Explainable Foundation Models for Epileptic Seizure Detection in Large-Scale Imbalanced EEG Data

Aitor Rodríguez-Dueñas[1⋆][0009−0004−8705−5887], Iana Miranda-Caramé[1][0009−0000−1946−4655], Ana Rodríguez-López[1][0009−0000−2672−9703], and Manuel J. Jiménez-Navarro[1][0000−0001−8514−4182]

[1] Department of Computer Languages and Systems, Universidad de Sevilla Av. Reina Mercedes S/N, 41012 Seville, Spain,
[2] SCORE Lab, I3US Institute, Universidad de Sevilla, Seville, Spain
{arodriguez56, mjimenez3, arodriguez24}@us.es, ianmircar@alum.us.es

Abstract. Automating the detection of epileptic seizures through electroencephalography analysis remains a significant challenge in clinical neurology, particularly due to the high volume of continuous recordings and the severe class imbalance between ictal and non-ictal events. In this work, tabular foundation models are evaluated for seizure detection under a strict patient-independent Leave-One-Subject-Out cross-validation framework, assessing their capacity to generalise across unseen patients without domain-specific fine-tuning. EEG signals are preprocessed and transformed into tabular feature representations, from which the 50 most discriminative descriptors are selected per fold. Two foundation models, TabPFN and TabICL, are benchmarked against classical machine learning classifiers including Random Forest, XGBoost, SVM and k-NN. Results show that TabPFN achieves the best overall performance, with a ROC-AUC of 0.87 and F1 Score of 0.56, outperforming all classical baselines under cross-subject evaluation. Explainability analysis via SHAP and LIME consistently identifies temporal EEG channels and dynamic signal descriptors as the most discriminative features for ictal activity, providing clinically interpretable model decisions. These findings highlight the potential of tabular foundation models as scalable and annotation-efficient tools for patient-independent EEG seizure detection.

Keywords: EEG signal processing · Seizure detection · Biomedical machine learning · Foundation models · Explainable AI.

1 Introduction

Epilepsy is one of the most common neurological disorders worldwide, affecting millions of individuals and significantly impacting quality of life due to the unpredictable nature of seizures [20]. Electroencephalography (EEG) remains one

⋆ Corresponding author.

of the primary non-invasive tools for monitoring brain activity and supporting the diagnosis and management of epileptic disorders. In recent years, automatic seizure detection and prediction systems based on machine learning have gained considerable attention because of their potential to assist clinicians, enable continuous patient monitoring and improve early warning systems in both hospital and wearable environments. Despite the progress achieved in automated EEG analysis, seizure detection remains a highly challenging task. EEG signals are inherently non-stationary, highly variable across patients and often contaminated by noise and motion artifacts [5]. Furthermore, epileptic datasets usually present severe class imbalance, where seizure events represent only a small fraction of the recorded data. Another major challenge is the strong inter-subject variability observed in EEG recordings, since patterns associated with seizures may differ significantly between patients due to differences in brain activity, electrode placement, age, medication and seizure type.

Traditional approaches for seizure detection have mainly relied on handcrafted feature extraction combined with classical machine learning algorithms. Commonly used features include statistical descriptors, spectral characteristics, entropy-based measures and wavelet representations, which are later processed by classifiers such as Support Vector Machines (SVM), Random Forests (RF), k-Nearest Neighbors (KNN), or Gradient Boosting methods. Although these approaches can achieve competitive performance in controlled scenarios, they often depend heavily on domain-specific feature engineering and may struggle to generalize across unseen patients [2].

Tabular foundation models have recently emerged as a promising paradigm for structured data analysis. Pretrained exclusively on synthetic datasets, models such as TabPFN [12] and TabICL [18] perform classification and regression through in-context learning, without task-specific fine-tuning, and show strong generalisation across heterogeneous tabular domains [17]. However, their application to biomedical EEG analysis remains largely unexplored, as most foundation models were designed for generic time-series forecasting rather than neurophysiological signal analysis. It is still unclear whether these pretrained architectures can capture seizure-related EEG patterns and improve cross-subject generalization under strict Leave-One-Subject-Out (LOSO) protocols. Moreover, their interpretability remains insufficiently investigated, despite explainability being a critical requirement in clinical decision-support systems [14].

Explainable Artificial Intelligence techniques such as SHAP and LIME have become increasingly important in medical machine learning applications because they provide insights into model decisions and improve transparency and trustworthiness. In EEG seizure detection, explainability can help identify relevant channels, temporal regions, or signal characteristics associated with epileptic activity. Nonetheless, most existing studies focus either on performance optimization or isolated explainability analyses, while comprehensive comparisons between classical machine learning methods and foundation models under explainable and patient independent settings remain scarce [19].

The main contributions of this work are summarized as follows:

- **Foundation Model Benchmarking:** A comparative evaluation of tabular foundation models and classical machine learning approaches for EEG-derived tabular feature representations for seizure detection under a strict patient-independent (LOSO) cross-validation framework.
- **Clinically-Oriented Explainability Analysis:** Joint application of SHAP and LIME to identify neurologically meaningful EEG channels and temporal descriptors associated with ictal activity, providing transparent and clinician-interpretable model decisions.
- **Reproducible End-to-End Pipeline:** Development and validation of a complete processing pipeline, from raw EEG preprocessing and feature engineering to model evaluation and interpretability analysis. The complete source code is publicly available at [1].

The remainder of this paper is organized as follows. Section 2 reviews related work on EEG-based seizure detection, tabular foundation models and explainable AI in clinical settings. Section 3 details the proposed methodology, including data preprocessing, feature engineering and model implementation. Section 4 presents the experimental findings and explainability analysis. Finally, Section 5 discusses the implications of the results and outlines future research directions.

2 Related work

Early automated seizure detection systems mainly relied on handcrafted feature extraction combined with traditional machine learning classifiers. These approaches typically extracted statistical, spectral and time-frequency features from EEG recordings, including entropy measures, power spectral density, wavelet coefficients, Hjorth parameters and energy-based descriptors. The extracted features were then used as inputs for classifiers such as Support Vector Machines (SVM), RF, KNN, Logistic Regression and Gradient Boosting methods [9].

Several studies reported competitive seizure detection performance using wavelet-based representations due to their ability to capture non-stationary EEG patterns in both time and frequency domains. Entropy-based methods were also widely adopted to characterize the irregular and chaotic behavior commonly associated with epileptic activity. More recently, large-scale feature extraction frameworks such as TSFresh [4] enabled the automatic computation of hundreds of temporal descriptors, reducing the need for manually designed features while maintaining compatibility with classical machine learning models [16]. Despite their effectiveness, classical machine learning approaches present several limitations. Most methods strongly depend on feature engineering and expert knowledge, which may reduce adaptability across datasets and patient populations. Furthermore, handcrafted features may fail to capture highly complex temporal dependencies present in EEG recordings. As a result, the generalization capability of these models under patient-independent evaluation settings often remains limited [10].

More recently, Transformer-based architectures have gained attention in EEG analysis for their self-attention mechanisms, which model global temporal relationships more efficiently than recurrent methods. Transformers have shown promising results in several biomedical time-series tasks, including seizure detection and prediction. Nevertheless, many deep learning approaches still require large annotated datasets and may show reduced robustness when evaluated on unseen patients under strict cross-subject protocols [3]. Although foundation models perform well in generic temporal applications, their use in biomedical EEG analysis remains limited, as most were designed for forecasting rather than clinical signal interpretation. Consequently, it is still unclear whether these architectures can capture seizure-related EEG patterns, especially under patient-independent methodologies such as LOSO validation [15]. Furthermore, current studies evaluating foundation models in healthcare often focus on performance metrics while offering limited analysis of robustness, interpretability and clinical applicability, creating a research gap in explainable, patient-independent seizure detection [13].

Explainability has become an essential requirement in medical artificial intelligence systems due to the need for transparency, reliability and clinician trust. In EEG seizure detection, explainable AI (XAI) techniques can help identify relevant EEG channels, temporal segments, or signal characteristics associated with epileptic activity, thereby facilitating clinical interpretation of model decisions. Model-agnostic approaches such as SHAP and LIME have been widely used to interpret machine learning predictions by estimating feature importance and local decision behavior. In deep learning systems, saliency maps, Grad-CAM and attention visualization techniques have also been applied to highlight discriminative EEG regions related to seizure activity [8].

While recent studies have begun exploring foundation models in EEG-related applications [17], these works typically rely on subject-dependent evaluation settings or do not assess cross-subject generalizability under strict LOSO protocols. Furthermore, comparative analyses between tabular foundation models and classical machine learning baselines under simultaneously explainable and patient-independent frameworks remain scarce in the current literature [11]. The present work addresses these gaps by benchmarking TabPFN and TabICL against classical classifiers under a rigorous 12 fold LOSO evaluation, and by jointly applying SHAP and LIME to assess the consistency and clinical interpretability of the resulting explanations.

3 Methodology

The methodology for this study follows a systematic pipeline designed to handle large EEG data while ensuring model interpretability and robust performance. The experimental pipeline is shown in Figure 1.

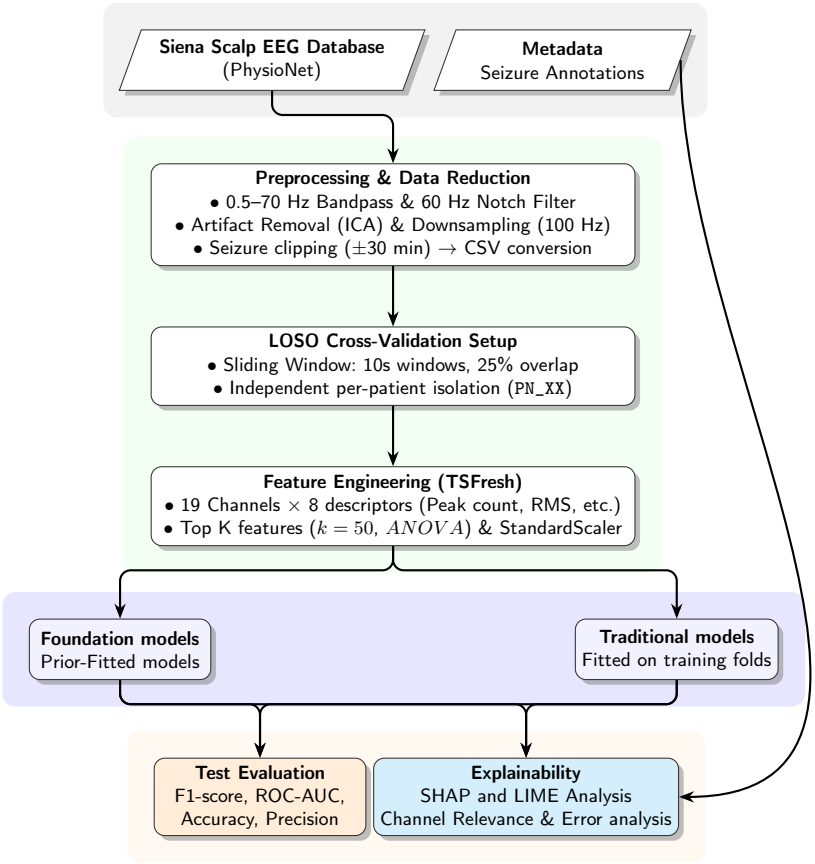


Fig. 1: Methodological pipeline for seizure classification, illustrating data acquisition, preprocessing, LOSO cross-validation, feature engineering and model evaluation with SHAP explainability.

3.1 Data Acquisition

The process begins with the acquisition of the Siena Scalp EEG Database [6,7] from PhysioNet. The dataset contains high-resolution EEG recordings from 14 patients, captured following the international 10–20 electrode placement system. Signals were recorded at a sampling frequency of 512 Hz and stored in European Data Format (.edf) files. The data is organized into patient-specific folders containing multiple recording sessions, along with metadata files describing session information and seizure onset and offset timestamps.

3.2 Preprocessing and Data Reduction

The preprocessing pipeline starts from the raw EDF recordings of the Siena Scalp EEG Database. Seizure annotations are extracted from metadata text files by

parsing seizure start and end times and converting them into timestamps relative to the recording. The signals are then filtered using a 0.5–70 Hz bandpass filter and a 60 Hz notch filter to remove noise and power-line interference. Independent Component Analysis (ICA) is applied to reduce artefacts such as eye movements and muscle activity and the cleaned signals are downsampled to 100 Hz. To reduce storage requirements and class imbalance, only a ±30-minute interval around seizure events is preserved.

3.3 LOSO Cross-Validation

The evaluation framework is based on a LOSO cross-validation strategy applied over the 12 patients with sufficient seizure annotations, yielding 12 independent folds, consequently, 2 patients were excluded due to insufficient seizure annotations. In each fold, one patient is held out for testing while the remaining patients are used for training, ensuring a fully patient-independent evaluation.

The pipeline within each fold follows a strict sequential order to prevent information leakage. Recordings are segmented into 10-second sliding windows with 25% overlap within each patient session, ensuring no window crosses patient or session boundaries. Each window is labelled as seizure if at least one ictal sample is present, a conservative criterion adopted to maximize sensitivity at seizure onset and offset boundaries. TSFRESH [4] then extracts over 15,000 candidate features per window across 19 channels, computed per patient independently before the LOSO loop begins. Within each fold, feature selection retaining the $k = 50$ most discriminative features via ANOVA scoring and standard normalisation are fitted exclusively on the training patients and applied to the test patient, ensuring that no information from the held-out subject influences any fitting step. For zero-shot foundation models, no training occurs, only the classification threshold is tuned on the training patients of each fold. All other models are trained from scratch in every fold.

Two foundation models are evaluated: TabPFN [12], a Prior-Fitted Network pre-trained on synthetic tabular data and TabICL [18], a transformer-based model that performs classification through context learning. For zero-shot foundation models, no training occurs, only the classification threshold is tuned on the training patients of each fold. All other models are trained from scratch in every fold.

Test Evaluation: After each fold, model performance is evaluated and reported using Accuracy, Recall, Precision, F1-score, F1-macro, F1-micro and ROC-AUC. Metric definitions and justification are provided in Section 4.

3.4 Explainability and Evaluation

To enhance clinical interpretability, both SHAP and LIME are applied to TabPFN, the best-performing model, using the test predictions aggregated across all 12 LOSO folds. The feature vectors and corresponding predictions from each fold are concatenated into a single pool and global feature importance is computed over this combined set. Both methods target the positive class (seizure), as the

clinical objective is to identify which EEG descriptors drive ictal predictions. Feature importance scores are aggregated by computing the mean absolute SHAP value and the mean LIME coefficient per feature across all samples and subsequently averaged by EEG channel to produce the channel-level importance rankings shown in Figure 3. This joint application of SHAP and LIME provides complementary and mutually reinforcing explanations of the classification decisions, increasing confidence that the identified patterns are not method dependent.

4 Results

This section presents the experimental evaluation of the proposed models for automated seizure detection.

4.1 Performance

Model performance is evaluated using the following metrics. Accuracy measures the proportion of correctly classified windows but is uninformative under severe class imbalance. Precision and Recall capture the trade-off between false positives and false negatives for the seizure class. The F1-score combines both into a single measure, while F1-macro averages F1 across both classes equally, penalizing poor minority-class detection regardless of majority-class performance. ROC-AUC summarizes discrimination capability across all decision thresholds. Given the extreme imbalance between seizure and non-seizure windows, F1-macro and ROC-AUC are considered the primary metrics throughout this work. Table 1 presents the LOSO evaluation across patients, a challenging setting that measures the ability of models to generalize to unseen subjects.

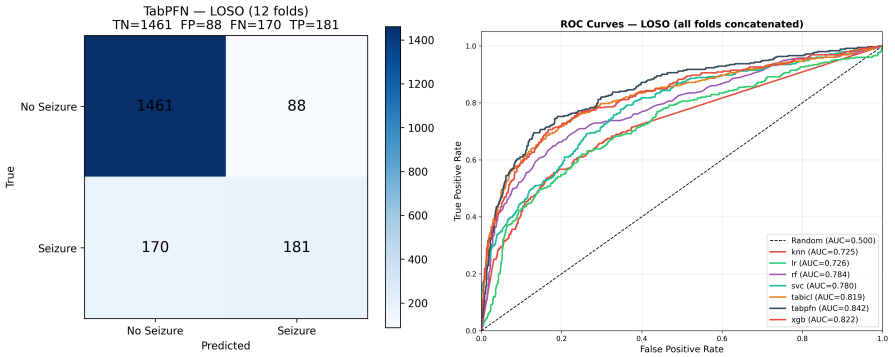
Table 1: Performance comparison of the evaluated models (mean ± std over LOSO folds)

Model	Accuracy	Precision	Recall	F1 Score	F1 Macro	ROC AUC
XGB	0.84 ± 0.07	0.61 ± 0.26	0.51 ± 0.26	0.51 ± 0.20	0.71 ± 0.11	0.84 ± 0.07
LR	0.80 ± 0.12	0.43 ± 0.35	0.45 ± 0.35	0.40 ± 0.30	0.63 ± 0.16	0.73 ± 0.18
KNN	0.77 ± 0.14	0.49 ± 0.23	0.45 ± 0.29	0.42 ± 0.22	0.63 ± 0.14	0.70 ± 0.12
RF	0.84 ± 0.09	0.68 ± 0.32	0.46 ± 0.28	0.48 ± 0.25	0.69 ± 0.14	0.82 ± 0.07
SVC	0.80 ± 0.10	0.45 ± 0.26	0.44 ± 0.34	0.40 ± 0.27	0.63 ± 0.14	0.77 ± 0.10
TabPFN	$\mathbf{0.85 \pm 0.09}$	$\mathbf{0.71 \pm 0.21}$	0.56 ± 0.24	$\mathbf{0.56 \pm 0.19}$	$\mathbf{0.73 \pm 0.12}$	$\mathbf{0.87 \pm 0.05}$
TabICL	0.84 ± 0.09	0.57 ± 0.24	$\mathbf{0.59 \pm 0.26}$	0.55 ± 0.22	0.72 ± 0.13	0.83 ± 0.06

Overall, classical models such as KNN, Logistic Regression and SVC achieve moderate accuracy but relatively low Recall and F1 Score, suggesting difficulties in correctly detecting seizure events in imbalanced EEG data. Ensemble

methods like Random Forest and XGBoost provide more stable and balanced performance, achieving higher ROC AUC and F1 values, which indicates better discrimination between seizure and non-seizure samples. The best overall performance is obtained by the tabular foundation models, especially TabPFN. This model achieves the highest Accuracy, Precision, F1 Score, F1 Macro and ROC-AUC, demonstrating strong generalization and discriminative capability under cross-subject evaluation. In contrast, TabICL obtains the highest Recall, meaning it detects a larger proportion of seizures, although with a slight reduction in Precision. The relatively high standard deviations across some metrics reflect the strong variability between patients, which is a common challenge in EEG-based seizure detection. Overall, the results suggest that tabular foundation models are promising approaches for EEG analysis, particularly when EEG signals are transformed into feature-based tabular representations.

The AUC reported in Table 1 (0.87 ± 0.05) represents the mean across the 12 LOSO folds, while the value shown in Figure 2b (0.842) is computed over all fold predictions concatenated into a single pool. TabPFN's curve dominates across the full range of false positive rates, followed by XGBoost (0.822) and TabICL (0.819), while classical models remain notably below (AUC $\approx$ 0.725). The confusion matrix over 12 LOSO folds (TN=1461, FP=88, FN=170, TP=181) shows that TabPFN correctly identifies the majority of non-seizure windows while achieving a seizure recall of 51.6%, maintaining a clinically meaningful decision boundary that prioritises specificity without entirely sacrificing sensitivity to ictal events.



(a) TabPFN confusion matrix (b) ROC AUC graph

Fig. 2: Results of the evaluation metrics

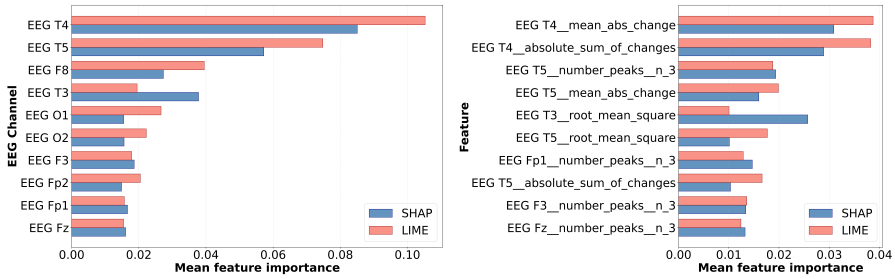
4.2 Explainability analysis

Two widespread explainability techniques have been employed to analyze TabPFN which demonstrated to be the best model. In this context, *root_mean_square*

represents the effective energy of the signal within each analysis window, capturing how strongly brain activity is expressed in terms of amplitude. The features *longest_strike_above_mean* and *longest_strike_below_mean* measure persistence, that is, how long the waveform remains consistently above or below its mean level, which reflects sustained activation or sustained suppression patterns.

In parallel, *number_peaks_n* quantifies how frequently salient transient events occur by counting local maxima with prominence above a predefined threshold (e.g., 3 or 5), thus describing burst-like or oscillatory activity. Finally, *absolute_sum_of_changes* summarizes the cumulative sample-to-sample variation, providing a compact measure of signal dynamism and short-term temporal complexity. The closely related descriptor *mean_abs_change* captures the average magnitude of consecutive sample differences, providing a normalised measure of the same phenomenon. Taken together, these features explain why the model is especially sensitive to both the intensity and the temporal structure of EEG activity.

Figure 3 indicates that the model bases its decisions on a combination of complementary EEG descriptors rather than on a single dominant predictor.



(a) Importance by channel.

(b) Importance by feature.

Fig. 3: Explainability Metrics: SHAP and LIME Feature Importance for TabPFN.

The two new global explainability plots are coherent and mutually reinforcing. First, SHAP and LIME provide a very similar ranking structure, which increases confidence that the interpretation is not method-dependent. Second, the dominant channels correspond mainly to temporal electrodes (T4, T5, T3), with additional but smaller contributions from frontal/occipital regions. At feature-family level, the model relies primarily on descriptors of signal change and temporal dynamics (e.g., absolute/mean changes and peak-related features), rather than static amplitude alone. This indicates that the classifier is learning seizure-relevant transient behavior and variability patterns over time. A useful global interpretation is therefore: TabPFN decisions are concentrated on temporal-channel dynamics and change-sensitive biomarkers, consistent with seizure dis-

crimination being more about evolving electrophysiological structure than single-point magnitude effects.

4.3 Error Analysis

The error analysis purpose is to find some anthropometric pattern which justifies the performance differences between the different groups for the TabPFN model. Figure 4 shows the ROC-AUC averaged for the different anthropometric properties. In general, it can be observed that most differences are driven more by patient-specific difficulty than by a single demographic variable. Both tabular models tend to succeed and fail on similar patients, which suggests they are responding to shared properties of the EEG data rather than random model instability.

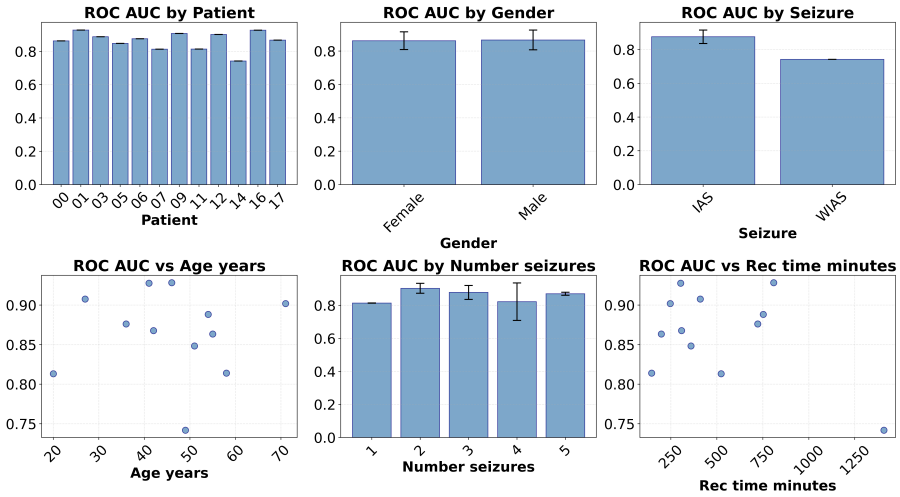


Fig. 4: TabPFN Error Analysis: Patient Metadata.

Regarding the example of patient 14, an important observation is that this patient is the subject with lower performance. This patient is the only one with Focal Onset Without Impaired Awareness in the studied dataset. Note that the sample size is not big enough to consider this kind of seizures as more complex to model, in fact, these seizures are commonly referred to as simple partial seizures. The patients with clearer performance degradation are better explained by a combination of factors such as atypical seizure profile, very long recordings, or fewer seizure events, which can make class separation harder in LOSO settings. Overall, the analysis does not support a conclusion on an isolated metadata attribute.

5 Conclusion

This work presents a comparative evaluation of tabular foundation models applied to handcrafted EEG-derived features for epileptic seizure detection under a strict patient-independent LOSO cross-validation framework, demonstrating that TabPFN consistently outperforms classical baselines with the highest ROC-AUC (0.87), Precision (0.71) and F1 Score (0.56), while TabICL achieves the highest Recall (0.59). The explainability analysis via SHAP and LIME identifies temporal channels T4, T5, and T3 as the most discriminative regions through dynamic signal descriptors, confirming that TabPFN captures ictal activity without domain-specific retraining, with residual performance variability explained by patient-specific factors such as atypical seizure profiles and recording duration rather than demographic attributes. To move toward clinically viable seizure detection systems, future work should explore foundation models designed for raw multivariate time-series data. These architectures could learn richer EEG representations directly from raw signals, potentially improving robustness and cross-subject generalization while reducing the dependence on handcrafted feature extraction.

Acknowledgments. This research has been supported by grant PID2023-146037OB-C21 funded by MICIU/AEI/10.13039/501100011033.

References

1. Explainable foundation models for epileptic seizure detection: Source code repository. https://anonymous.4open.science/r/TFM_EEG_SEIZURE_CLASSIFICATION-8FF1/ (2024), anonymous repository for peer review
2. Abir, S.I., Shoha, S., Hossain, M.M.: Machine learning and deep learning techniques for EEG-based prediction of psychiatric disorders. Journal of Computer Science and Technology Studies **7**(1), 46–63 (1 2025). https://doi.org/10.32996/jcsts.2025.7.1.4
3. Chen, Y., Ren, K., Song, K., Wang, Y., Wang, Y., Li, D., Qiu, L.: Eegformer: Towards transferable and interpretable large-scale eeg foundation model (2024). https://doi.org/10.48550/arXiv.2401.10278, https://arxiv.org/pdf/2401.10278
4. Christ, M., Braun, N., Neuffer, J., Kempa-Liehr, A.W.: Time series feature extraction on basis of scalable hypothesis tests (tsfresh – a python package). Neurocomputing **307**, 72–77 (2018)
5. Darankoum, D., Villalba, M., Allioux, C., Caraballo, B., Dumont, C., Gronlier, E., Roucard, C., Roche, Y., Habermacher, C., Grudinin, S., Volle, J.: From epilepsy seizure classification to detection: A deep learning-based approach for raw eeg signals. Neuroscience Informatics (2026). https://doi.org/10.1016/J.NEURI.2026.100263
6. Detti, P.: Siena scalp EEG database. PhysioNet (2020). https://doi.org/10.13026/5d4a-j060, https://doi.org/10.13026/5d4a-j060, version 1.0.0
7. Detti, P., Vatti, G., de Lara, G.Z.M.: EEG synchronization analysis for seizure prediction: A study on data of noninvasive recordings. Processes **8**(7), 846 (7 2020). https://doi.org/10.3390/pr8070846

8. Dokare, I., Gupta, S.: Shap-driven explainable ai with simulated annealing for optimized seizure detection using multichannel eeg signal. Cognitive Neurodynamics (2025). https://doi.org/10.1007/s11571-025-10269-3, https://pubmed.ncbi.nlm.nih.gov/40485846/
9. Farooq, M.S., Zulfiqar, A., Riaz, S.: Epileptic seizure detection using machine learning: Taxonomy, opportunities, and challenges. Diagnostics **13**(6), 1058 (2023). https://doi.org/10.3390/diagnostics13061058
10. Gao, Y., Xie, W., Luo, Z., Houston, M., Zhang, Y.: Multi-domain feature analysis of MI-EEG signals using tensor train decomposition and projected gradient nonnegative matrix factorization. Neurocomputing **623**, 129410 (3 2025). https://doi.org/10.1016/j.neucom.2025.129410
11. Gurmessa, D.K., Jimma, W.: A comprehensive evaluation of interpretable artificial intelligence for epileptic seizure diagnosis using an electroencephalogram: A systematic review. Digital Health **11** (2025). https://doi.org/10.1177/20552076251325411, https://journals.sagepub.com/doi/10.1177/20552076251325411
12. Hollmann, N., Müller, S., Purucker, L., Krishnakumar, A., Körfer, M., Hoo, S.B., Schirrmeister, R.T., Hutter, F.: Accurate predictions on small data with a tabular foundation model. Nature (2025). https://doi.org/10.1038/s41586-024-08328-6
13. Imrie, F., Davis, R., van der Schaar, M.: Multiple stakeholders drive diverse interpretability requirements for machine learning in healthcare. Nature Machine Intelligence pp. 824–829 (2023). https://doi.org/10.1038/s42256-023-00698-2, https://www.nature.com/articles/s42256-023-00698-2
14. Kuruppu, G., Wagh, N., Kremen, V., Varatharajah, Y.: Eeg foundation models: a critical review of current progress and future directions. Journal of neural engineering (2026). https://doi.org/10.1088/1741-2552/AE4455
15. Liu, D., Chen, Y., Chen, Z., Cui, Z., Wen, Y., An, J., Luo, J., Wu, D.: Eeg foundation models: Progresses, benchmarking, and open problems. arXiv preprint arXiv:2601.17883 (2026). https://doi.org/10.48550/arXiv.2601.17883, https://arxiv.org/abs/2601.17883
16. Mursalin, M., Zhang, Y., Chen, Y., Chawla, N.V.: Automated epileptic seizure detection using improved correlation-based feature selection with random forest classifier. Neurocomputing **241**, 204–214 (6 2017). https://doi.org/10.1016/j.neucom.2017.02.053
17. Obaido, G., Esenogho, E.: Evaluating eeg-based seizure classification using foundation and classical ensemble models. Applied Sciences 2026, Vol. 16, Page 3120 p. 3120 (2026). https://doi.org/10.3390/APP16073120
18. Qu, J., Holzmüller, D., Varoquaux, G., Le Morvan, M.: Tabicl: A tabular foundation model for in-context learning on large data. In: Proceedings of the 42nd International Conference on Machine Learning (ICML) (2025)
19. Rehman, A., Mun, S.: Explainable ai-enhanced ensemble protocol using gradient-boosted models for zero-false-alarm seizure detection from eeg. Sensors p. 863 (2026). https://doi.org/10.3390/s26030863, https://www.mdpi.com/1424-8220/26/3/863
20. Zhong, S., Lian, Y., Yi, H., Jia, K., Zhang, N., Kwan, P., Yang, Y., Tian, X.: Global burden of epilepsy attributable to neonatal disorders in children from 1990 to 2021. iScience (2026). https://doi.org/10.1016/j.isci.2026.115296