Computationally Efficient and Generalizable Machine Learning Algorithms for Seizure Detection from EEG Signals

Zheyun Shou
KTH Royal Institute of Technology
zheyuns@kth.se

Erik Fransén
KTH Royal Institute of Technology
erikf@kth.se

Gonzalo Uribarri
Stockholm University
gonzalo.uribarri@dsv.su.se

Abstract

Automated seizure detection from scalp EEG is critical for epilepsy management, yet existing algorithms often struggle to balance computational efficiency, predictive performance, and generalizability across diverse patient populations. This study investigates the ROCKET framework and complementary state-of-the-art frameworks in time series classification and seizure detection tasks, including Detach-ROCKET, Detach Ensemble, STFT-based feature transform, catch22 features and EEGNet. Models were trained on the TUSZ dataset and rigorously evaluated on both TUSZ and the independent Siena dataset to assess inter-subject and cross-dataset generalizability. The Detach Ensemble model achieved strong performance with median event-wise F1 scores of 0.89 on TUSZ and 0.53 on cross-data evaluation, while maintaining low false positive rates and exceptional computational efficiency. These results demonstrate that ROCKET-based models are competitive in achieving strong predictive performance, computational efficiency, and generalizability promising for practical clinical deployment. The implementation is available at https://github.com/zheyun-shou/eeg_seizure.

1 Introduction

Epilepsy is a chronic neurological disorder characterized by an enduring predisposition to generate recurrent, unprovoked epileptic seizures, which are transient episodes of abnormal, excessive, or synchronous neuronal activity in the brain [10, 11].Seizures can manifest through diverse symptoms ranging from subtle behavioral changes and brief lapses in awareness to dramatic convulsive episodes, depending on the brain regions involved and the pattern of electrical discharge propagation [28]. Beyond the immediate physical risks associated with seizure events, epilepsy imposes substantial burdens on patients through cognitive impairments, psychological comorbidities, and social limitations that diminish quality of life [9, 14, 28]. The unpredictable nature of seizures necessitates continuous monitoring and rapid clinical response, making accurate and timely seizure detection a valuable technique for patient care and the prevention of potentially life-threatening complications [8].

Electroencephalography (EEG) is the most widely used and reliable method for monitoring brain electrical activity and detecting seizures in clinical practice [18]. However, the manual interpretation of EEG recordings is a time-intensive process that requires specialized expertise, making it susceptible to inter-observer variability and potential delays in critical decision-making [21]. The increasing availability of long-term EEG monitoring systems has further amplified the volume of data requiring analysis, creating an urgent need for automated seizure detection algorithms that can assist clinicians in providing timely and accurate diagnoses.

Recent advances in machine learning and time series analysis have opened new possibilities for automated EEG analysis. Traditional approaches rely heavily on handcrafted features and conventional machine learning algorithms [1, 25], while more recent developments have explored deep

39th Conference on Neural Information Processing Systems (NeurIPS 2025) Workshop: Time Series for Health.

learning methodologies [2, 23]. However, many existing approaches face significant challenges in terms of computational efficiency, generalizability across diverse patient populations, and practical deployment in resource-constrained clinical environments [13].

The ROCKET family of algorithms has emerged as a promising paradigm for time series classification, offering state-of-the-art accuracy while maintaining exceptional computational efficiency [5, 6]. These methods transform time series data using randomly generated convolutional kernels, creating discriminative features that can be effectively utilized by simple linear classifiers. Recent extensions, including Detach-ROCKET [31] and Detach Ensemble [24], have further enhanced this approach through feature selection and ensemble methodologies.

2 Methods

2.1 Data Processing

We employed two datasets in this study: the TUH EEG Seizure Corpus(TUSZ) [22] and the Siena Scalp EEG Database [7], comprising recordings from 459 and 14 subjects, respectively. Each EEG recording consists of 19 channels corresponding to the standard 10–20 system scalp electrode montage. To support model training across multiple datasets, we convert each dataset to the Seizure Community Open-Source Research Evaluation (SzCORE) standardized format [4] for data and seizure annotations. Models are trained on the TUSZ dataset, including 1,134 recordings with seizures and 4,585 background recordings. All recordings were downsampled to 128 Hz using FFT-based resampling methods. Before training, we evaluated different window lengths ranging from 2 to 20 seconds using Detach ROCKET models to determine the optimal window length, as it has been shown to notably influence classification performance in previous studies [19, 30]. An 8-second window was selected based on its enhanced test accuracy and reduced memory requirements. Non-overlapping windows (epochs) were then categorized as seizure, interictal, or background, corresponding to segments extracted from seizure events, intervals between annotated seizure events, and background recordings, respectively. Both interictal and background epochs were labeled as non-seizure.

2.2 Models

We evaluate ROCKET-based and complementary state-of-the-art methods for EEG seizure detection, including Detach-ROCKET, Detach Ensemble, STFT-based feature transform, catch22 features and EEGNet. ROCKET (**R**and**O**m **C**onvolutional **KE**rnel **T**ransform) achieves state-of-the-art classification accuracy for time series classification while requiring only a fraction of the computational resources used by most existing methods [5]. It transforms time series using random convolutional kernels and uses the transformed features to train a linear classifier. The method was later reformulated as MINIROCKET, which is up to 75 times faster than ROCKET on large datasets and is almost deterministic, while maintaining nearly the same accuracy [6]. Although ROCKET and MINIROCKET are efficient and computationally lightweight, many of the randomly generated features are redundant or non-informative, increasing the computational burden and may reduce the model's generalizability. Detach-ROCKET addresses this issue by introducing Sequential Feature Detachment (SFD), a method designed to identify and prune the non-essential features from ROCKET-based models [31]. In SFD, the transformed features are ranked according to their contribution to the model's decisions. At each iteration, a fixed proportion (5% in our implementation) of the least informative features is discarded. Given the high dimensionality of multivariate time series such as the 19-channel EEG data, relying on a single set of randomly generated kernels may inadequately capture the complex spatio-temporal patterns and inter-channel relationships critical for seizure detection. To address this limitation, we employed the Detach Ensemble model which involves training N independent Detach-ROCKET models. Each model is pruned using SFD and then assigned a weight based on its performance on the training set [24]. In our implementation, we use an ensemble of $N = 8$ Detach-MINIROCKET models. This number was chosen as a balance between seeking improved performance through ensemble diversity and managing computational resources.

Beyond ROCKET-based models, we include three representative baselines spanning complementary seizure detection paradigms: (i) frequency-domain features via short-time Fourier transform(STFT). Power spectrum density (PSD) on brainwave frequency bands are extracted and fed to a Support Vector Machine (SVM) [29]; (ii) catch22 (Canonical Time-series Characteristics) features with RandomForest, as a state-of-the-art framework in TSC, provides a concise set of 22 highly discriminative and computationally inexpensive time-series features, selected from the larger hctsa

feature set [12, 17] to maximize performance while minimizing redundancy. This focus on feature selection makes it a relevant comparison for our Detach-ROCKET methodology; and (iii) EEGNet, a compact convolutional neural network that uses depthwise and separable convolutions to achieve computational efficiency and generalizability [16].These baselines respectively represent handcrafted features, classical time-series classification methods, and deep learning approaches, offering comprehensive and state-of-the-art comparisons to the Detach-ROCKET methodology. A summary of the configurations and parameters of all models is provided in Appendix A.

2.3 Evaluation of models

The performance of the models are assessed through two complementary evaluation frameworks: epoch-wise and event-wise analysis. An epoch-wise evaluation is conducted based on the 8-second non-overlapping windows. Standard classification metrics including sensitivity, precision, F1-score and False Positive Rate(per 24h) are computed on the test set. This epoch-based assessment provides views of the model's ability to correctly classify individual epochs. Moreover, recognizing that clinical assessment of epilepsy monitoring often focuses on the detection of seizure events rather than isolated time epochs, an event-wise evaluation is performed. During the evaluation procedure, event-based sensitivity, precision, F1-score and FPR are calculated base on the open-source `timescoring` library [3]. We use the same rules to combine the epochs to seizure events as configured by the 2025 Seizure Detection Challenge [27]. Figure 1a illustrates the epoch and event definitions used in our evaluation framework and the Challenge, with detailed rules described in Appendix B. For the specific application of seizure detection, event-wise metrics are considered more clinically meaningful than epoch-wise metrics. Therefore, we evaluate model performance using both epoch-wise and event-wise F1 scores and false positive rates (FPR), with greater emphasis placed on event-wise performance to assess the model's clinical utility and practical applicability.

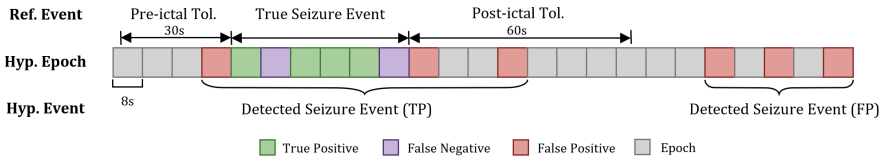
3 Experiments

3.1 Model Performance

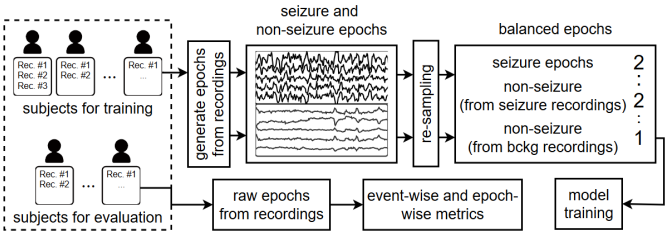
We trained and evaluated our selected models on the TUSZ dataset. We allocated 80% of the subjects for training and 20% for evaluation to ensure that performance metrics reflect the model's ability to generalize across different subjects under the same dataset. Figure 1b illustrates the overall pipeline for data preparation and experiments on the TUSZ dataset. Extracted epochs are balanced before model training. To assess models' generalization capabilities on completely unseen data, we subsequently evaluated performance on the Siena dataset. This independent dataset, which was never used during model development or parameter selection, provides a rigorous test of model performance across different patient populations, recording equipment, and potential variations in EEG signal characteristics.

Table 1 presents performance metrics for all models trained on the TUSZ dataset and evaluated on both datasets. Metrics were computed for each evaluation recording and then averaged across all recordings. We report median values and min-max ranges derived from 5-fold cross-validation. The selected models, particularly Detach-ROCKET, demonstrated strong epoch-wise and event-wise performance on the TUSZ evaluation set, indicating effective generalization across different subjects under similar recording conditions. On the independent Siena dataset, the Detach Ensemble model achieved notably better F1 scores at both levels with a low event-wise FPR compared to other models. When compared to results from the Seizure Detection Challenge 2025 [26], where proposed algorithms were trained on publicly available datasets and evaluated on a private EEG dataset ranked by event-wise F1 score, the Ensemble model's lowest F1 score on the independent Siena dataset (0.48) exceeded the top-performing competition algorithm (F1 = 0.43). Given that the Siena dataset represents a distinct data distribution from the TUSZ training set, this performance indicates the model's competitive generalization capability and robust seizure detection across diverse datasets and recording conditions. The full table is shown in Appendix C.

To examine the model's generalization capabilities across individuals in greater detail, we computed subject-wise performance metrics on the TUSZ evaluation set to investigate the extent of inter-subject variability in prediction performance(Appendix D). Considerable inter-subject variability was observed across our models. This heterogeneity is anticipated and reflects the inherent challenges of EEG-based seizure detection. The observed variation likely stems from multiple factors, including differences in seizure types among test subjects, variations in electrode positioning, and inter-subject differences in scalp thickness, skull conductivity, and electrode-scalp impedance [15, 20].



(a)



(b)

Figure 1: (a) Epoch-wise and event-wise evaluation framework for seizure detection. Detected events are matched against ground truth reference events (Ref. Event) within specified tolerance windows to determine final classification outcomes. And (b) Overview of the experiment pipeline on TUSZ. Training epochs are balanced to create a balanced and diverse training set.

Table 1: Evaluation performance across models, evaluated on TUSZ and Siena datasets at epoch- and event-levels. Values are *median (range)*. RF = RandomForest, SVM = Support Vector Machine.

Model	Level	F1 (TUSZ)	FPR/24h (TUSZ)	F1 (Siena)	FPR/24h (Siena)
D-ROCKET	Epoch	**0.69 (0.58 - 0.70)**	7067 (5562 - 7558)	0.31 (0.30 - 0.38)	5313 (672 - 5982)
	Event	**0.90 (0.87 - 0.91)**	**28 (22 - 33)**	0.39 (0.37 - 0.43)	52 (31 - 68)
D-ROCKET Ens.	Epoch	0.57 (0.56 - 0.62)	7785 (5249 - 11237)	**0.43 (0.39 - 0.47)**	**839 (433 - 4619)**
	Event	0.89 (0.86 - 0.92)	32 (19 - 42)	**0.53 (0.48 - 0.60)**	**24 (19 - 33)**
STFT + SVM	Epoch	0.40 (0.37 - 0.47)	7307 (6234 - 8709)	0.26 (0.26 - 0.29)	2953 (2737 - 3236)
	Event	0.82 (0.76 - 0.89)	47 (24 - 66)	0.31 (0.29 - 0.36)	79 (68 - 81)
catch22 + RF	Epoch	0.53 (0.45 - 0.60)	**6958 (4734 - 11524)**	0.35 (0.34 - 0.36)	1556 (1424 - 1801)
	Event	0.86 (0.82 - 0.90)	30 (18 - 47)	0.46 (0.42 - 0.49)	47 (43 - 55)
EEGNet	Epoch	0.55 (0.46 - 0.56)	10263 (1864 - 21692)	0.28 (0.14 - 0.44)	4839 (781 - 16952)
	Event	0.89 (0.88 - 0.89)	22 (8 - 31)	0.43 (0.20 - 0.61)	67 (20 - 145)

3.2 Computational efficiency

The computational efficiency of the models, particularly their achievable prediction speed, is a critical consideration for real-world deployment of seizure detection systems. We evaluated prediction times using a representative 1,200-second EEG recording on a distributed server(See Appendix E for more details). The Detach-ROCKET model achieved a prediction time of 0.74 seconds, while the ensemble of 8 Detach-MINIROCKET models required 5.27 seconds. Given the superior performance provided by ROCKET-based algorithms, these results highlight the computational advantages of ROCKET-based approaches and demonstrate their strong potential for applications requiring near real-time processing and decision-making.

4 Conclusions

This study aimed to develop and evaluate computationally efficient and generalizable algorithms for seizure detection from scalp EEG recordings. By leveraging ROCKET-based feature transformation algorithms and complementary state-of-the-art TSC and seizure detection algorithms, this work demonstrates that ROCKET-based models effectively balance high predictive performance with computational efficiency and generalizability. Under our experimental framework, the Detach Ensemble model demonstrated strong inter-subject and cross-dataset generalizability through rigorous evaluation on both datasets, achieving median event-wise F1 scores of 0.89 on TUSZ and 0.53 on cross-dataset evaluation while maintaining low false positive rates. The rapid model prediction times confirm the computational efficiency of ROCKET-based approaches, highlighting their potential for implementation in near real-time clinical monitoring systems.

References

[1] U Rajendra Acharya, S Vinitha Sree, G Swapna, Roshan Joy Martis, and Jasjit S Suri. Automated diagnosis of epileptic eeg using entropies. *Entropy*, 15(10):4116–4130, 2013.

[2] Saeed Anwar, Adel Al-Jumaily, and Khurshid Khurshid. Deep learning methods for electroencephalogram (eeg) data analysis: a review. *Multimedia Tools and Applications*, 81(1):1–34, 2022.

[3] Jonathan Dan et al. Library for measuring performance of time series classification. `https://github.com/esl-epfl/timescoring`, 2023. Accessed: 2025-05-04.

[4] Jonathan Dan, Una Pale, Alireza Amirshahi, William Cappelletti, Thorir Mar Ingolfsson, Xiaying Wang, Andrea Cossettini, Adriano Bernini, Luca Benini, Sándor Beniczky, David Atienza, and Philippe Ryvlin. Szcore: Seizure community open-source research evaluation framework for the validation of electroencephalography-based automated seizure detection algorithms. *Epilepsia*, n/a(n/a).

[5] Angus Dempster, François Petitjean, and Geoffrey I. Webb. Rocket: exceptionally fast and accurate time series classification using random convolutional kernels. *Data Mining and Knowledge Discovery*, 34(5):1454–1495, July 2020.

[6] Angus Dempster, Daniel F. Schmidt, and Geoffrey I. Webb. Minirocket: A very fast (almost) deterministic transform for time series classification. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery amp; Data Mining*, KDD '21, page 248–257. ACM, August 2021.

[7] P. Detti. Siena scalp eeg database (version 1.0.0). `https://doi.org/10.13026/5d4a-j060`, 2020. PhysioNet.

[8] Orrin Devinsky, Annamaria Vezzani, Terence J O'Brien, Marc Dichter, Emilio Perucca, Wolfgang Löscher, and Ingrid E Scheffer. Epilepsy. *Nature reviews Disease primers*, 4(1):1–23, 2018.

[9] Kirsten M Fiest, Jonathan Dykeman, Scott B Patten, Samuel Wiebe, Gilaad G Kaplan, Colleen J Maxwell, Andrew G M Bulloch, and Nathalie Jetté. Depression in epilepsy: a systematic review and meta-analysis. *Neurology*, 80(6):590–599, 2013.

[10] Robert S Fisher, Carlos Acevedo, Alexis Arzimanoglou, Alicia Bogacz, J Helen Cross, Christian E Elger, Jerome Engel Jr, Lars Forsgren, Jacqueline A French, Mike Glynn, et al. Ilae official report: a practical clinical definition of epilepsy. *Epilepsia*, 55(4):475–482, 2014.

[11] Robert S Fisher, J Helen Cross, Jacqueline A French, Norimichi Higurashi, Edouard Hirsch, Floor E Jansen, Lieven Lagae, Solomon L Moshé, Jukka Peltola, Eliane Roulet Perez, et al. Operational classification of seizure types by the international league against epilepsy: position paper of the ilae commission for classification and terminology. *Epilepsia*, 58(4):522–530, 2017.

[12] Ben D Fulcher, Max A Little, and Nick S Jones. Highly comparative time-series analysis: the empirical structure of time series and their methods. *Journal of the Royal Society Interface*, 10(83):20130048, 2013.

[13] Kenneth Han, Chris Liu, and Daniel Friedman. Artificial intelligence/machine learning for epilepsy and seizure diagnosis. *Epilepsy Behavior*, 155:109736, 2024.

[14] Barbara C Jobst and Catherine A Schevon. Cognitive dysfunction in adults with epilepsy. *CNS spectrums*, 15(S8):12–19, 2010.

[15] Y. Lai, W. van Drongelen, L. Ding, K. E. Hecox, V. L. Towle, D. M. Frim, and B. He. Estimation of in vivo human brain-to-skull conductivity ratio from simultaneous extra- and intra-cranial electrical potential recordings. *Clinical neurophysiology : official journal of the International Federation of Clinical Neurophysiology*, 116(2):456–465, Feb 2005.

[16] Vernon J Lawhern, Amelia J Solon, Nicholas R Waytowich, Stephen M Gordon, Chou P Hung, and Brent J Lance. Eegnet: a compact convolutional neural network for eeg-based brain–computer interfaces. *Journal of neural engineering*, 15(5):056013, 2018.

[17] Carl H Lubba, Sarab S Sethi, Philip Knaute, Simon R Schultz, Ben D Fulcher, and Nick S Jones. catch22: Canonical time-series characteristics. *Data Mining and Knowledge Discovery*, 33:1821–1852, 2019.

[18] Ernst Niedermeyer and Fernando Lopes Da Silva. *Electroencephalography: Basic principles, clinical applications, and related fields*. Lippincott Williams & Wilkins, 5th edition, 2005.

[19] Delin Ouyang, Yufei Yuan, Guofa Li, and Zizheng Guo. The effect of time window length on eeg-based emotion recognition. *Sensors*, 22(13):4939, 2022.

[20] Roberto D. Pascual-Marqui, Dietrich Lehmann, Thomas Koenig, Kieko Kochi, Marco C.G. Merlo, Daniel Hell, and Martha Koukkou. Low resolution brain electromagnetic tomography (loreta) functional imaging in acute, neuroleptic-naive, first-episode, productive schizophrenia. *Psychiatry Research: Neuroimaging*, 90(3):169–179, 1999.

[21] S Roy, I Kiral-Kornek, and S Harrer. A review on epileptic seizure detection using machine learning classifiers. *IEEE Reviews in Biomedical Engineering*, 14:249–264, 2021.

[22] Vikram Shah, Ethan von Weltin, Sarah Lopez, John McHugh, Luis Veloso, Mohammad Golmohammadi, Ibrahim Obeid, and Joseph Picone. The temple university hospital seizure detection corpus. *Frontiers in Neuroinformatics*, 12:83, 2018.

[23] Arman Shoeibi, Mahboobeh Khodatars, Nima Ghassemi, Marjaneh Jafari, Parisa Moridian, Roohallah Alizadehsani, Maryam Panahiazar, Fahimeh Khozeimeh, Atefeh Zare, Hossein Hosseini-Nejad, et al. Epileptic seizures detection using deep learning techniques: A review. *International Journal of Environmental Research and Public Health*, 18(11):5780, 2021.

[24] Adrià Solana, Erik Fransén, and Gonzalo Uribarri. Classification of raw meg/eeg data with detach-rocket ensemble: An improved rocket algorithm for multivariate time series analysis, 2024.

[25] Abdulhamit Subasi and M Ilk Gursoy. Practical guide for implementing machine learning for eeg signal processing. *IEEE Sensors Journal*, 19(21):9649–9667, 2019.

[26] SzCORE Organizers. Algorithm scoring results - seizure detection challenge (2025). `https://szcore.org/challenges/challenge_2025/results/`, 2024. Accessed on: May 17, 2025.

[27] SzCORE Organizers. Seizure detection challenge (2025). `https://szcore.org/challenges/challenge_2025/`, 2024. Accessed on: May 17, 2025.

[28] Roland D Thijs, Rainer Surges, Terence J O'Brien, and Josemir W Sander. *Epilepsy in adults*, volume 393. Elsevier, 2019.

[29] Alexandros T. Tzallas, Markos G. Tsipouras, and Dimitrios I. Fotiadis. Epileptic seizure detection in eegs using time–frequency analysis. *IEEE Transactions on Information Technology in Biomedicine*, 13(5):703–710, 2009.

[30] Katerina D Tzimourta, Nikolaos Giannakeas, Alexandros T Tzallas, Loukas G Astrakas, Theodora Afrantou, Panagiotis Ioannidis, Nikolaos Grigoriadis, Pantelis Angelidis, Dimitrios G Tsalikakis, and Markos G Tsipouras. Eeg window length evaluation for the detection of alzheimer's disease over different brain regions. *Brain sciences*, 9(4):81, 2019.

[31] Gonzalo Uribarri, Federico Barone, Alessio Ansuini, and Erik Fransén. Detach-rocket: Sequential feature selection for time series classification with random convolutional kernels, 2024.

Appendix A Model Architectures and Parameters

Table 2: Summary of the main architectural parameters used for each model tested in this work. Default values refer to those in the respective published implementations.

Model	Parameters / Architecture
D-ROCKET	10,000 random convolutional kernels (default). Sequential Feature Detachment (SFD) with 5% pruning per iteration (default).
D-ROCKET Ens.	Ensemble of $N = 8$ Detach-MINIROCKET models.
STFT + SVM	STFT features computed over standard EEG frequency bands: δ (0.5–4 Hz), θ (4–8 Hz), α (8–13 Hz), β (13–20 Hz, 20-30 Hz), γ (30–40 Hz, 40-60 Hz). SVM with RBF kernel using default parameters from scikit-learn.
catch22 + RF	22 canonical time-series features (catch22). Random Forest with 100 trees, max depth = 16.
EEGNet	Default implementation of the official repository github.com/vlawhern/arl-eegmodels. Hyperparameters (dropout rate, kernel length, F1, D, F2, and epochs) are explored to obtain optimized results for fair comparison: Chans = 19, Samples = 1024, dropoutRate = 0.01, kernLength = 64, F1 = 8, D = 2, F2 = 16, dropoutType = 'Dropout', epoch = 100.

Appendix B Rules for Combining Seizure Epochs into Events

The following rules were applied in this study to merge individual seizure epochs into seizure events:

Minimum Overlap: Any temporal overlap, however brief, between a reference event and a hypothesis event is sufficient to consider it a potential match.

Pre-ictal Tolerance: A hypothesis event starting up to 30 seconds before the onset of a reference event can still be considered a detection of that event.

Post-ictal Tolerance: A hypothesis event ending up to 60 seconds after the end of a reference event can still be considered part of the detection of that event.

Minimum Duration: Reference or hypothesis events separated by less than 90 seconds are merged into a single, longer event before scoring. This duration corresponds to the sum of the pre- and post-ictal tolerances, preventing closely spaced detections from being penalized multiple times.

Appendix C Detailed performance comparison

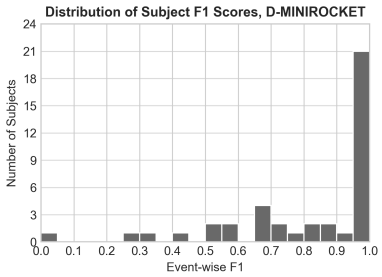
Table 3: Performance comparison of different models for seizure detection on TUSZ dataset. The median values and the range of results from the 5-fold cross-validation are reported for the main metrics.

Model	Level	Sens.	Prec.	F1	FPR/24h
D-ROCKET	Epoch	**0.74 (0.61–0.76)**	**0.73 (0.70–0.73)**	**0.69** (0.58–0.70)	7067 (5562–7558)
	Event	0.94 (0.91–0.96)	0.89 (0.87–0.89)	**0.90 (0.87–0.91)**	28 (22–33)
D-ROCKET Ens.	Epoch	0.64 (0.56–0.69)	0.68 (0.65–0.76)	0.57 (0.56–0.62)	7785 (5249–11237)
	Event	**0.95 (0.90–0.96)**	0.88 (0.83–0.91)	0.89 (0.86–0.92)	32 (19–42)
STFT + SVM	Epoch	0.41 (0.36–0.48)	0.56 (0.54–0.66)	0.40 (0.37–0.47)	7307 (6234–8709)
	Event	0.91 (0.83–0.93)	0.80 (0.73–0.89)	0.82 (0.76–0.89)	47 (24–66)
catch22 + RF	Epoch	0.53 (0.49–0.63)	0.69 (0.65–0.76)	0.53 (0.45–0.60)	**6958 (4734–11524)**
	Event	0.92 (0.88–0.95)	0.88 (0.80–0.91)	0.86 (0.82–0.90)	30 (18–47)
EEGNet	Epoch	0.62 (0.39–0.77)	0.72 (0.52–0.89)	0.55 (0.46–0.56)	10262 (1864–21692)
	Event	0.92 (0.85–0.97)	**0.90 (0.88–0.95)**	0.89 (0.88–0.89)	**22 (8–31)**

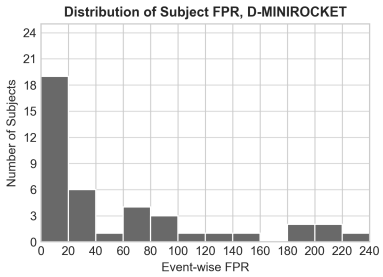
Table 4: Performance comparison of different models for seizure detection on Siena dataset. The median values and the range of results from the 5-fold cross-validation are reported for the main metrics.

Model	Level	Sens.	Prec.	F1	FPR/24h
D-ROCKET	Epoch	0.50 (0.37–0.53)	0.33 (0.32–0.41)	0.31 (0.30–0.38)	5313 (672–5982)
	Event	0.89 (0.74–0.92)	0.32 (0.30–0.35)	0.39 (0.37–0.43)	52 (31–68)
D-ROCKET Ens.	Epoch	0.46 (0.42–0.50)	**0.54 (0.45–0.58)**	**0.43 (0.39–0.47)**	**839 (433–4619)**
	Event	0.84 (0.77–0.88)	**0.47 (0.39–0.54)**	**0.53 (0.48–0.60)**	**24 (19–33)**
STFT + SVM	Epoch	0.39 (0.39–0.42)	0.27 (0.24–0.31)	0.26 (0.26–0.29)	2953 (2737–3236)
	Event	0.84 (0.83–0.86)	0.22 (0.21–0.28)	0.31 (0.29–0.36)	79 (68–81)
catch22 + RF	Epoch	0.45 (0.44–0.46)	0.41 (0.36–0.42)	0.35 (0.34–0.36)	1556 (1424–1801)
	Event	0.89 (0.89–0.92)	0.38 (0.34–0.42)	0.46 (0.42–0.49)	47 (43–55)
EEGNet	Epoch	**0.57 (0.24–0.79)**	0.40 (0.10–0.59)	0.28 (0.14–0.44)	4839 (781–16952)
	Event	**0.91 (0.79–0.99)**	0.37 (0.12–0.59)	0.43 (0.20–0.61)	66 (20–145)

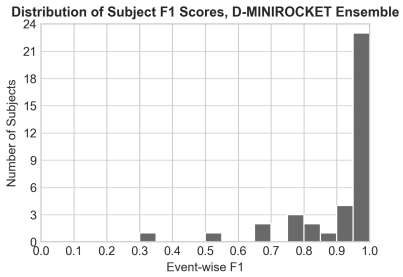
Appendix D Subject Variation Analysis



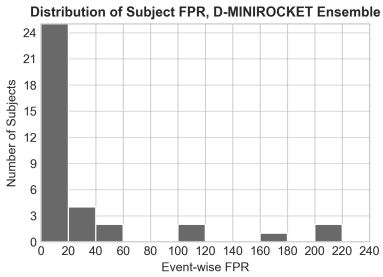
(a) Subject-wise event F1 score distribution for the Detach-MINIROCKET on the TUSZ evaluation set.



(b) Subject-wise event FPR distribution for the Detach-MINIROCKET on the TUSZ evaluation set.



(c) Subject-wise event F1 score distribution for the Detach Ensemble on the TUSZ evaluation set.



(d) Subject-wise event FPR distribution for the Detach Ensemble model on the TUSZ evaluation set.

Figure 2: Distribution of subject-wise performance metrics on the TUSZ test set for different model configurations. Subjects who contributed exclusively background recordings to the test set (i.e., had no annotated seizure events) are excluded. For these subjects, True Positives (TP) and False Negatives (FN) are inherently zero, so the F1 score cannot be calculated.

Appendix E Computational Performance and Hardware Configuration

All models were trained and evaluated on a distributed NVIDIA DGX-H100 machine using an allocated pod with 64GB RAM and 20GB GPU memory. ROCKET-based models and catch22 utilized both CPU and GPU resources, STFT ran exclusively on CPU, and EEGNet ran exclusively on GPU.

Table 5: Model prediction time (seconds) on a 1200s EEG recording.

D-MINIROCKET	**D-MINIROCKET Ens.**	**STFT**	**Catch22**	**EEGNet**
0.80	5.27	4.57	14.20	0.78