From Epilepsy Seizures Classification to Detection: A Deep Learning-based Approach for Raw EEG Signals

Davy Darankoum, Manon Villalba, Clélia Allioux, Baptiste Caraballo, Carine Dumont, Eloïse Gronlier, Corinne Roucard, Yann Roche, Chloé Habermacher, Sergei Grudinin, et al.

From epilepsy seizures classification to detection: A deep learning-based approach for raw EEG signals

Davy Darankoum[1,2], Manon Villalba[1], Clélia Allioux[1], Baptiste Caraballo[1], Carine Dumont[1], Eloïse Gronlier[1], Corinne Roucard[1], Yann Roche[1], Chloé Habermacher[1], Sergei Grudinin[2,*], Julien Volle[1,*]

1. SynapCell SAS, 38330 Saint-Ismier, France
2. Univ. Grenoble Alpes, CNRS, Grenoble INP, LJK, 38000 Grenoble, France
***. Both authors contributed equally to this work.**

Corresponding authors: Sergei.Grudinin@univ-grenoble-alpes.fr, jvolle@synapcell.fr

Abstract

Epilepsy represents the most prevalent neurological disease in the world. One-third of people suffering from mesial temporal lobe epilepsy (MTLE) exhibit drug resistance, urging the need to develop new treatments. A core part in anti-seizure medication (ASM) development is the capability of detecting and quantifying epileptic seizures occurring in electroencephalogram (EEG) signals, which is crucial for treatment efficacy evaluation.
In this study, we introduced a novel seizure detection pipeline and trained several deep-learning models on raw EEG signals. The pipeline integrates (i) a new pre-processing technique, which segments continuous raw EEG signals without prior distinction between seizure and seizure-free activities; (ii) a post-processing algorithm developed to reassemble EEG segments, which allows the identification of seizures' beginnings/ends; (iii) and finally, a new evaluation procedure based on a strict seizure events comparison between predicted and manually annotated labels. We constructed and trained convolutional and transformer-based models using a data-splitting strategy, addressing potential data leakage. We demonstrated the fundamental distinction between the seizure classification and seizure detection tasks and showed the difference in the performance between them. Finally, we illustrated the generalization capabilities of our best architecture, trained on animal EEGs, and tested on human EEGs, which achieved an F1-score of 93%.

Key words: Epilepsy, Raw EEG, Seizure classification, Seizure detection, Deep neural networks, CNN, Transformers

1. Introduction

Epilepsy affects more than 50 million individuals worldwide and is characterized by recurring seizures arising from abnormal brain activity, profoundly impacting daily functioning and quality of life (Beghi, 2020; Thijs et al., 2019). Epilepsy can appear through several syndromes, underscoring the need for precise and effective diagnosis to orient epileptic patients toward appropriate healthcare treatments (Nabbout and Kuchenbuch, 2020). Anti-seizure medications (ASMs) suffer from high patient-dependent responses. The development of accurate tools to extract specific information from epileptic patients appears then to be paramount for developing better-suited medications.

Electroencephalogram (EEG) stands as a pivotal tool in epilepsy diagnosis, thanks to one of its capacities to capture substantial alterations in brain electrical activity during and in proximity to epileptic seizures (Mesraoua et al., 2019; Noachtar and Rémi, 2009; Shoji et al., 2021). Neurologists classify brain activity into four distinguishable phases based on EEG inspection. The preictal phase represents the period preceding a seizure; the ictal phase corresponds to the actual seizure event; the postictal phase encompasses the time following a seizure episode; and the interictal phase constitutes the interval between seizure occurrences, distinct from the other states. Different analysis procedures involving some or all of these four phases help neurologists diagnose the proper type of epilepsy (Toraman, 2021). Moreover, epileptic seizure identification and quantification are broadly used to evaluate the efficacy of new ASMs and disease-modifying therapies.

Among the different epilepsy syndromes, Mesial Temporal Lobe Epilepsy (MTLE), characterized by refractory seizures, is the most common type of focal epilepsy in adults. Approximately 30-50% of patients with MTLE develop drug resistance (Ammothumkandy et al., 2022; Paschen et al., 2020). Considerable research endeavors presently focus on optimizing the preclinical stage of drug development for epilepsy to enhance translational success and improve the likelihood of therapeutic candidates advancing through clinical trials. The preclinical phase designates a stage where studies are conducted on laboratory animals to identify the best treatment candidates among many and to determine safe doses in order to accelerate and increase success chances on the tests carried out on humans, namely the «clinical phase». The translational properties of EEG signals make them a valuable tool for monitoring brain activity in animal models, facilitating the extrapolation of findings to human brain function. However, due to a lack of accurate analysis tools, neurologists often review and interpret EEG signals manually, which leads to misidentification of epileptic seizures, inefficiencies, and subjectivity. Therefore, there is a need to develop automated techniques to identify seizures accurately and minimize diagnostic errors. This automation task is rather challenging given the complex characteristics of EEG signals, including their low signal-to-noise ratio, high-frequency dimension, non-stationarity, non-linearity, variability, and the presence of artifacts.

Thanks to the constant progress in machine learning-based techniques, many repetitive data-annotation tasks can now be automated, thus affording minimal room for errors and liberating

oneself from protracted, time-consuming activities (Ebrahim et al., 2020; Jumper et al., 2021; LeCun et al., 2015; Min et al., 2016). These advances have also extended to seizure detection in EEG signals (Chen et al., 2023; Choi et al., 2019; Hussain et al., 2019). We can divide approaches that automate epileptic seizure detection through machine learning into two main categories. The first includes research focusing on hand-crafted feature extraction from EEG signals, followed by training classical machine-learning or deep-learning models to result in an epileptic seizure detection tool (Mahjoub et al., 2020; Shoeb and Guttag, 2010). The second category tackles the task by training deep-learning models to automatically extract meaningful features in EEG signals and simultaneously perform seizure detection.

One can manually extract features from the EEG signals in multiple ways: in the time domain, frequency domain, time-frequency domain, or with nonlinear analysis (Durongbhan et al., 2019; Singh and Krishnan, 2023). A study introduced by Guo et al. (2010) presents an epileptic seizure detection pipeline based on the computation of line-length features from wavelet transform-based signal decomposition, a technique that allows feature extraction from the time-frequency domain of EEG signals. Then, the authors used the extracted features with a Multi-Layer Perceptron Neural Network (MLPNN) to perform a seizure detection task. Their model achieved very high performance (98% accuracy on the classification into seizure and non-seizure segments) on the Bonn University dataset published by Andrzejak et al. (2001). In another study, Wang et al. (2018) proposed a real-time seizure detection algorithm based on Short Time Fourier Transform (STFT), another technique to extract time-frequency domain features from the EEG, and then trained them with a Support Vector Machine (SVM), a machine learning-based model. The authors tested their pipeline on the CHB-MIT Scalp EEG database and achieved 98% sensitivity on the seizure detection task. Mursalin et al. (2017) performed a correlation-based feature selection from the time domain and frequency domain of EEGs and then applied an ensemble of random forest classifiers (machine learning-based models) to detect seizures on EEG. On the Bonn University dataset, their model achieved 97% accuracy. Despite the high reported accuracies, the models developed for epileptic seizure detection and based on manual feature extraction exhibit several limitations, mainly regarding their capacity to generalize across diverse subject profiles and conditions (Einizade et al., 2020). This poor generalization indicates that some features invariant to variabilities across subjects' EEG signals might not be learned in the trained models. Moreover, the subject-dependent signal-to-noise ratio causes a high variability of feature importance for seizure detection. Some researchers attempted to overcome these limitations by developing deep-learning models that automatically extract meaningful features in EEG signals by learning invariant embeddings, which we briefly describe below.

Deep-learning models, such as Convolutional Neural Networks (CNNs), with their capacity for extracting local features, and Recurrent Neural Networks (RNNs), capable of capturing long-range relationships, serve as relevant tools for acquiring consistent and translational features essential for EEG classification into seizure and non-seizure segments. Acharya et al. (2018) proposed one of the first CNN-based networks trained on raw EEG time series to classify EEGs into seizure and non-seizure activities. They achieved 89% accuracy on the Bonn University dataset. Roy et al. (2018) applied different EEG pre-processing techniques coupled

to several neural network architectures, namely, a 1D CNN, a 2D CNN, and a 1D CNN-GRU (Gated Recurrent Unit), to classify EEG signals into normal and abnormal activities. Their best model (1D-CNN-GRU) demonstrated 99% accuracy on the TUH Abnormal EEG Corpus (Obeid and Picone, 2016). Cho and Jang (2020) compared four input modalities (raw time series EEG, periodograms that reflect the spectral density of EEG signals, 2D images from STFT coefficients, and 2D images from raw EEG waveforms) and different neural networks for an epileptic seizure detection task. They trained fully connected neural networks, RNN and CNN, on unique or combined input types listed above. Their best pipeline led to a 99% accuracy on UPenn and Mayo Clinic's Seizure Detection challenge datasets.

This study presents the development of deep-learning-based pipelines to automatically detect epileptic seizures in EEG signals from an animal model of MTLE, the intra-hippocampal kainate mouse model. We focused on MTLE mainly due to the prevalence of this epileptic form worldwide (Tatum, 2012) and because MTLE manifests by seizures occurring in a specific brain region, thus limiting the use of common multi-electrodes set-ups for EEG recording. MTLE is a focal epilepsy type characterized by recurrent seizures with an onset involving the amygdalohippocampal complex and parahippocampal region. Consequently, seizures can be captured only using a single electrode positioned near the onset site. In contrast, non-focal epilepsy types involve seizures occurring across multiple brain regions, allowing for the use of multiple electrodes to capture them. Our second goal was the development of accurate tools for enhancing preclinical studies' workflow. We trained and validated the developed neural network architectures using EEG signals recorded on MTLE mice. We evaluated the generalization performance of our top-performing models by applying them to signals from human patients. Our exploration of different neural network architectures included convolutional neural networks, recurrent neural networks, segmentation models based on the U-Net architecture, and attention-based networks. We further developed a post-processing algorithm that concatenates overlapping signal segments into continuous time series, which allows the detection of seizures in a real-world scenario. Finally, we identified two evaluation strategies for assessing model performance, which we believe are more effective for benchmarking the efficacy of automated seizure detection tools.

2. Materials and Methods

2.1. Animals - MTLE mice model

Animal experiments were approved by the ethical committee of the Grenoble Institute of Neuroscience, University Grenoble Alpes, and performed by SynapCell in accordance with the European Committee Council directive of September 22, 2010 (2010/63/EU). Duveau and Roucard (2017) has previously described detailed information about the generation of the MTLE mouse model. Briefly, adult male C57Bl/6J mice (11 weeks of age) receive a kainic acid injection in the right dorsal hippocampus (AP = -2, ML = -1.5, DV = -2 mm relative to bregma) (Paxinos and Franklin, 2001). During the surgical procedure, a bipolar electrode is

positioned in the right dorsal hippocampus (AP = -2.4, ML = -1.5, DV = -2 mm relative to bregma). The implant is secured to the skull using dental cement to allow tethered EEG recordings in freely moving animals. After surgery, animals are left in their home cage for at least one week of recovery. After an epileptogenesis period lasting four weeks, mice got accustomed to the recording conditions, and EEGs were recorded to assess each animal. Different criteria (number of HPD = Hippocampal Paroxysmal Discharges, sufficient signal-to-noise ratio), allowing a distinct determination of the beginning and end of events, are used to enroll animals in a study. For all experiments, the criterion of inclusion is rigorously the same. During the study, animals are connected to an amplifier by a recording cable that does not restrict their movement. The EEG signal is band-pass filtered between 0.8 Hz and 1 kHz and digitized at 512 Hz (SDLTM128 Channels; Micromed, France). EEGs are stored for offline analysis, allowing experts to evaluate all the EEG traces and annotate the boundaries of each HPD.

2.2 Datasets

In our study, we used two datasets. Both contain EEG signals recorded on subjects (mice for the first dataset and humans for the second one) who suffer from MTLE.

Dataset #1: This dataset results from a selection of EEG recorded in ten different studies conducted at SynapCell. It includes 1440 hours of EEG signals recorded in 136 MTLE mice: 1190 hours of seizure-free activity and 250 hours of epileptic seizures. Each mouse is recorded following an average of 3 different sessions, with each session lasting approximately 3 hours. Each signal has been reviewed by an expert scorer, who labeled ictal activities with the assistance of a commercial software for seizure detection (Deltamed Coherence, Natus Medical Incorporated, USA).

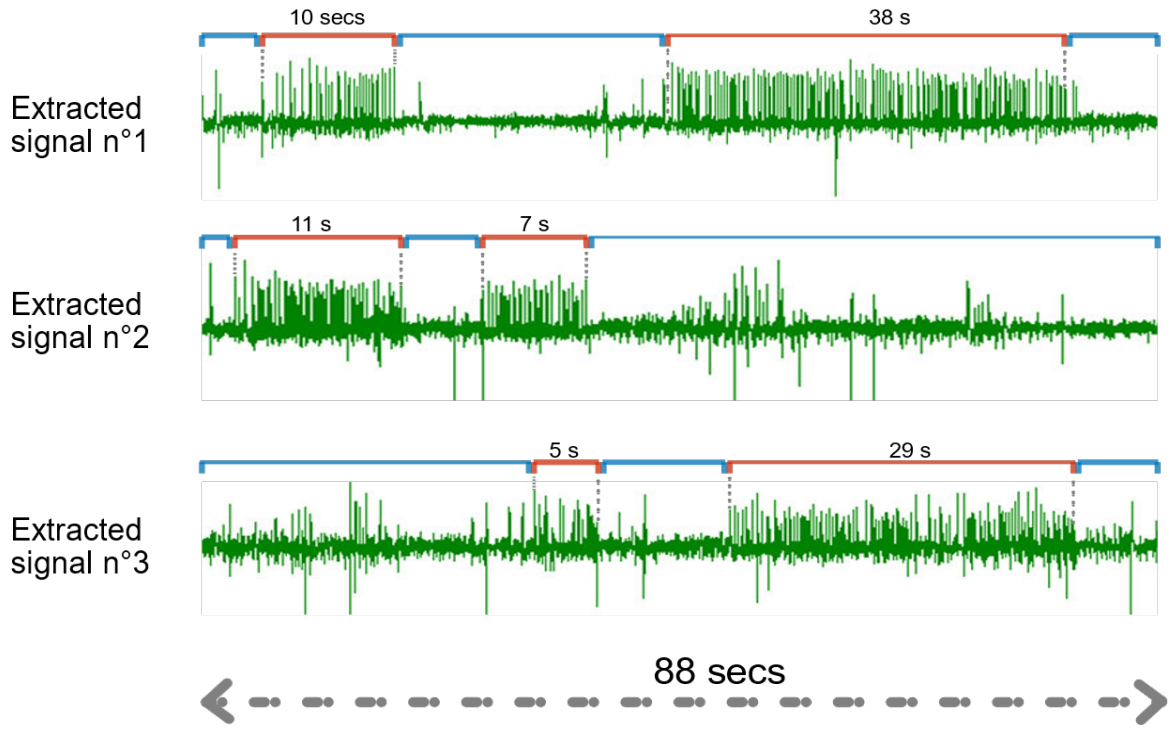

Figure 1. Examples of EEG signals from Dataset #1. Snapshots of 3 portions of EEG signals measured in 3 mice and labeled by the same expert. Labels in red indicate detected seizures, and blue labels represent seizure-free activities. The total duration of the snapshots is 88 seconds.

Dataset #2: The second dataset is public from Bonn University (Andrzejak *et al.*, 2001). The dataset contains EEG signals recorded on healthy humans and humans suffering from MTLE through a multi-channel set-up, but they provide only one-channel data. We chose it for our study as it exhibits the same EEG data constraints of single-channel measurements as dataset #1. It also allows assessing the generalization abilities of our models from the preclinical setting to the clinical environment.

The dataset comprises five different sets denoted from A to E. Each contains 100 EEG segments of 23.6 sec. These segments result from continuous EEG recordings that were processed by the authors to remove artifacts. In total, they recorded EEG signals from ten humans: five healthy and five diagnosed with MTLE. The EEG signals recorded through the head surface of the five healthy volunteers formed sets A and B. Set A consists of EEGs recorded with eyes open and set B with eyes closed. Presurgical EEGs from five patients suffering from MTLE were used to constitute sets C, D, and E. Set D comprises EEG recorded from the epileptogenic zone. Set C comprises EEG measured from the hippocampal structure of the opposite hemisphere. Segments in sets C and D contain only the brain activity measured during seizure-free intervals, whereas set E contains EEG segments recorded through the epileptogenic zone (the hippocampal structure on the hemisphere from which the seizures originate) during seizure activity. The authors recorded all the EEG signals from these sets at a sampling rate of 173.61 Hz and a bandpass filter to keep only frequencies between 0.53 and 40 Hz.

2.3. Pre-processing and post-processing pipelines

2.3.1. Dataset #1

We used Dataset #1 to train models for two tasks: seizure classification and seizure detection. **Table 1** lists the processing stages, including data pre-processing, post-processing, and model evaluation.

Mode	Task	
	Classification (**C**)	Detection (**D**)
Training	Pre-processing **C** ⬇ Model training	Pre-processing **C** ⬇ Model training
Inference time	Pre-processing **C** ⬇ Model application ⬇ Evaluation **C**	Pre-processing **D** ⬇ Model application ⬇ Post-processing **D** ⬇ Evaluation **D**

Table 1. Task-based (Classification and Detection) pipelines summary.

Pre-processing with prior identification of seizure/seizure-free activity (Pre-processing C)

We pre-process each EEG signal individually. Firstly, the signal is resampled from 512 Hz to 100 Hz. The resulting downsampled signal is filtered using a bandpass finite impulse response (FIR) filter between 1 and 20 Hz. We selected downsampling and filtering parameters based on their impact on the trained models' accuracy. Then, we annotated continuous ranges of epileptic and seizure-free activity according to the onset and offset intervals of seizures labeled by experts. We then perform a Z-score normalization of the signal amplitudes. The mean and the variance for the normalization are calculated on the selection of all the amplitudes extracted from the seizure-free activity of the signal. Finally, we segment each range of seizure or seizure-free activities separately into 2- or 4-second blocks with variable overlap (or shift) size. The overlap is the common signal part between two consecutive blocks and the shift refers to the signal between the beginnings of two consecutive blocks. This pre-processing leads to 2- or 4-second blocks that do not contain mixed activities (seizure and seizure-free, please see **Fig 2.**).

Pre-processing without prior identification of seizure/seizure-free activity (Pre-processing D)

Here, we also pre-process each EEG signal individually. The same resampling from 512 Hz to 100 Hz and bandpass filter from 1 to 20 Hz is applied. We then perform the Z-score normalization of the signal amplitudes as before. The mean and variance for the normalization are, however, computed using only the first 5 minutes of the signal. Finally, using a sliding window starting at time zero, the signal is segmented into overlapped 2- or 4-second blocks. It

is worth noting that during this pre-processing procedure, the segmentation into 2- or 4-second blocks is applied without any prior distinction between seizure and seizure-free activities to mimic a real-world scenario. Such pre-processing will lead to 2- or 4-second blocks containing mixed activities (seizure and seizure-free, please see **Fig 3.**).

Post-processing for signal reconstitution (Post-processing D)

Following the application of a model at inference time to segments pre-processed without prior identification of seizure/seizure-free activity (Pre-processing D), we obtain a list of segments with labels predicted as seizures or non-seizures. Then, an ad-hoc post-processing algorithm is used to combine these segments and reconstruct the original signal. Segments with overlapping ranges are merged into a single range, and the predominant label is assigned to it. The procedure is iteratively applied until it produces a new set of labels on the continuous reconstructed EEG signal. We then compare this reconstructed signal to the original one following the evaluation D strategy.

Evaluation strategy for seizure classification (Evaluation C)

After model application at inference time to segments pre-processed with prior identification of seizure/seizure-free activity (Pre-processing C), we obtain a list of segments classified with seizure or seizure-free activity labels, one label per segment. Following a simple binary classification strategy, each predicted label is characterized as a correct or incorrect prediction, which allows the computation of true positives (TP: segment predicted as “seizure” while the true label is also “seizure”), true negatives (TN: segment predicted as “non-seizure” while the true label is also “non-seizure”), false positives (FP: segment predicted as “seizure” while the true label is “non-seizure”), and false negatives (FN: segment predicted as “non-seizure” while the true label is “seizure”).

Evaluation strategy for seizure detection (Evaluation D)

This evaluation strategy aims to build a reliable metric for the seizure detection task. After signal reconstruction following the post-processing D method, we used an event-based metric previously introduced by (Mesaros et al., 2016) to evaluate models. Adapted to the seizure detection task, this evaluation method compares each event (seizure) detected by the model to events labeled by the expert.

TP (True Positive): an event detected by the model overlaps an event labeled by the expert within a tolerance range of 1 second. The start of the detected event must be in a range of ± 1 second of the beginning of the labeled event and likewise for the end of the two events.

FN (False Negative): an event labeled by the expert has not found a corresponding event detected by the model within the (+-) 1–second tolerance.

FP (False Positive): an event detected by the model has not found a corresponding event labeled by the expert within the (+-) 1–second tolerance.

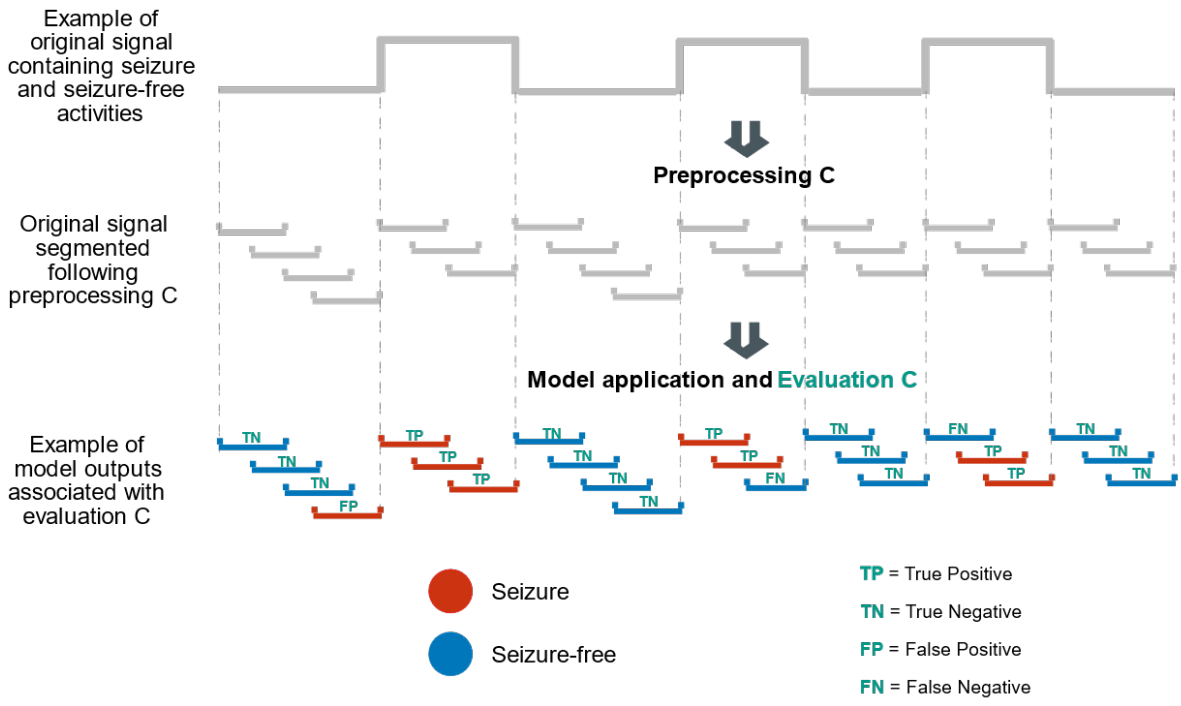

Figure 2. Illustration of the classification task pipeline at inference time. Segments built out of pre-processing C do not overlap across two activities. Segments colored in blue or red reflect an example of classification by the trained model. Blue color corresponds to classification into seizure-free activity. The red color indicates classification into seizure activity.

TP : Segment labeled as seizure and detected as seizure by the model

TN : Segment labeled as seizure-free and detected as seizure-free

FP : Segment labeled as seizure-free and detected as seizure by the model

FN : Segment labeled as seizure and detected as seizure-free

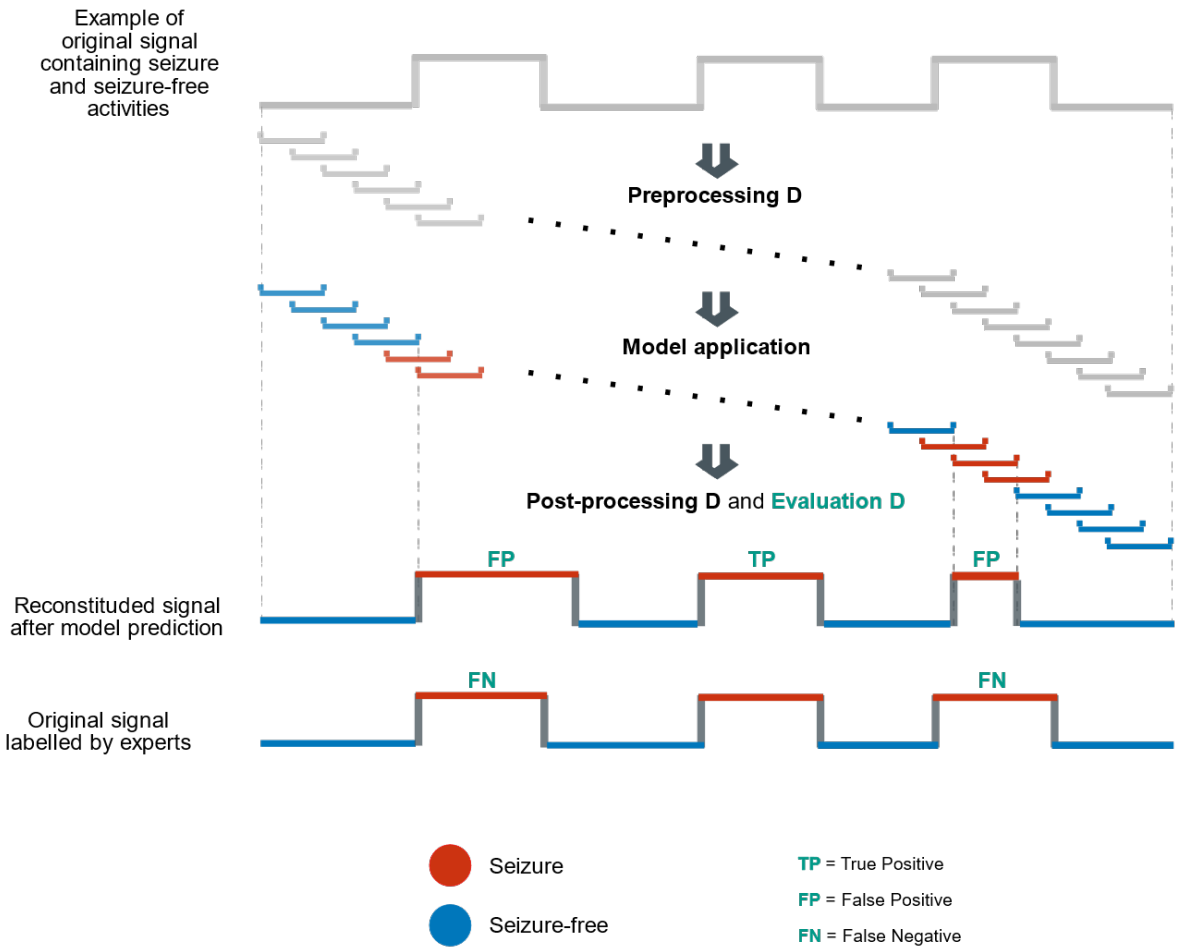

Figure 3. Illustration of the detection task pipeline at inference time. Segments built out of pre-processing D do overlap across activities exhibiting a real-world scenario. Segments predicted through the model application are reassembled into a continuous signal using post-processing D. Finally, following evaluation D strategy, events formed in the reconstituted signal are compared with events labeled by the expert.

TP : A seizure start & end labeled by the expert matches a seizure start & end detected by the model*
FP : A seizure start & end detected by the model did not find a match with any seizure start & end labeled by the expert*
FN : A seizure start & end labeled by the expert did not find a match with any seizure start & end detected by the model*

***With a tolerance range of 1 sec.**

2.3.2. Dataset #2

Dataset #2 (Bonn public dataset) consists of signals already separated into seizure or non-seizure activities. Therefore, we could only perform a classification task on this dataset, as the

detection task requires processing continuous signals containing both seizure and non-seizure activities.

To facilitate a comparison with dataset #1, we also downsampled each signal of 23.6 seconds, from 173.61 Hz to 100 Hz. Then, a Z-score normalization is applied, using the mean and the variance calculated from sets A, B, C, or D or the combination of some of them (refer to the subset description below). Initially, the Bonn dataset was unbalanced, with 100 segments (Set E) related to the seizure activity and 400 segments (Sets A, B, C, and D) labeled as seizure-free activity. To evaluate the robustness of our models, we introduced three balance ratios of the two classes by selecting different set combinations as follows:
Subset 1: a dataset constructed of sets A, B, C, D, and E. It is unbalanced, with 80% of seizure-free activity and 20% of seizure activity.
Subset 2: a dataset constructed of sets A, B, and E. It is unbalanced, with 66.6% of seizure-free activity and 33.3% of seizure activity.
Subset 3: a dataset formed of sets C and E. The balance is 50% of seizure-free activity and 50% of seizure activity.
In all the assembled subsets, we used the segmentation into blocks of 4 seconds with 50% overlap. We finally used the evaluation for seizure classification (Evaluation C) strategy to compute TP, TN, FP, and FN metrics.

2.3.3. Training, Validation, and Test sets

We applied a rigorous approach to data splitting to ensure data integrity and prevent data leakage in dataset #1. Specifically, if a signal from animal A was allocated to the training set, we guaranteed that no other signals from the same animal A would be included in the validation or test sets.
Among the 136 animals, 100 were randomly selected to constitute the training set, 19 others for the validation set and, 17 left for the test set. EEG signals recorded from the selected animals were assigned to each of the three groups.

- Training set: EEG recorded from the selected 100 animals correspond to 184 hours of epileptic seizure activity and 870 hours of seizure-free activity. To balance the training set, we selected 200 hours of seizure-free activity among the 870 hours.
- Validation set: EEG recorded from the selected 19 animals correspond to 44 hours of epileptic seizure activity and 210 hours of seizure-free activity. To facilitate a more straightforward evaluation of the model during the training process, we selected 50 hours of seizure-free activity among the 210 hours to balance the validation set.
- Test set: EEG recorded from the selected 17 animals correspond to 22 hours of seizure activity and 110 hours of seizure-free activity. We kept the test set unbalanced to evaluate the models' generalization capabilities in a real-world scenario.

	Seizure activity (hours)	Seizure-free activity (hours)
Training set	184 (73%)	200 (55%)
Validation set	44 (17%)	50 (14%)

Test set	22 (10%)	110 (31%)
Dataset #1 selected	250 (100%)	360 (100%)

Table 2. Training, validation and test set repartition from dataset #1.

In this study, dataset #2 is used entirely as a test set to evaluate the generalization capabilities of our best models.

2.4. Network architectures

To address the challenge of accurate seizure detection in EEG signals, we leveraged the capabilities of deep learning models. As introduced above, EEG signals, by nature, present a high variability across subjects. From one animal to another, before and after the administration of a pharmaceutical condition, significant differences can be observed in the EEG signals according to the amplitudes or the signal-to-noise ratio. Model architectures employed in this study for automatic seizure detection on EEG signals can be categorized into three groups: (1) convolutional neural networks (CNN)-based models, (2) models that combine CNN and recurrent neural networks (RNN)-based architectures, and (3) models that combine CNN and transformer architectures.

2.4.1. CNN-based architectures

Convolutional neural networks excel in identifying local patterns in images or time series data (Roy et al., 2018). The principal element constituting this network is a convolutional layer followed by a nonlinear activation function and, very often, resolution reduction operations like maximum/average pooling layers.

Classical CNN architectures

We constructed these network architectures by combining convolutional layers, batch normalization layers, Rectified Linear Unit (ReLU), max-pooling layers, SoftMax activation functions, and dense layers. We rigorously applied the following order to all the constructed CNN-based architectures. They all start with a convolutional layer followed by a batch normalization layer, a ReLu activation function and a max pooling layer. These blocks of four grouped layers are linked in a consecutive manner **multiple** times (**3**, **5**, **6**, **12**, or **16**). These blocks are followed by two dense layers separated by a ReLU function, and finally, the learned embeddings are followed by a SoftMax activation function for the final classification into epileptic/seizure-free activity.

Customized U-Time architectures

U-Time (Perslev et al., 2019) is an architecture formed by a downstream network similar to a CNN-based network with blocks of convolutional, pooling, batch normalization layers, and

activation functions. It is followed by an upstream network, also based on CNN-based networks but with pooling layers replaced by upsampling layers that increase the data dimensionality previously decreased by pooling layers on the downStream part. U-Time is a modified version of U-Net (Ronneberger et al., 2015), a network where the term “U” refers to the shape of the network architecture. Such networks are called segmentation-based networks. The original U-Network allows us to achieve image segmentation tasks. U-Time is the adapted version for time series data like EEG signals. To adjust the U-Time network to our seizure detection task, we removed the classifier segment.

2.4.2. CNN+RNN-based architecture

Recurrent neural network (RNN) is a type of neural network characterized by a bi-directional flow. The output of some nodes combined with the following inputs of the same nodes guarantees a dependence between inputs and outputs as opposed to feed-forward networks like CNNs. Another distinguishable characteristic is that they share parameters across the layers in the network and have the same weight parameters within each layer, whereas feed-forward network types have different weights across each node. Typically, RNN-based networks are tailored to extract long-term relationships on time series data (Roy et al., 2018).
Classical RNN layers are prone to vanishing gradients. This limitation led to the development of long short-term memory (LSTM) layers (Sak et al., 2014) and gated-recurrent unit (GRU) layers (Cho et al., 2014), inspired by classical RNN layers and less exposed to the gradient vanishing problem. We have combined CNN layers with either LSTM or GRU layers for our classification and seizure detection tasks.

2.4.3. CNN+transformer architecture

The transformer architecture described in (Vaswani et al., 2023) comprises an encoder and a decoder part. In our study, we only used the encoder part, as our tasks do not evolve into data generation. The encoder architecture contains a multi-head attention network, performing the computation of attention scores on pairs of EEG sequences and a position-wise fully connected feed-forward network. Each of the aforementioned networks is followed by a residual connection and a layer normalization.
Our CNN+transformer architecture comprises a two-head attention network with embeddings from the raw EEG combined with positional encodings as inputs. The input embeddings are learned through two CNN-based architectures with six convolution blocks. The first CNN architecture contains convolution kernels of size 3, and the second is made of kernels of size 10 to capture frequency information on different scales (Eldele et al., 2021). The positional encodings are built with the RoPE rotary position embedding introduced by (Su et al., 2023).

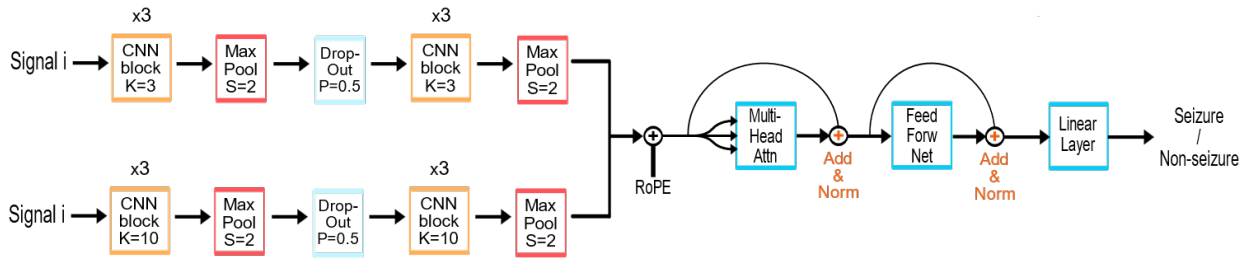

Figure 4. CNN+transformer architecture. A CNN block contains a convolution layer, batch normalization layer and ReLu activation function. K=Kernel size, S=Stride, P=Probability.

2.5. Computational details

We trained all the models for 100 epochs using the PyTorch framework and the Adam optimizer to optimize the learnable parameters. The hyperparameters optimization of the models (learning rate, number of epochs) was carried out using the validation set, and we specifically used the binary cross-entropy (BCE) loss function to tune the training performance. The computing resource used to train the models was an NVIDIA Tesla V100 GPU cluster with a single node containing 32 Go RAM.

2.6. Performance metrics

We considered the "seizure" state as the positive class and the "seizure-free" state as the negative class. Our main metrics are accuracy, sensitivity, precision, and F1-score.

Accuracy indicates the ratio of correct predictions made by the model to the total number of predictions:

$$Accuracy = \frac{Total\ Number\ of\ Predictions}{Number\ of\ Correct\ Predictions}$$

Eq. 1

Sensitivity, also called Recall, is defined as the proportion of correctly categorized positive segments to all positive segments:

$$Recall = \frac{True\ Positives}{True\ Positives + False\ Negatives}$$

Eq. 2

Precision to get the ratio of the number of segments predicted as "seizure" with respect to the number of segments labeled as "seizure" by experts is defined as:

$$Precision = \frac{True\ Positives}{True\ Positives + False\ Positives}$$

Eq. 3

Finally, F1-score is a metric that combines precision and sensitivity ratio to evaluate the performance of a model. It considers false positives and false negatives in the estimation. It indicates how well a model performs with respect to all predicted classes:

$$F1 - score \quad = \frac{2 \times Precision \times Recall}{Precision + Recall}$$

Eq. 4

3. Results

3.1. Dataset #1 models' performance on seizure classification task

We evaluated the models trained on dataset #1 using the test set representing 10% of the entire dataset.

Table 3 lists the success rates of the CNN-based architectures of different depths, from 3 to 16 convolutional layers. The model rendering the best performance is the one with 6 convolutional layers, which also appears to be the one containing the smallest number of parameters.

Architecture	Segmentation parameters	Nb trained parameters	Recall	F1-score on classification task
CNN-3 layers	Size 4s shift 2s	2.5 M	0.909	**0.818**
CNN-5 layers	Size 4s shift 2s	159 K	0.919	**0.818**
CNN-6 layers	Size 4s shift 2s	23 K	**0.926**	**0.818**
CNN-12 layers	Size 4s shift 2s	216 K	0.918	0.816
CNN-16 layers	Size 4s shift 2s	258 K	0.907	0.807

Table 3. Success rates of the CNN-based architectures on a seizure "classification" task on dataset #1.

U-Time-based models did not manage to outperform the CNN-based models, as listed in Table 4. Increasing the network depth resulted in a slightly better performance. However, F1-scores indicate lower performances compared to the CNN-based models.

Architecture	Segmentation parameters	Nb trained parameters	Recall	F1-score on classification task

Customized U-Time using 22 Conv layers	Size 4s shift 2s	943 K	**0.9**	**0.726**
Customized U-Time using 17 Conv layers	Size 4s shift 2s	236 K	0.876	0.656

Table 4. Success rates of the customized U-Time architecture on a seizure "classification" task on dataset #1.

Table 5 lists the success rates observed when combining the best CNN-based architecture from Table 3 with either a Long-Short Term Memory, a Gated Recurrent Unit, or a transformer layer. Architectures that combine CNN with RNN-based layers do not outperform CNN-only architectures while the number of their parameters increases drastically. On the other hand, the CNN+transformer architecture demonstrates the best performance (according to recall and F1 score metrics) on the seizure classification task.

Architecture	**Segmentation parameters**	**Nb trained parameters**	**Recall**	**F1-score on classification task**
CNN-6 layers + biLSTM	Size 4s shift 2s	2.6 M	0.888	0.814
CNN-6 layers + GRU	Size 4s shift 2s	2.6 M	0.895	0.808
CNN + Transformer	Size 4s shift 2s	158 K	**0.898**	**0.868**

Table 5. Success rates of CNN+LSTM, CNN+GRU and CNN+transformer architectures on a seizure "classification" task on dataset #1.

3.2. Dataset #1 models' performance on seizure classification vs seizure detection tasks

Table 6 lists the results for the best architectures selected in Tables 3, 4, and 5. These are CNN-6 layers, CNN-6 layers+biLSTM, CNN+transformer and Customized U-Time – 22. Each of these networks has been trained using two segmentation schemes (one training with segments of 4 seconds shifted by 2 seconds and another with segments of 2 seconds shifted by 1 second). The results shown in Table 6 illustrate the performance of the trained models in the classification and detection of seizure tasks.

During the detection task, models are evaluated using segments built with different shift sizes. Indeed, the detection task requires a post-processing step consisting in the concatenation of overlapping segments to form a continuous signal. The seizures detected with this post-processing method can vary according to the overlap (or shift) sizes, which justifies our experiments conducted with different shifts.
Such experiments are not relevant in the classification task because segments are evaluated solely and, therefore, different segment shift sizes do not impact the classification performance.

CNN+transformer outperforms all the proposed architectures on both tasks. In the classification task, it reached the F1-score of 87% with segments of 4 seconds shifted by 2 seconds. In the detection task, it reached the F1-score of 56% with segments of 4 seconds shifted by 0.5 seconds.

The success rates in the classification task are much higher than those in the detection task, demonstrating the complexity of the latter. Regarding all the architectures, the best scores in the detection task are obtained with segments built using 0.5-second shifts. This shows that larger overlapping windows allow a higher precision in the signals reconstitution, and, therefore, better detections of the beginning and end of seizures (events).

Architecture	**F1-score on classification task**		**F1-score on detection task**			
	Size 4s shift 2s	Size 2s shift 1s	Size 4s shift 2s	Size 4s shift 0.5s	Size 2s shift 1s	Size 2s shift 0.5s
CNN-6 layers	0.818	0.722	0.0	0.182	0.381	0.496
CNN-6 layers + biLSTM	0.814	0.725	0.0	0.16	0.242	0.377
Customized U-Time	0.726	0.694	0.0	0.0	0.0	0.0
CNN-6 layers + Transformer	**0.868**	**0.818**	**0.363**	**0.565**	**0.481**	**0.529**

Table 6. Comparison of models' performance on seizure classification vs seizure detection tasks on dataset #1.

3.3. Models trained on dataset #1 and tested on dataset #2

One of the main goals of preclinical studies is to enhance risk management and predict treatment outcomes on experiments which will be conducted in human patients. Following such a vision, we trained our best architectures (CNN-6 layers and CNN+transformer) on mice EEG (dataset #1) and tested them on human EEG (dataset #2). **Table 7** lists the obtained results. The

high performance obtained using the CNN+transformer model demonstrates its robustness and generalization capabilities. The CNN-6 layers-based model exhibits high recall scores. However, the F1-scores obtained in the experiments conducted with unbalanced data are low and, therefore, show the limits of this model. Regarding these experiments, we can confidently conclude that the CNN+transformer model translates signatures learned on brain activity from mice to humans.

	Subset 1 Seiz : Set E Non-seiz : Sets A, B, C, D		**Subset 2** Seiz : Set E Non-seiz : Sets A, B		**Subset 3** Seiz : Set E Non-seiz : Set C	
	Recall	F1-score	Recall	F1-score	Recall	F1-score
CNN-6 layers	0.956	0.627	0.952	0.638	0.95	0.908
CNN + Transformer	0.904	**0.888**	0.896	**0.852**	0.893	**0.935**

Table 7. Performance of our two best architectures when trained on dataset #1 and evaluated on dataset #2.

4. Discussions and perspectives

The application of machine-learning algorithms, including deep learning neural networks, has witnessed significant progress in detecting seizure activity from electroencephalography (EEG) recordings over the past few decades. This advancement holds promise for enhancing clinical treatment outcomes and deepening our understanding of the underlying neurobiological mechanisms. Notably, since the early 90s, numerous studies have consistently demonstrated the ability of machine learning to identify seizures with high sensitivity, typically exceeding 95% (Jandó et al., 1993; Medvedev and Lehmann, 2024). However, these findings often have elevated rates of false positive detections. As a result, they lack generalization across subjects, highlighting the need for continued refinement of these approaches.

The increasing adoption of deep learning methods for raw EEG data analysis has been hindered by the limited size of many EEG datasets, which challenges training reliable models. Collecting large datasets can be a resource-intensive and time-consuming endeavor, often beyond the capabilities of smaller research centers. Before the recent trend of applying CNNs to raw electrophysiology data for automated feature extraction, the conventional approach involved manual feature creation and the application of machine learning or deep-learning methods with traditional explainability techniques (Chen et al., 2023; Ellis et al., 2023; Ince et al., 2008; Kwon et al., 2018; Ruffini et al., 2019). These manually extracted features typically captured time-domain and/or frequency-domain aspects of the data. Although effective, these approaches are inherently limited by the size of the feature space from which they can learn.

In contrast, deep-learning methods, like CNNs, can automatically collect features by learning robust embedding spaces, making them an attractive solution for raw EEG analysis (Oh et al., 2019; Shoeibi et al., 2021). In the present study, we leveraged the availability of a large dataset collected under comparable conditions (hardware and recording parameters) in an animal model of MTLE, enabling the use of CNNs and transformers paired with raw electrophysiological signals.

One of the key strengths of this study lies in its consideration of the potential for data leakage during dataset training. A common pitfall in EEG-based studies is the random assignment of segments to training and test sets, which results in data samples from individual subjects being parts of both sets. Such assignments can lead to data leakage, where EEG segments from a single subject appear in both the training and test sets, thus artificially inflating model performance. A recent study (Brookshire et al., 2024) highlighted the importance of addressing this issue by comparing the performance of deep neural network (DNN) classifiers using segment-based holdout (where segments from one subject can appear in both sets) versus subject-based holdout (where all segments from one subject are exclusive to either the training or test set). The authors demonstrated that segment-based holdout can lead to a significant overestimation of the model performance on previously unseen subjects. Alarmingly, they found that most translational DNN-EEG studies employ segment-based holdout, which may result in a dramatic overestimation of the model performance on new subjects (Rasheed et al., 2021; Shoeibi et al., 2021). To ensure an accurate assessment of our model, we designed a rigorous approach by exclusively including each subject's data to only the training or test sets, but never both.

All existing methods for seizure detection involve pre-processing by isolating seizure activities from non-seizure ones before segmenting the data into small blocks to constitute training and test sets. As described above, such a pre-processing technique generates an ideal scenario where small blocks obtained after segmentation contain either pure seizure or pure seizure-free activities. In this study, we have demonstrated that such a pre-processing pipeline overestimates seizure detection performance. To mimic real-world challenges related to automatic seizure detection, the developed pipelines must be conceptualized without any prior distinction between the seizure and seizure-free activities (refer to Pre-processing D method). Moreover, models must be primarily evaluated regarding their capacity to detect accurately the onset and offset of seizures on continuous EEG signals (refer to Post-processing D and evaluation D methods). Post-processing D algorithm allowed the reconstitution of continuous EEG signals out of overlapped segments following a temporal resolution of 500ms. It helped to increase the model’s precision and facilitate the onset/offset comparison between predicted and labelled data. The analysis of two distinct evaluation strategies (Evaluation C and Evaluation D) shed light on the fundamental differences between the classification and the detection tasks of seizures on EEG signals.

One of the most exciting aspects of this study is the robustness of our approach, which yields comparable performance metrics to some commercial systems (Koren et al., 2021), even when faced with modifications to the recording setup, changes in recording conditions, or differences in environment (experimental/clinical) or species (mouse/human) (Besné et al., 2022). This versatility underscores the potential of our approach to transcend traditional boundaries and

facilitate seamless translation between preclinical research and clinical applications. Our results demonstrate the capabilities of the proposed approach in both clinical and research environments, offering a valuable tool to aid experts in alleviating the burden of annotating extensive hours-long EEG recordings. Furthermore, the trans-species adaptability of our approach may facilitate a deeper understanding of the differences and similarities between human diseases and animal models. Notably, to the best of our knowledge, this study is one of the few to validate a high-performance detection algorithm for HPDs on comprehensive EEG datasets from both animal (mice MTLE dataset) and human (Bonn dataset) subjects. The detection performance of our proposed method suggests that this approach can be reliably applied in preclinical research and clinical settings, paving the way for future studies to explore its potential in real-world applications. Future work should focus on validating this framework using extended EEG data with diverse seizure types to evaluate its specificity and expand its applications. Additionally, it would be beneficial to validate our pipeline on other animal models to enable its broader use in preclinical research. Although our methodology shows promising results regarding generalization capabilities, it would be interesting to apply explainability methods to understand features involved in our model predictions. Our approach based on deep neural networks trained directly on raw EEG complexified such a goal. It exhibits one of the main disadvantages compared to pipelines based on models trained with extracted features where relatively simple studies can highlight features' contribution to seizure detection (Sturm et al., 2016).

Conclusion

In this work, we have introduced a novel seizure detection pipeline exhibiting pre-processing and post-processing techniques, which integrate real-world scenarios in EEG signal segmentation without prior knowledge of the labels and the predicted seizure evaluation. The experiments that we conducted showed that when considering a pipeline that does not incorporate these real cases, we end up with an overestimation of model performance. We implemented multiple CNN, RNN, segmentation, and transformer-based architectures on the seizure detection task and demonstrated the generalization of our best architecture combining a CNN and a transformer encoder. Trained with raw EEG signals from animals and tested on raw EEG from humans, the CNN+transformer model demonstrated high generalization capabilities with F1 score and recall higher than 0.85.

CRediT authorship contribution statement

Davy Darankoum: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. Manon Villalba: Investigation. Clélia Allioux: Investigation. Baptiste Caraballo: Investigation. Carine Dumont: Investigation. Eloïse Gronlier: Investigation. Corinne Roucard: Writing - Review & Editing, Project administration. Yann Roche: Writing - Review & Editing, Funding acquisition. Chloé Habermacher: Writing - Review & Editing,

Validation. Julien Volle: Writing - Review & Editing, Supervision, Conceptualization. Sergei Grudinin: Writing - Review & Editing, Supervision, Conceptualization.

Funding statement

This work was funded by SynapCell SAS through the Cortex project which has been awarded at the 9th edition of the i-Nov competition organized for French companies. The research has been conducted using both SynapCell and Laboratoire Jean Kuntzmann (CNRS/UGA) resources.

Declaration of Competing Interest

Davy Darankoum, Manon Villalba, Clélia Allioux, Baptiste Caraballo, Carine Dumont, Eloïse Gronlier, Corinne Roucard, Yann Roche, Chloé Habermacher, and Julien Volle are employees of SynapCell SAS. Sergei Grudinin have no conflicts of interest to declare.

Data availability statement

Dataset #1 can be made available to independent researchers after receipt of a valid research proposal, data analysis plan, and summary of researcher qualifications. Requests may be submitted to SynapCell at hello@synapcell.fr. Provision of data is contingent on business feasibility and execution of a data use agreement.

Dataset #2 is available at https://www.upf.edu/web/ntsa/downloads/-/asset_publisher/xvT6E4pczrBw/content/2001-indications-of-nonlinear-deterministic-and-finite-dimensional-structures-in-time-series-of-brain-electrical-activity-dependence-on-recording-regi.

References

1. Acharya, U.R., Oh, S.L., Hagiwara, Y., Tan, J.H., Adeli, H., 2018. Deep convolutional neural network for the automated detection and diagnosis of seizure using EEG signals. Comput. Biol. Med. 100, 270–278. doi:10.1016/j.compbiomed.2017.09.017

2. Ammothumkandy, A., Ravina, K., Wolseley, V., Tartt, A.N., Yu, P.-N., Corona, L., Zhang, N., Nune, G., Kalayjian, L., Mann, J.J., Rosoklija, G.B., Arango, V., Dwork, A.J., Lee, B., Smith, J.A.D., Song, D., Berger, T.W., Heck, C., Chow, R.H., Boldrini,

M., Liu, C.Y., Russin, J.J., Bonaguidi, M.A., 2022. Altered adult neurogenesis and gliogenesis in patients with mesial temporal lobe epilepsy. Nat. Neurosci. 25 4 , 493–503. doi:10.1038/s41593-022-01044-2

3. Andrzejak, R.G., Lehnertz, K., Mormann, F., Rieke, C., David, P., Elger, C.E., 2001. Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state. Phys. Rev. E 64 6 , 061907. doi:10.1103/PhysRevE.64.061907

4. Beghi, E., 2020. The EpidemiologVency of Epilepsy. Neuroepidemiology 54 2 , 185–191. doi:10.1159/000503831

5. Besné, G.M., Horrillo-Maysonnial, A., Nicolás, M.J., Capell-Pascual, F., Urrestarazu, E., Artieda, J., Valencia, M., 2022. An interactive framework for the detection of ictal and interictal activities: Cross-species and stand-alone implementation. Comput. Methods Programs Biomed. 218, 106728. doi:10.1016/j.cmpb.2022.106728

6. Brookshire, G., Kasper, J., Blauch, N.M., Wu, Y.C., Glatt, R., Merrill, D.A., Gerrol, S., Yoder, K.J., Quirk, C., Lucero, C., 2024. Data leakage in deep learning studies of translational EEG. Front. Neurosci. 18, 1373515. doi:10.3389/fnins.2024.1373515

7. Chen, W., Wang, Y., Ren, Y., Jiang, H., Du, G., Zhang, J., Li, J., 2023. An automated detection of epileptic seizures EEG using CNN classifier based on feature fusion with high accuracy. BMC Med. Inform. Decis. Mak. 23 1 , 96. doi:10.1186/s12911-023-02180-w

8. Cho, K., van Merrienboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., Bengio, Y., 2014. Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation.

9. Tatum, W.O., 2012. Mesial Temporal Lobe Epilepsy. J. Clin. Neurophysiol. 29 5 .

10. Duveau, V., Roucard, C., 2017. A Mesiotemporal Lobe Epilepsy Mouse Model. Neurochem. Res. 42 7 , 1919–1925. doi:10.1007/s11064-017-2239-3

11. Cho, K.-O., Jang, H.-J., 2020. Comparison of different input modalities and network structures for deep learning-based seizure detection. Sci. Rep. 10 1 , 122. doi:10.1038/s41598-019-56958-y

12. Choi, G., Park, C., Kim, J., Cho, K., Kim, T.-J., Bae, H., Min, K., Jung, K.-Y., Chong, J., 2019. A Novel Multi-scale 3D CNN with Deep Neural Network for Epileptic Seizure Detection, in: 2019 IEEE International Conference on Consumer Electronics (ICCE). Presented at the 2019 IEEE International Conference on Consumer Electronics (ICCE), IEEE, Las Vegas, NV, USA, pp. 1–2. doi:10.1109/ICCE.2019.8661969

13. Durongbhan, P., Zhao, Yifan, Chen, L., Zis, P., De Marco, M., Unwin, Z.C., Venneri, A., He, X., Li, S., Zhao, Yitian, Blackburn, D.J., Sarrigiannis, P.G., 2019. A Dementia Classification Framework Using Frequency and Time-Frequency Features Based on EEG Signals. IEEE Trans. Neural Syst. Rehabil. Eng. 27 5 , 826–835. doi:10.1109/TNSRE.2019.2909100

14. Ebrahim, S.A., Poshtan, J., Jamali, S.M., Ebrahim, N.A., 2020. Quantitative and Qualitative Analysis of Time-Series Classification Using Deep Learning. IEEE Access 8, 90202–90215. doi:10.1109/ACCESS.2020.2993538

15. Eldele, E., Chen, Z., Liu, C., Wu, M., Kwoh, C.-K., Li, X., Guan, C., 2021. An Attention-Based Deep Learning Approach for Sleep Stage Classification With Single-Channel EEG. IEEE Trans. Neural Syst. Rehabil. Eng. 29, 809–818. doi:10.1109/TNSRE.2021.3076234

16. Ellis, C.A., Miller, R.L., Calhoun, V.D., 2023. Evaluating Augmentation Approaches for Deep Learning-based Major Depressive Disorder Diagnosis with Raw Electroencephalogram Data * (preprint). Neuroscience. doi:10.1101/2023.12.15.571938

17. Guo, L., Rivero, D., Dorado, J., Rabuñal, J.R., Pazos, A., 2010. Automatic epileptic seizure detection in EEGs based on line length feature and artificial neural networks. J. Neurosci. Methods 191 1 , 101–109. doi:10.1016/j.jneumeth.2010.05.020

18. Hussain, W., Iqbal, M.S., Xiang, J., Wang, B., Niu, Y., Gao, Y., Wang, X., Sun, J., Zhan, Q., Cao, R., Mengni, Z., 2019. Epileptic Seizure Detection With Permutation Fuzzy Entropy Using Robust Machine Learning Techniques. IEEE Access 7, 182238–182258. doi:10.1109/ACCESS.2019.2956865

19. Ince, N.F., Goksu, F., Pellizzer, G., Tewfik, A., Stephane, M., 2008. Selection of spectro-temporal patterns in multichannel MEG with support vector machines for schizophrenia classification, in: 2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. Presented at the 2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, IEEE, Vancouver, BC, pp. 3554–3557. doi:10.1109/IEMBS.2008.4649973

20. Jandó, G., Siegel, R.M., Horváth, Z., Buzsáki, G., 1993. Pattern recognition of the electroencephalogram by artificial neural networks. Electroencephalogr. Clin. Neurophysiol. 86 2 , 100–109. doi:10.1016/0013-4694(93)90082-7

21. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S.A.A., Ballard, A.J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., Back, T., Petersen, S., Reiman, D., Clancy, E., Zielinski, M., Steinegger, M., Pacholska, M., Berghammer, T., Bodenstein, S., Silver, D., Vinyals, O., Senior, A.W., Kavukcuoglu, K., Kohli, P., Hassabis, D., 2021. Highly accurate protein structure prediction with AlphaFold. Nature 596 7873 , 583–589. doi:10.1038/s41586-021-03819-2

22. Koren, J., Hafner, S., Feigl, M., Baumgartner, C., 2021. Systematic analysis and comparison of commercial seizure-detection software. Epilepsia 62 2 , 426–438. doi:10.1111/epi.16812

23. Kwon, Y.-H., Shin, S.-B., Kim, S.-D., 2018. Electroencephalography Based Fusion Two-Dimensional (2D)-Convolution Neural Networks (CNN) Model for Emotion Recognition System. Sensors 18 5 , 1383. doi:10.3390/s18051383

24. LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521 7553 , 436–444. doi:10.1038/nature14539

25. Mahjoub, C., Le Bouquin Jeannès, R., Lajnef, T., Kachouri, A., 2020. Epileptic seizure detection on EEG signals using machine learning techniques and advanced preprocessing methods. Biomed. Eng. Biomed. Tech. 65 1 , 33–50. doi:10.1515/bmt-2019-0001

26. Medvedev, A.V., Lehmann, B., 2024. The detection of absence seizures using cross-frequency coupling analysis with a deep learning network. doi:10.21203/rs.3.rs-4178484/v1

27. Mesaros, A., Heittola, T., Virtanen, T., 2016. Metrics for Polyphonic Sound Event Detection. Appl. Sci. 6 6 , 162. doi:10.3390/app6060162

28. Mesraoua, B., Deleu, D., Al Hail, H., Melikyan, G., Boon, P., Haider, H.A., Asadi-Pooya, A.A., 2019. Electroencephalography in epilepsy: look for what could be beyond the visual inspection. Neurol. Sci. 40 11 , 2287–2291. doi:10.1007/s10072-019-04026-8

29. Min, S., Lee, B., Yoon, S., 2016. Deep learning in bioinformatics. Brief. Bioinform. bbw068. doi:10.1093/bib/bbw068

30. Mursalin, M., Zhang, Y., Chen, Y., Chawla, N.V., 2017. Automated epileptic seizure detection using improved correlation-based feature selection with random forest classifier. Neurocomputing 241, 204–214. doi:10.1016/j.neucom.2017.02.053

31. Nabbout, R., Kuchenbuch, M., 2020. Impact of predictive, preventive and precision medicine strategies in epilepsy. Nat. Rev. Neurol. 16 12 , 674–688. doi:10.1038/s41582-020-0409-4

32. Noachtar, S., Rémi, J., 2009. The role of EEG in epilepsy: A critical review. Epilepsy Behav. 15 1 , 22–33. doi:10.1016/j.yebeh.2009.02.035

33. Obeid, I., Picone, J., 2016. The Temple University Hospital EEG Data Corpus. Front. Neurosci. 10. doi:10.3389/fnins.2016.00196

34. Oh, S.L., Vicnesh, J., Ciaccio, E.J., Yuvaraj, R., Acharya, U.R., 2019. Deep Convolutional Neural Network Model for Automated Diagnosis of Schizophrenia Using EEG Signals. Appl. Sci. 9 14 , 2870. doi:10.3390/app9142870

35. Paschen, E., Elgueta, C., Heining, K., Vieira, D.M., Kleis, P., Orcinha, C., Häussler, U., Bartos, M., Egert, U., Janz, P., Haas, C.A., 2020. Hippocampal low-frequency stimulation prevents seizure generation in a mouse model of mesial temporal lobe epilepsy. eLife 9, e54518. doi:10.7554/eLife.54518

36. Perslev, M., Jensen, M., Darkner, S., Jennum, P.J., Igel, C., 2019. U-time: A fully convolutional network for time series segmentation applied to sleep staging. Adv. Neural Inf. Process. Syst. 32.

37. Rasheed, K., Qayyum, A., Qadir, J., Sivathamboo, S., Kwan, P., Kuhlmann, L., O'Brien, T., Razi, A., 2021. Machine Learning for Predicting Epileptic Seizures Using

EEG Signals: A Review. IEEE Rev. Biomed. Eng. 14, 139–155. doi:10.1109/RBME.2020.3008792

38. Ronneberger, O., Fischer, P., Brox, T., 2015. U-Net: Convolutional Networks for Biomedical Image Segmentation, in: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (Eds.), Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015, Lecture Notes in Computer Science. Springer International Publishing, Cham, pp. 234–241. doi:10.1007/978-3-319-24574-4_28

39. Roy, S., Kiral-Kornek, I., Harrer, S., 2018. Deep Learning Enabled Automatic Abnormal EEG Identification, in: 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). Presented at the 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), IEEE, Honolulu, HI, pp. 2756–2759. doi:10.1109/EMBC.2018.8512756

40. Ruffini, G., Ibañez, D., Castellano, M., Dubreuil-Vall, L., Soria-Frisch, A., Postuma, R., Gagnon, J.-F., Montplaisir, J., 2019. Deep Learning With EEG Spectrograms in Rapid Eye Movement Behavior Disorder. Front. Neurol. 10, 806. doi:10.3389/fneur.2019.00806

41. Sak, H., Senior, A., Beaufays, F., 2014. Long Short-Term Memory Based Recurrent Neural Network Architectures for Large Vocabulary Speech Recognition.

42. Shoeb, A., Guttag, J., 2010. Application of Machine Learning To Epileptic Seizure Detection.

43. Shoeibi, A., Sadeghi, D., Moridian, P., Ghassemi, N., Heras, J., Alizadehsani, R., Khadem, A., Kong, Y., Nahavandi, S., Zhang, Y.-D., Gorriz, J.M., 2021. Automatic Diagnosis of Schizophrenia in EEG Signals Using CNN-LSTM Models. Front. Neuroinformatics 15, 777977. doi:10.3389/fninf.2021.777977

44. Shoji, T., Yoshida, N., Tanaka, T., 2021. Automated detection of abnormalities from an EEG recording of epilepsy patients with a compact convolutional neural network. Biomed. Signal Process. Control 70, 103013. doi:10.1016/j.bspc.2021.103013

45. Singh, A.K., Krishnan, S., 2023. Trends in EEG signal feature extraction applications. Front. Artif. Intell. 5, 1072801. doi:10.3389/frai.2022.1072801

46. Sturm, I., Lapuschkin, S., Samek, W., Müller, K.-R., 2016. Interpretable deep neural networks for single-trial EEG classification. J. Neurosci. Methods 274, 141–145. doi:10.1016/j.jneumeth.2016.10.008

47. Su, J., Lu, Y., Pan, S., Murtadha, A., Wen, B., Liu, Y., 2023. RoFormer: Enhanced Transformer with Rotary Position Embedding.

48. Thijs, R.D., Surges, R., O'Brien, T.J., Sander, J.W., 2019. Epilepsy in adults. The Lancet 393 10172 , 689–701. doi:10.1016/S0140-6736(18)32596-0

49. Toraman, S., 2021. Automatic recognition of preictal and interictal EEG signals using 1D-capsule networks. Comput. Electr. Eng. 91, 107033. doi:10.1016/j.compeleceng.2021.107033

50. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., Polosukhin, I., 2023. Attention Is All You Need.

51. Wang, H., Shi, W., Choy, C.-S., 2018. Hardware Design of Real Time Epileptic Seizure Detection Based on STFT and SVM. IEEE Access 6, 67277–67290. doi:10.1109/ACCESS.2018.2870883