Degree Project in Technology

Second cycle, 30 credits

Efficient and Generalizable Seizure Detection Algorithm

A subtitle in the language of the thesis

ZHEYUN SHOU

Contents

Chapter 2

Background

2.1 Seizure and Epilepsy

A seizure represents a fundamental disruption of normal brain function, characterized by a transient surge of abnormal, excessive, or synchronous neuronal activity within the brain [1, 2]. These paroxysmal events can manifest through a wide spectrum of signs and symptoms, the nature of which is intrinsically linked to the specific cerebral networks involved in the ictal discharge and its subsequent propagation [3]. While an isolated seizure may be provoked by acute systemic insults or transient neurological stressors, the term epilepsy designates a more profound neurological condition: an enduring predisposition to generate unprovoked epileptic seizures, along with the associated neurobiological, cognitive, psychological, and social ramifications [4]. Clinically, epilepsy is typically diagnosed after at least one unprovoked seizure, especially when factors suggest a high likelihood of recurrence.

The global impact of seizures and epilepsy is considerable. Affecting an estimated 50 million individuals worldwide, epilepsy stands as one of the most prevalent serious neurological disorders [5]. Beyond the immediate risk of physical injury during an ictal event, seizures can precipitate or exacerbate cognitive difficulties, particularly in domains of memory and attention, and are frequently associated with psychological comorbidities such as anxiety and depression [6, 7]. The societal impact is also profound, with individuals often confronting stigma, discrimination, and limitations in education, employment, and personal autonomy [8]. The consequences of recurrent seizures permeate virtually every aspect of an individual's life.

The diagnostic process for seizures heavily relies on meticulous clinical

history-taking, including detailed eyewitness accounts, as the ictal events themselves are often brief and unpredictable, making direct observation by clinicians rare [9]. This dependence on subjective reporting, while indispensable, introduces potential inaccuracies and challenges in distinguishing seizures from other paroxysmal events [10]. Accurate seizure detection is not merely an academic exercise; it is critical for guiding appropriate therapeutic interventions, predicting long-term outcomes, and informing genetic counseling [11].

2.2 Electroencephalography for Brain Activity Monitoring

Electroencephalography (EEG) is a cornerstone neurophysiological technique used to record the electrical activity generated by the brain. It provides a non-invasive, direct measure of brain function by detecting voltage fluctuations resulting from ionic current flows within the neurons of the cerebral cortex [12, 13]. Specifically, EEG signals primarily reflect the summed postsynaptic potentials (both excitatory and inhibitory) of large populations of synchronously active pyramidal neurons oriented radially to the scalp [13, 14]. Due to its excellent temporal resolution, typically in the millisecond range, EEG is uniquely suited for capturing the rapidly changing dynamics of brain activity, making it an invaluable tool in both clinical neurology and neuroscience research, particularly for the assessment of conditions characterized by abnormal electrical discharges, such as epilepsy [15].

The acquisition of EEG data involves placing electrodes on the scalp, typically made of conductive materials like silver/silver-chloride (Ag/AgCl). These electrodes detect the minute electrical potentials (on the order of microvolts, μV) generated by the brain, which are then significantly amplified by differential amplifiers to make them suitable for digitization and subsequent analysis [16]. To ensure reproducibility and comparability of EEG recordings across different laboratories and individuals, standardized electrode placement systems are employed. The most widely adopted is the International 10-20 system and its extensions [17, 18]. This system positions electrodes at locations that are 10% or 20% of the total front-to-back or right-to-left distance of the skull, ensuring proportional spacing relative to cranial landmarks (nasion, inion, and preauricular points). Each electrode is labeled with a letter indicating the underlying brain lobe (e.g., F for frontal, P for parietal, T for temporal, O for occipital) and a number or another letter to denote its specific

position [17].

The resulting EEG waveforms are complex and are traditionally analyzed in terms of their frequency, amplitude, morphology, and topography. Clinically relevant information is often contained within specific frequency bands, such as delta (0.5-4 Hz), theta (4-8 Hz), alpha (8-13 Hz), beta (13-30 Hz), and gamma (>30 Hz), each associated with different brain states or cognitive processes [13]. In the context of epilepsy, EEG is crucial for identifying interictal epileptiform discharges (IEDs) and for characterizing the electrographic signature of seizures themselves (ictal patterns) [19]. These patterns provide critical information for diagnosing epilepsy, classifying seizure types, localizing the seizure onset zone, and guiding treatment decisions.

However, the utility of EEG is often challenged by the inherent low signal-to-noise ratio (SNR) and its susceptibility to various artifacts [16, 20]. EEG signals are frequently contaminated by physiological artifacts originating from non-cerebral sources, such as eye movements and blinks (electrooculogram, EOG), muscle activity (electromyogram, EMG) particularly from scalp and facial muscles, and cardiac activity (electrocardiogram, ECG) [20, 21]. Non-physiological artifacts also pose significant problems, including 50/60 Hz power line interference, electrode impedance issues (e.g., electrode pop or poor contact), and movement artifacts from patient motion or cable sway [21]. These artifacts can mimic or obscure true epileptiform activity, leading to potential misinterpretations if not properly addressed. Consequently, rigorous data preprocessing, encompassing techniques such as filtering, artifact detection, and artifact removal or suppression, is an indispensable yet often complex step before any meaningful analysis or automated interpretation of EEG data can be performed [21, 22]. The challenge of reliably distinguishing pathological neural signals from this pervasive noise, especially in long-term recordings or in ambulatory settings, underscores the need for robust and sophisticated signal processing and machine learning algorithms, forming a critical motivation for the research presented in this thesis.

2.3 Time-series classification

Time Series Classification (TSC) is a specialized area within machine learning concerned with assigning predefined categorical labels to unlabeled time series data [23]. A time series itself is a sequence of data points indexed in time order, commonly encountered in diverse domains such as finance, healthcare, environmental science, and industrial processes [24]. The primary

objective of TSC is to build a model that can learn discriminative patterns or features from a collection of labeled time series, enabling it to accurately predict the class of new, unseen time series instances, and providing a powerful framework for automated detection tasks [23, 25].

Early TSC methods often relied on distance-based measures, with Dynamic Time Warping (DTW) being a prominent example, which calculates similarity between two temporal sequences that may vary in time or speed [26]. Other approaches focused on extracting statistical or structural features from the time series (e.g., mean, variance, spectral properties, shapelets) and then applying standard static classification algorithms [25, 27]. While effective in certain contexts, these methods can sometimes struggle with the high dimensionality, inherent noise, and complex temporal dependencies often present in real-world time series data.

More recently, deep learning architectures have demonstrated state-of-the-art performance on many TSC benchmarks [28]. Prominent among these are variants of Convolutional Neural Networks (CNNs) and Residual Networks (ResNet), adapted for time series, and have achieved remarkable success by effectively capturing local and global patterns [29]. Methods like InceptionTime also utilizes modules with multiple convolutional filter sizes to capture features at different temporal scales [30]. Recurrent Neural Networks (RNNs), particularly Long Short-Term Memory (LSTM) units and Gated Recurrent Units (GRUs), are inherently suited for modeling temporal dependencies and have also been widely applied, though sometimes at a higher computational cost [31]. Furthermore, Transformer-based models, leveraging self-attention mechanisms, have shown significant promise in capturing long-range dependencies and complex relationships within time series data [32]. Other approaches also includes ensemble methods, which combine several heterogeneous classifiers into a large meta-ensemble [33, 34, 35]. However, training these models can be computationally intensive, require large datasets, which poses limits in a sparse data scenario.

Alongside the advancements in deep learning, a distinct and highly competitive paradigm has gained prominence, focusing on transform-based feature engineering coupled with efficient classification. This approach decouples the often complex task of feature extraction from the subsequent classification step, allowing for specialized, powerful feature generation techniques to be paired with simpler, faster classifiers. Early and comprehensive exemplars of this philosophy include toolkits like hctsa (Highly Comparative Time-Series Analysis) [36] and catch22 [37]. hctsa facilitates the extraction of thousands of diverse time-series features derived

from a wide array of scientific literature, enabling extensive data-driven exploration and the potential discovery of novel discriminative patterns [36]. In contrast, catch22 offers a more concise, curated set of 22 canonical time-series characteristics selected for their broad applicability and high discriminative power in general-purpose TSC tasks, aiming for robust performance with minimal redundancy [37]. Both these methodologies transform raw time series into a rich feature space, thereby empowering relatively simple classifiers to perform effectively.

Another highly influential branch within this "transform-then-classify" paradigm emerged with ROCKET (RandOm Convolutional KErnel Transform) [38]. ROCKET demonstrated that convolving time series with a large number of randomly generated, diverse convolutional kernels and extracting simple summary statistics (typically the maximum value and proportion of positive values) could yield highly discriminative features. These features then enable robust linear classifiers, such as ridge regression, to achieve state-of-the-art accuracy with exceptional speed, bypassing the need for complex backpropagation and extensive hyperparameter tuning associated with many deep learning models [38, 39]. The success of ROCKET spurred further innovations: MiniROCKET significantly reduced computational overhead by optimizing kernel parameters and focusing primarily on the proportion of positive values feature [40], while MultiROCKET enhanced performance by incorporating a richer set of summary statistics from the convolved outputs and an improved kernel generation strategy [41]. Building upon this lineage, Detach-ROCKET further reinforces the modular nature of the approach by separating, or "detaching" the redundant random kernels from the classifier training stage [42]. This design preserves computational efficiency and enhances the flexibility of feature generation, thereby readily supporting ensemble methods such as Detach-Ensemble [43], where diverse feature sets can be strategically combined. The collective strength of these methods lies in their computational efficiency, reduced reliance on massive datasets, and potential for greater interpretability, making them particularly well-suited for challenging biomedical signal analysis tasks.

Chapter 3

Method

3.1 Data Processing

3.1.1 Data

In this project, two datasets are employed for model training and evaluation. The first dataset is the TUH EEG Seizure Corpus [44] (TUSZ), which comprises recordings from a total of 459 subjects. Within this corpus, a subset of the EEG recordings contains annotated seizure events, while the remainder consists of background (BCKG) recordings included to balance the dataset and to better assess the system's false alarm performance. Here BCKG recordings are defined as those that do not exhibit any of the following patterns: spike and/or sharp waves, periodic lateralized epileptiform discharges, generalized periodic epileptiform discharges, eye movements, or artifacts. Each EEG recording consists of 19 channels corresponding to the standard 10–20 system scalp electrode montage (see Figure 3.1), with a sampling rate of 256 Hz. See the publication of Obeid et al. [45] for further details on data collection.

Another dataset we used is the Siena Scalp EEG Database [46]. This database contains EEG recordings from 14 patients, collected at the Unit of Neurology and Neurophysiology at the University of Siena. The participants include 9 males (ages 25–71) and 5 females (ages 20–58). Each subject was monitored using Video-EEG at a sampling rate of 256 Hz. The recordings include 19 EEG channels, based on the standard 10–20 system scalp electrode montage (see Figure 3.1). See the publication of Detti et al. [47] for further details on data collection.

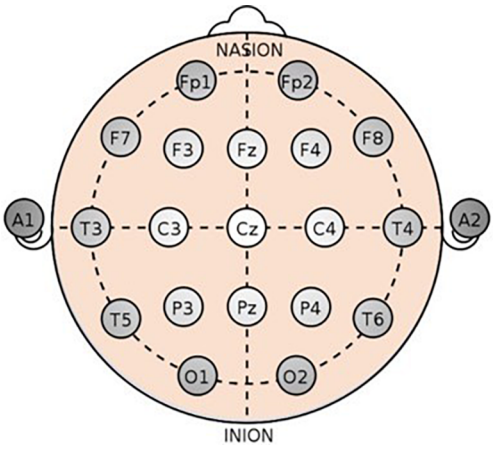

Figure 3.1: Electrode locations for EEG signals. From Shah, V., von Weltin, E., Lopez. S., McHugh, J., Veloso, L., Golmohammadi, M., Obeid, I., and Picone, J. (2018). The Temple University Hospital Seizure Detection Corpus. Frontiers in Neuroinformatics. 12:83. doi: 10.3389/fninf.2018.00083

3.1.2 Data preprocessing

To support model training across multiple datasets, we convert each dataset to the Seizure Community Open-Source Research Evaluation (SzCORE) standardized format [48] for data and seizure annotations. This format is compatible with the Brain Imaging Data Structure (BIDS).

The model is trained on the large Temple University Seizure Corpus (TUSZ) dataset, which includes 1,134 recordings with seizures and 4,585 background recordings. All recordings are downsampled to 128 Hz. Each recording is segmented into non-overlapping 10-second epochs. These epochs are then categorized as seizure, interictal, or background (bckg). Seizure epochs are sampled from periods annotated as seizures. Interictal epochs are sampled from intervals between annotated seizure events. Background epochs are sampled exclusively from recordings labeled as BCKG. Seizure epochs are assigned the label 1 (seizure), while interictal and background epochs are assigned the label 0 (non-seizure).

3.2 Model

3.2.1 ROCKET and MINIROCKET

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 [38]. 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 (**MINI**mally **R**and**O**m **C**onvolutional **KE**rnel **T**ransform), which is up to 75 times faster than ROCKET on large datasets and is almost deterministic, while maintaining nearly the same accuracy [40]. MINIROCKET is significantly faster than other methods with similar accuracy and provides significantly better accuracy than methods with comparable computational efficiency.

3.2.2 Detach-ROCKET

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 [42]. In SFD, the transformed features are ranked according to their contribution to the model's decisions. At each iteration, a fixed proportion of the least informative features is discarded. Let $\mathbb{F}$ denote the complete set of features generated by ROCKET's random convolutional kernels. The subset of currently active features is represented by $\mathbb{S} \subseteq \mathbb{F}$, which is initially set to $\mathbb{F}$.

At each step t, a ridge classifier is trained on the active feature set $\mathbb{S}_t$ by solving the following optimization problem:

$$\hat{\theta}_t^{\text{ridge}} = \underset{\theta}{\operatorname{argmin}} \left\{ \sum_{i=1}^{N} \left(y_i - \theta_0 - \sum_{k \in \mathbb{S}_t} x_{ik}\theta_k \right)^2 + \lambda \sum_{k \in \mathbb{S}_t} \theta_k^2 \right\} \tag{3.1}$$

The training process produces a set of optimal coefficients $\hat{\theta}_t^{\text{ridge}} = \{\hat{\theta}_k\}$, where each coefficient is proportional to the contribution of a corresponding feature on the classifier's decision. The features are then ranked according

to the absolute values of their coefficients. Let p denote the proportion of features to be removed. At each step, the lowest $100 \cdot p\%$ of ranked features discarded, and the remaining $100 \cdot (1-p)\%$ are retained to form the updated feature set $\mathbb{S}_{t+1}$. Since p controls the trade-off between computational cost and model accuracy, it is set to 0.05 to ensure a conservative yet computationally affordable pruning procedure, meaning that 5% of the features are removed at each step.

To determine the optimal number of features to finally retain in our dataset, we solve the following optimization problem:

$$Q_c \;=\; \underset{q}{argmax}\;\; f_c(q) = \;\underset{q}{argmax}\;\; \{\alpha(q) + c \cdot q\}\,. \tag{3.2}$$

In this equation, Q_c represents the accuracy curve obtained by evaluating the model at each pruning step, q represents the proportion of pruned features, and $\alpha(q)$ the accuracy of the pruned model on the validation set. The hyperparameter c is the weighting factor between accuracy and data size, in which a smaller c value favors accuracy and larger c value favors data size. We set $c = 0.1$, as this achieves the optimal performance according to Fig.4 in [42].

3.2.3 Detach Ensemble

Given the high dimensionality of multivariate time series such as the 19-channel EEG data in this project, 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 and improve model generalization and robustness, an ensemble approach based on Detach-ROCKET is employed. The Detach Ensemble involves training N independent Detach-ROCKET models. For each model, a subset of the training data is used to determine the optimal pruning size, and the model is then pruned using SFD. Each pruned model is then assigned a weight based on its performance on the training set [43].

For classifying a given input instance, predictions from each individual model are aggregated via a weighted average, using the predetermined model weights, to yield a final ensemble probability score. This probability is then thresholded (commonly at 0.5) to produce the definitive classification label (seizure or non-seizure).

In our implementation, we use an ensemble of $N = 10$ Detach-MiniROCKET models. This number was chosen as a balance between seeking improved performance through ensemble diversity and managing

computational resources, as training significantly more models becomes resource-intensive. Each of the 10 model is trained independently using the procedure described in Section 3.2.2, including the application of SFD with $p = 0.05$ and the feature selection criterion $c = 0.1$.

3.3 Evaluation of Models

The performance of the trained Detach-ROCKET and Detach Ensemble model are assessed through two complementary evaluation frameworks: epoch-wise and event-wise analysis.

Firstly, an epoch-wise evaluation is conducted based on the 10-second non-overlapping segments described in Section 3.1.2. Standard classification metrics including accuracy, sensitivity, precision, and the F1-score are computed on the test set. The confusion matrix is also examined to understand the distribution of classification errors. This epoch-based assessment provides views of the model's ability to correctly classify individual epochs.

However, recognizing that clinical assessment of epilepsy monitoring often focuses on the detection of seizure episodes (events) rather than isolated time epochs, an event-wise evaluation is performed. During the evaluation procedure, event-based sensitivity, precision, and F1-score are calculated base on the open-source `timescoring` library [49].

3.3.1 Evaluation metrics

To evaluate the performance both epoch-wise and event-wise, we first define and explain the True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN) in this study: True Positive (TP): An instance correctly identified as belonging to the positive class (seizure); True Negative (TN): An instance correctly identified as belonging to the negative class (non-seizure);False Positive (FP): An instance incorrectly identified as belonging to the positive class when it belongs to the negative class (a non-seizure instance classified as seizure);False Negative (FN): An instance incorrectly identified as belonging to the negative class when it belongs to the positive class (a seizure instance classified as non-seizure).

For epoch-wise evaluation, a TP is a correctly classified seizure epoch, an FP is a non-seizure epoch classified as seizure, etc. For event-wise evaluation, the following parameters were configured to combine the epochs to seizure events, as suggested by the 2025 Seizure Detection Challenge:

- Minimum Overlap: Any temporal overlap, however brief, between a reference event and a hypothesis event is sufficient to consider it a potential match. This setting maximizes sensitivity to detecting any part of a seizure.
- 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.

In this configuration, a reference event refers to a labeled seizure event, a hypothesis event refers to a model predicted event. Based on these criteria, an event-based TP occurs when a reference seizure event is correctly matched with one or more hypothesis seizure events according to the overlap and tolerance rules. An event-based FP corresponds to a hypothesis seizure event that does not match any reference event. An event-based FN represents a reference seizure event that is not matched by any hypothesis event. For the specific application of seizure detection, event-wise metrics are considered more clinically meaningful than epoch-wise metrics. Consequently, while both will be reported, event-wise performance will be the primary focus for evaluating the model's practical utility.

Based on these setting, we define the accuracy, sensitivity, precision, f1-score and false alarm rate as following:

- Accuracy: the proportion of total instances that were correctly classified.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{3.3}$$

- Sensitivity: the proportion of true positive instances that were correctly identified by the model.

$$\text{Sensitivity} = \frac{TP}{TP + FN} \tag{3.4}$$

- Precision: the proportion of instances predicted as positive that were actually positive.

$$\text{Precision} = \frac{TP}{TP + FP} \tag{3.5}$$

- F1 Score: the harmonic mean of Precision and Sensitivity, providing a single metric that balances both concerns.

$$\text{F1 Score} = 2 \times \frac{\text{Precision} \times \text{Sensitivity}}{\text{Precision} + \text{Sensitivity}} = \frac{2 \times TP}{2 \times TP + FP + FN} \tag{3.6}$$

- False positive rate (FPR): the number of false positives per 24 hours. A low FPR is crucial for clinical usability, as frequent false alarms can reduce clinician trust in the system.

$$\text{FPR} = \frac{FP}{FP + TN} \tag{3.7}$$

3.4 Channel Relevance Estimation

To estimate the relevance of individual EEG channels within our Detach-ROCKET Ensemble, we adopt the methodology proposed by Solana et al. [43]. This process first assesses channel relevance on each Detach-MINIROCKET model:

1. **Kernel Selection:** Sequential Feature Detachment (SFD) identifies and selects the most relevant kernels within the MiniROCKET model.
2. **Channel Retrieval:** For each kernel selected by SFD, the specific EEG channels it processed are retrieved.
3. **Importance Weighting:** Each retrieved channel receives an importance score proportional to its kernel's weight (θ_i). It is then divided by the number of channels in that same kernel.
4. **Relevance Aggregation:** The weighted importance scores for all channels are summed across the selected kernels. These sums are then normalized to create a relative channel relevance histogram.

why divided by the number of channels?

The final ensemble relevance for each channel is then determined by taking the median of relevancies from base models and normalizing across all channels.

References

[1] R. S. Fisher, C. Acevedo, A. Arzimanoglou, A. Bogacz, J. H. Cross, C. E. Elger, J. Engel Jr, L. Forsgren, J. A. French, M. Glynn *et al.*, "Ilae official report: a practical clinical definition of epilepsy," *Epilepsia*, vol. 55, no. 4, pp. 475–482, 2014. [Page 3.]

[2] R. S. Fisher, J. H. Cross, J. A. French, N. Higurashi, E. Hirsch, F. E. Jansen, L. Lagae, S. L. Moshé, J. Peltola, E. 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*, vol. 58, no. 4, pp. 522–530, 2017. [Page 3.]

[3] R. D. Thijs, R. Surges, T. J. O'Brien, and J. W. Sander, *Epilepsy in adults*. Elsevier, 2019, vol. 393, no. 10172. [Page 3.]

[4] A. T. Berg, S. F. Berkovic, M. J. Brodie, J. Buchhalter, J. H. Cross, W. van Emde Boas, J. Engel, J. French, T. A. Glauser, G. W. Mathern *et al.*, "Revised terminology and concepts for organization of seizures and epilepsies: report of the ilae commission on classification and terminology, 2005–2009," *Epilepsia*, vol. 51, no. 4, pp. 676–685, 2010. [Page 3.]

[5] World Health Organization, "Epilepsy," February 2023, accessed on [Insert Date of Access Here]. [Page 3.]

[6] B. C. Jobst and C. A. Schevon, "Cognitive dysfunction in adults with epilepsy," *CNS spectrums*, vol. 15, no. S8, pp. 12–19, 2010. [Page 3.]

[7] K. M. Fiest, J. Dykeman, S. B. Patten, S. Wiebe, G. G. Kaplan, C. J. Maxwell, A. G. M. Bulloch, and N. Jetté, "Depression in epilepsy: a systematic review and meta-analysis," *Neurology*, vol. 80, no. 6, pp. 590–599, 2013. [Page 3.]

[8] H. M. de Boer, S. L. Moshé, E. Gaily, R. Grünewald, W. Mühlnickel, M. Siniatchkin, J. W. Sander, and H. Meinardi, "The global burden and stigma of epilepsy," *Epilepsia*, vol. 49, pp. 50–54, 2008. [Page 3.]

[9] P. E. Smith, "The initial diagnosis of epilepsy," *British journal of hospital medicine*, vol. 76, no. 12, pp. 690–695, 2015. [Page 4.]

[10] W. C. LaFrance Jr and O. Devinsky, "Psychogenic nonepileptic seizures," *Current neurology and neuroscience reports*, vol. 8, no. 4, pp. 306–313, 2008. [Page 4.]

[11] O. Devinsky, A. Vezzani, T. J. O'Brien, M. Dichter, E. Perucca, W. Löscher, and I. E. Scheffer, "Epilepsy," *Nature reviews Disease primers*, vol. 4, no. 1, pp. 1–23, 2018. [Page 4.]

[12] E. Niedermeyer and F. L. Da Silva, *Electroencephalography: Basic principles, clinical applications, and related fields*, 5th ed. Lippincott Williams & Wilkins, 2005. [Page 4.]

[13] E. R. Kandel, J. H. Schwartz, T. M. Jessell, S. A. Siegelbaum, and A. Hudspeth, *Principles of neural science*. McGraw-Hill, New York, 2013, vol. 5. [Pages 4 and 5.]

[14] S. Sanei and J. A. Chambers, "Eeg signal processing," 2007. [Page 4.]

[15] S. Noachtar and A. S. Peters, "The semiology of epileptic seizures: a critical review," *Epilepsy Behavior*, vol. 15, no. 1, pp. 2–10, 2009. [Page 4.]

[16] M. Teplan, "Fundamentals of eeg measurement," *Measurement science review*, vol. 2, no. 2, pp. 1–11, 2002. [Pages 4 and 5.]

[17] G. H. Klem, H. O. Lüders, H. H. Jasper, and C. Elger, "The ten-twenty electrode system of the international federation," *Electroencephalography and clinical neurophysiology. Supplement*, vol. 52, pp. 3–6, 1999. [Pages 4 and 5.]

[18] R. Oostenveld and P. Praamstra, "Guidelines for the use of eeg and meg in cognitive neuroscience research: a checklist of good practice," *Clinical Neurophysiology*, vol. 112, no. 4, pp. 575–577, 2001. [Page 4.]

[19] E. K. St Louis and N. Foldvary-Schaefer, "Epilepsy and arousals from sleep," *Sleep medicine reviews*, vol. 13, no. 4, pp. 297–307, 2009. [Page 5.]

[20] M. K. Islam, A. Rastegarnia, and Z. Yang, "Eeg artifacts handling in brain-computer interface applications: a comprehensive review," *Neural Processing Letters*, vol. 48, pp. 877–900, 2018. [Page 5.]

[21] X. Jiang, G.-B. Bian, and Z.-A. Tian, "Removal of artifacts from eeg signals: a review," *Sensors*, vol. 19, no. 5, p. 987, 2019. [Page 5.]

[22] M. Fatourechi, A. Bashashati, R. K. Ward, and G. E. Birch, "Emg and eog artifacts in practical bci systems: a survey," *Clinical neurophysiology*, vol. 118, no. 3, pp. 486–504, 2007. [Page 5.]

[23] A. Bagnall, J. Lines, A. Bostrom, J. Large, and E. Keogh, "The great time series classification bake off: a review and experimental evaluation of recent algorithmic advances," *Data Mining and Knowledge Discovery*, vol. 31, pp. 606–660, 2017. [Pages 5 and 6.]

[24] Z. Xing, J. Pei, and E. Keogh, "A brief review on time series analysis," *ACM SIGKDD Explorations Newsletter*, vol. 12, no. 1, pp. 40–48, 2010. [Page 5.]

[25] P. Esling and C. Agon, "Time-series data mining," *ACM computing surveys (CSUR)*, vol. 45, no. 1, pp. 1–34, 2012. [Page 6.]

[26] T. Rakthanmanon, B. Campana, A. Mueen, G. Batista, B. Westover, Q. Zhu, J. Zakaria, and E. Keogh, "Searching and mining trillions of time series subsequences under dynamic time warping," *ACM Transactions on Knowledge Discovery from Data (TKDD)*, vol. 7, no. 3, pp. 1–45, 2013. [Page 6.]

[27] L. Ye and E. Keogh, "Time series shapelets: a new primitive for data mining," pp. 947–956, 2009. [Page 6.]

[28] H. I. Fawaz, G. Forestier, J. Weber, L. Idoumghar, and P.-A. Muller, "Deep learning for time series classification: a review," *Data mining and knowledge discovery*, vol. 33, pp. 917–963, 2019. [Page 6.]

[29] Z. Wang, W. Yan, and T. Oates, "Time series classification from scratch with deep neural networks: A strong baseline," pp. 1578–1585, 2017. [Page 6.]

[30] H. I. Fawaz, B. Lucas, G. Forestier, C. Pelletier, D. F. Schmidt, J. Weber, G. I. Webb, L. Idoumghar, P.-A. Muller, and J. Lines, "Inceptiontime: Finding alexnet for time series classification," *Data*

Mining and Knowledge Discovery, vol. 34, pp. 1936–1962, 2020. [Page 6.]

[31] H. Ismail Fawaz, G. Forestier, J. Weber, L. Idoumghar, and P.-A. Muller, "Deep learning for time series classification: A review," *Data Mining and Knowledge Discovery*, vol. 33, no. 4, pp. 917–963, 2019. [Page 6.]

[32] G. Zerveas, S. Jayaraman, D. Patel, A. Bhamidipaty, and C. Eickhoff, "A transformer-based framework for multivariate time series representation learning," pp. 2114–2124, 2021. [Page 6.]

[33] A. Bagnall, J. Lines, J. Hills, and A. Bostrom, "Time-series classification with COTE: The collective of transformation-based ensembles," *IEEE Transactions on Knowledge and Data Engineering*, vol. 27, no. 9, pp. 2522–2535, 2015. [Page 6.]

[34] J. Lines, S. Taylor, and A. Bagnall, "Time series classification with HIVE-COTE: The hierarchical vote collective of transformation-based ensembles," *ACM Transactions on Knowledge Discovery from Data (TKDD)*, vol. 12, no. 5, pp. 1–35, 2018. [Page 6.]

[35] M. Middlehurst, J. Large, M. Flynn, J. Lines, A. Bostrom, and A. Bagnall, "HIVE-COTE 2.0: a new meta ensemble for time series classification," *Machine Learning*, vol. 110, no. 11-12, pp. 3211–3243, 2021. [Page 6.]

[36] B. D. Fulcher, M. A. Little, and N. S. Jones, "Highly comparative time-series analysis: the empirical structure of time series and their methods," *Journal of the Royal Society Interface*, vol. 10, no. 83, p. 20130048, 2013. [Pages 6 and 7.]

[37] C. H. Lubba, S. S. Sethi, P. Knaute, S. R. Schultz, B. D. Fulcher, and N. S. Jones, "catch22: Canonical time-series characteristics," 2019. [Online]. Available: https://arxiv.org/abs/1901.10200 [Pages 6, 7, and 21.]

[38] A. Dempster, F. Petitjean, and G. I. Webb, "Rocket: exceptionally fast and accurate time series classification using random convolutional kernels," *Data Mining and Knowledge Discovery*, vol. 34, no. 5, p. 1454–1495, Jul. 2020. doi: 10.1007/s10618-020-00701-z. [Online]. Available: http://dx.doi.org/10.1007/s10618-020-00701-z [Pages 7 and 11.]

[39] A. P. Ruiz, M. Flynn, J. Large, M. Middlehurst, and A. Bagnall, "The great multivariate time series classification bake off: a review and experimental evaluation of recent algorithmic advances," *Data Mining and Knowledge Discovery*, vol. 35, no. 2, pp. 401–449, 2021. [Page 7.]

[40] A. Dempster, D. F. Schmidt, and G. 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*, ser. KDD '21. ACM, Aug. 2021. doi: 10.1145/3447548.3467231 p. 248–257. [Online]. Available: http://dx.doi.org/10.1145/3447548.3467231 [Pages 7 and 11.]

[41] C. W. Tan, A. Dempster, C. Bergmeir, and G. I. Webb, "Multirocket: multiple pooled representations from random convolutional kernels for fast and accurate time series classification," *Data Mining and Knowledge Discovery*, vol. 36, no. 5, pp. 1738–1762, 2022. [Page 7.]

[42] G. Uribarri, F. Barone, A. Ansuini, and E. Fransén, "Detach-rocket: Sequential feature selection for time series classification with random convolutional kernels," 2024. [Online]. Available: https://arxiv.org/abs/2309.14518 [Pages 7, 11, 12, and 21.]

[43] A. Solana, E. Fransén, and G. Uribarri, "Classification of raw meg/eeg data with detach-rocket ensemble: An improved rocket algorithm for multivariate time series analysis," 2024. [Online]. Available: https://arxiv.org/abs/2408.02760 [Pages 7, 12, and 15.]

[44] V. Shah, E. von Weltin, S. Lopez, J. McHugh, L. Veloso, M. Golmohammadi, I. Obeid, and J. Picone, "The temple university hospital seizure detection corpus," *Frontiers in Neuroinformatics*, vol. 12, p. 83, 2018. doi: 10.3389/fninf.2018.00083. [Online]. Available: https://doi.org/10.3389/fninf.2018.00083 [Page 9.]

[45] I. Obeid and J. Picone, "The temple university hospital eeg data corpus," in *Augmentation of Brain Function: Facts, Fiction and Controversy. Volume I: Brain-Machine Interfaces*, 1st ed., C. Guger, B. Z. Allison, and G. Edlinger, Eds. Lausanne, Switzerland: Frontiers Media S.A., 2018, pp. 394–398. [Page 9.]

[46] P. Detti, "Siena scalp eeg database (version 1.0.0)," https://doi.org/10.13026/5d4a-j060, 2020, physioNet. [Page 9.]

[47] P. Detti, G. Vatti, and G. Z. M. de Lara, "EEG Synchronization Analysis for Seizure Prediction: A Study on Data of Noninvasive Recordings," *Processes*, vol. 8, no. 7, p. 846, 2020. doi: 10.3390/pr8070846. [Online]. Available: https://doi.org/10.3390/pr8070846 [Page 9.]

[48] J. Dan, U. Pale, A. Amirshahi, W. Cappelletti, T. M. Ingolfsson, X. Wang, A. Cossettini, A. Bernini, L. Benini, S. Beniczky, D. Atienza, and P. Ryvlin, "Szcore: Seizure community open-source research evaluation framework for the validation of electroencephalography-based automated seizure detection algorithms," *Epilepsia*, vol. n/a, no. n/a. doi: https://doi.org/10.1111/epi.18113. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1111/epi.18113 [Page 10.]

[49] J. Dan *et al.*, "Library for measuring performance of time series classification," https://github.com/esl-epfl/timescoring, 2023, accessed: 2025-05-04. [Page 13.]

If you do not have an appendix, do not include the \cleardoublepage command below; otherwise, the last page number in the metadata will be one too large.