. Tanaka, "Epileptic seizure detection by using interpretable machine learning models," *Journal of Neural Engineering*, vol. 20, no. 1, p. 015002, 2023.

[15] K. D. Tzimourta, A. T. Tzallas, N. Giannakeas, L. G. Astrakas, D. G. Tsalikakis, P. Angelidis, and M. G. Tsipouras, "A robust methodology for classification of epileptic seizures in EEG signals," *Health and Technology*, vol. 9, pp. 135–142, 2019.

[16] A. Van Esbroeck, L. Smith, Z. Syed, S. Singh, and Z. Karam, "Multi-task seizure detection: addressing intra-patient variation in seizure morphologies," *Machine Learning*, vol. 102, no. 3, pp. 309–321, 2016.

[17] A. Vilamala, K. H. Madsen, and L. K. Hansen, "Deep convolutional neural networks for interpretable analysis of EEG sleep stage scoring," in *2017 IEEE 27th international workshop on machine learning for signal processing (MLSP)*. IEEE, 2017, pp. 1–6.

[18] J. Adebayo, J. Gilmer, M. Muelly, I. J. Goodfellow, M. Hardt, and B. Kim, "Sanity checks for saliency maps," in *Neural Information Processing Systems*, 2018.

[19] P. Sturmfels, S. Lundberg, and S.-I. Lee, "Visualizing the impact of feature attribution baselines," *Distill*, 2020, https://distill.pub/2020/attribution-baselines.

[20] S. Wong, A. Simmons, J. Rivera-Villicana, S. Barnett, S. Sivathamboo, P. Perucca, Z. Ge, P. Kwan, L. Kuhlmann, R. Vasa *et al.*, "EEG datasets for seizure detection and prediction—A review," *Epilepsia Open*, vol. 8, no. 2, pp. 252–267, 2023.

[21] 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.

[22] 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.

[23] D. Mantini, M. G. Perrucci, S. Cugini, A. Ferretti, G. L. Romani, and C. Del Gratta, "Complete artifact removal for eeg recorded during continuous fmri using independent component analysis," *Neuroimage*, vol. 34, no. 2, pp. 598–607, 2007.

[24] 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.

[25] 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.

[26] 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.

[27] A. E. Hramov, A. A. Koronovskii, V. A. Makarov, V. A. Maksimenko, A. N. Pavlov, and E. Sitnikova, *Wavelets in Neuroscience*. Springer Nature, 2021.

[28] L. Kralj, S. Kurkin, H. Lenasi, A. Pavlov, A. Stožer, A. E. Hramov, M. Perc, and M. Gosak, "Wavelet analysis in biomedical research: A versatile tool for multiscale signal characterization and functional connectivity analysis," *Physics of Life Reviews*, vol. 59, pp. 19–47, 2026.

[29] P. Jiruska, M. De Curtis, J. G. Jefferys, C. A. Schevon, S. J. Schiff, and K. Schindler, "Synchronization and desynchronization in epilepsy:

Authorized licensed use limited to: Kant Baltic Federal Univ. Downloaded on October 23,2026 at 15:23:08 UTC from IEEE Xplore. Restrictions apply.

controversies and hypotheses," *The Journal of physiology*, vol. 591, no. 4, pp. 787–797, 2013.
[30] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," in *Computer Vision – ECCV 2014*. Springer International Publishing, 2014, pp. 818–833.
[31] J. Quiñonero-Candela, M. Sugiyama, A. Schwaighofer, and N. D. Lawrence, *Dataset shift in machine learning*. Mit Press, 2022.
[32] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual explanations from deep networks via gradient-based localization," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 618–626.
[33] Y. Zhao, L. Cao, Y. Ji, B. Wang, and W. Wu, "Interpretable EEG emotion classification via CNN model and gradient-weighted class activation mapping," *Brain Sciences*, vol. 15, no. 8, p. 886, 2025.
[34] H. Daoud and M. Bayoumi, "Deep learning approach for epileptic focus localization," *IEEE transactions on biomedical circuits and systems*, vol. 14, no. 2, pp. 209–220, 2019.
[35] Y. Tsybina, I. Kastalskiy, V. B. Kazantsev, A. E. Hramov, and S. Gordleeva, "Beyond the neuron: Unveiling the role of reactive astrocytes in epileptic seizure dynamics through self-organized bistability," *Computers in Biology and Medicine*, vol. 197, p. 110980, 2025.
[36] N. B. Danielson, J. N. Guo, and H. Blumenfeld, "The default mode network and altered consciousness in epilepsy," *Behavioural neurology*, vol. 24, no. 1, pp. 55–65, 2011.
[37] S. A. Kurkin, A. N. Pisarchik, L. A. Mayorova, and A. E. Hramov, "Evolution of methods for assessing fMRI-based functional networks: from classical pairwise connectivity to higher-order interactions," *Physics Reports*, vol. 1174, pp. 1–66, 2026.
[38] S. Luca, P. Karsmakers, K. Cuppens, T. Croonenborghs, A. V. de Vel, B. Ceulemans, L. Lagae, S. V. Huffel, and B. Vanrumste, "Detecting rare events using extreme value statistics applied to epileptic convulsions in children," *Artificial intelligence in medicine*, vol. 60 2, pp. 89–96, 2014.

Authorized licensed use limited to: Kant Baltic Federal Univ. Downloaded on October 23,2026 at 15:23:08 UTC from IEEE Xplore. Restrictions apply.