Computer Methods and Programs in Biomedicine 218 (2022) 106728

Contents lists available at ScienceDirect

Computer Methods and Programs in Biomedicine

journal homepage: www.elsevier.com/locate/cmpb

An interactive framework for the detection of ictal and interictal activities: Cross-species and stand-alone implementation

Guillermo M. Besné[a], Alejandro Horrillo-Maysonnial[b], María Jesús Nicolás[a], Ferran Capell-Pascual[a], Elena Urrestarazu[b,c], Julio Artieda[a,c], Miguel Valencia[a,c,d,*]

[a] Program of Neuroscience, Universidad de Navarra, CIMA, Avenida Pío XII, 55, 31008 Navarra, Pamplona, Spain
[b] Clinical Neurophysiology Section, Clínica Universidad de Navarra, Pamplona, Spain
[c] IdiSNA, Instituto de Investigación Sanitaria de Navarra, Pamplona, Spain
[d] Institute of Data Science and Artificial Intelligence, Universidad de Navarra, Pamplona, Spain

ARTICLE INFO

Article history:
Received 9 August 2021
Revised 3 February 2022
Accepted 1 March 2022

Keywords:
Machine learning
Epilepsy
Dravet
Seizures
Interictal
Stand-alone

ABSTRACT

Background and objective: Despite advances on signal analysis and artificial intelligence, visual inspection is the gold standard in event detection on electroencephalographic recordings. This process requires much time of clinical experts on both annotating and training new experts for this same task. In scenarios where epilepsy is considered, the need for automatic tools is more prominent, as both seizures and interictal events can occur on hours- or days-long recordings. Although other solutions have already been proposed, most of them are not integrated on clinical and basic science environments due to their complexity and required specialization. Here we present a pipeline that arises from coordinated efforts between life-science researchers, clinicians and data scientists to develop an interactive and iterative workflow to train machine-learning tools for the automatic detection of electroencephalographic events in a variety of scenarios.

Methods: The approach consists on a series of subsequent steps covering data loading and configuration, event annotation, model training/re-training and event detection. With slight modifications, the combination of these blocks can cope with a variety of scenarios. To illustrate the flexibility and robustness of the approach, three datasets from clinical (patients of Dravet Syndrome) and basic research environments (mice model of the same disease) were evaluated. From them, and in response to researchers' daily needs, four real world examples of interictal event detection and seizure classification tasks were selected and processed.

Results: Results show that the current approach was of great aid for event annotation and model development. It was capable of creating custom machine-learning solutions for each scenario with slight adjustments on the analysis protocol, easily accessible to users without programming skills. Final annotator similarity metrics reached values above 80% on all cases of use, reaching 92.3% on interictal event detection on human recordings.

Conclusions: The presented framework is easily adaptable to multiple real world scenarios and the interactive and ease-to-use approach makes it manageable to clinical and basic researches without programming skills. Nevertheless, it is conceived so data scientists can optimize it for specific scenarios, improving the knowledge transfer between these fields.

1. Introduction

Epilepsy is a chronic neurologic condition marked by the recurrence of unprovoked seizures. Seizures are paroxysmal events that result from abnormal neuronal discharges and manifest in the form of sudden, stereotyped episodes with accompanying changes in motor activity, sensation, and behavior that expose the patient to life-threatening situations. Epilepsy being among the most common neurological disorders, much research has been centered on studying, detecting, and predicting the nature of these events [1,2].

One of the main clinical tools for the diagnosis and follow-up of patients with epilepsy is the electroencephalogram (EEG),

* Corresponding author at: Program of Neuroscience, Universidad de Navarra, CIMA, Avenida Pío XII, 55, 31008 Navarra, Pamplona, Spain.
E-mail address: mvustarroz@unav.es (M. Valencia).

https://doi.org/10.1016/j.cmpb.2022.106728

that records the electrical activity generated by the brain [3]. During seizures, the pattern of normal brain activity changes and EEG shows aberrant activities in the form of ictal discharges. Between seizures other pathological distinctive waves and complexes distinguished from background [4,5] also occur. These activities, designated as interictal epileptiform discharges (IEDs) have been recorded both on patients and animal models of epilepsy as well as other pathologies such as Alzheimer's disease [6].

Visual analysis is still the gold standard in clinical EEG for the detection of seizures and IEDs. Despite its demonstrated reliability and worldwide consensus, this process depends on the expertise of the physician and requires a long assisted learning process [7–9]. As described on the bibliography, this visual analysis and event recognition is time consuming; especially when recordings last several hours (even days) and the agreement between annotators is not assured [10]. Several approximations to ease IED and seizure annotation have been proposed. Some of them are focused on analyzing scalp EEG from human beings [11–17], but there are also developments aimed at detecting seizures and interictal activities in non-human recordings [18–21]. However, these solutions are not implemented as user-friendly integrated tools and often require a certain level of programming and data analysis skills. As a result, they are only partially adopted and do not result suitable for most basic and clinical research environments.

Nevertheless, clinical practice and preclinical research require widely accepted analytical tools to assess and followup the severity of the disease and the effect of therapeutic interventions. Here, we propose an intuitive, fast, reliable, and easy-to-use framework to detect ictal and interictal events in long EEG recordings. This framework arises from coordinated efforts between life-science researchers, clinicians, and data scientists. It has been implemented in an integrated, stand-alone tool developed in MATLAB and built as an interactive and iterative loop that includes event annotation, model training and test cycles, so customized solutions are easily generated.

In a recent work, we characterized a mice model of Dravet syndrome (DS) and observed that DS –but not wild-type- mice showed spontaneous IEDs on electrophysiological recordings over multiple structures [22]. DS is an early-onset epileptic encephalopathy [23] characterized by febrile seizures, high risk of sudden unexpected death in epilepsy (SUDEP), psychomotor delay and treatment refractoriness [24–26]. To illustrate our framework, we use multisite electrophysiological recordings in this mice model and show that the proposed approach is not only robust for the detection of IEDs along different electrode configurations, but it is also able to detect seizures and to classify ongoing activity into pre-ictal, ictal and post-ictal episodes. Finally, we further demonstrate the translational ability of our approach by detecting IEDs in clinical EEG recordings performed on patients with DS.

2. Material and methods

2.1. Proposed framework

Fig. 1 summarizes the proposed interactive framework. It is aimed at building customized ML models for the detection of ictal and interictal activities on electrophysiological recordings. First, a data-loading step is initiated; files are selected and certain information is requested from the user (channels to use, training time window, prediction time window, etc.). Afterwards, an interactive annotation tool is provided to annotate event times with or without assistance (providing/not providing candidates for model training). With enough events, a new model can be generated or, alternatively, already available models can be re-trained. This procedure generates a performance report and a model file for future use. Finally, the event detection process can be run on the whole recordings, with the possibility of including additional files. Across the whole process, information about the loaded files (configuration, marked events, etc.) is stored and can be exported for further iterative use.

All these functionalities have been integrated in an interface using MATLAB GUI generator. At the current stage, Spike2 files (format .smr or .smrx) are required for the use of the app. However, interfaces for other file formats such as .txt and .edf could be easily integrated in future versions. The code was implemented following a modular structure with independent functions for each functionality: file management, configuration management, event marking for experts, model training and full-recording event detection. Each functionality is called upon by a main function (main_app.m) that manages all the different modules. These capabilities are further explained in the following section by their use on multiple scenarios and datasets.

2.2. Datasets

The functionality and performance of the proposed framework have been assessed by analyzing three datasets coming from two different clinical and experimental scenarios. All the procedures were performed in accordance with the protocols approved by the ethical committee (CEEA-021–21 and 059–16) and following institutional guidelines.

The first dataset (*dMice1*) consists of seven electrophysiological recordings from four DS mice. Animals were implanted with three multi-electrode probes (50 µm tungsten electrodes) aiming at the hippocampus, hypothalamus and prefrontal cortex: three electrodes in CA1, two in the dentate gyrus (Gd), two in the preoptic area (PoA) and one in the prefrontal cortex (PFC). Animals were recorded under a thermal challenge where a heating chamber gradually increased the ambient temperature. Animals were placed in the heating chamber and recorded during 15 min at room temperature. After this period, the temperature was gradually increased until a seizure was observed (electrically and behaviorally). Examples of IED and non-IED activities (also including muscle artifacts and other non-IED spiky transients) were annotated by two independent experts and later unified. IEDs were manually annotated in time windows of 2 s, then the precise timing was shifted to the sharpest point of the corresponding waveform (Fig. 2, A). Seizure classification was based on the mice Racine Scale [27]. These mice model shows multifocal seizures, with different origins within and across individuals, and intra/inter-hemispheric propagation that often generalize to tonic-clonic seizures. They show clustered patterns of generalized, high amplitude-low frequency activity with altering low amplitude waveforms. Accordingly, four distinct classes for pre-critical (*PreSeiz*, Fig. 2, B, white), seizure beginning (*Seiz1*, dark blue), seizure ending (*Seiz2*, green), and inter-critical (*InterSeiz*, light blue) stages were defined.

For robustness, a second set of nine recordings from nine DS mice was analyzed (*dMice2*). In this set, animals were implanted with a different electrode configuration: three tungsten electrodes in the external globus pallidus and one steel screw targeting the PFC (0.95 mm ø). In this case, animals were recorded for 15 min in an openfield arena (freely moving conditions in a 40 $\times$ 40 cm box).

Finally, performance on human data was assessed on two routine EEG recordings from two patients with DS (*dHuman*). Twelve to thirteen hour-long recordings were composed of 18 valid EEG channels distributed according to the 10:20 international system. Interictal activities were annotated by a clinical expert that reviewed EEG recordings on windows of 15 s and placed markers identifiable IEDs. Extra markers as reference of non-IEDs (free of epileptiform activity) were also included, reaching at least a 50:50

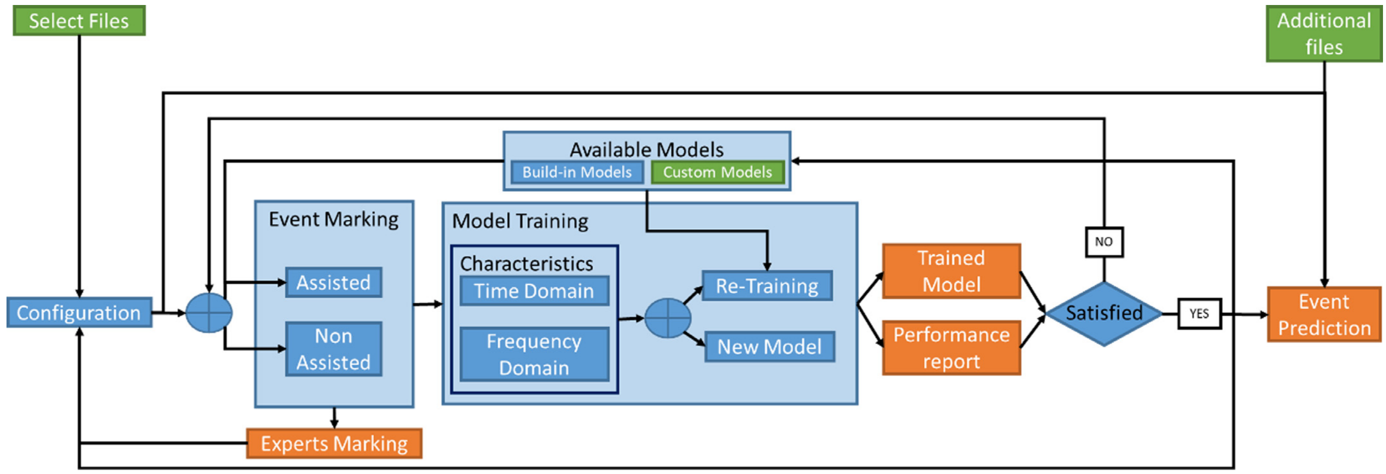


Fig. 1. Workflow representing all available pathways on this app where the colored boxes represent inputs (green), processes and decisions (blue), and outputs (orange).

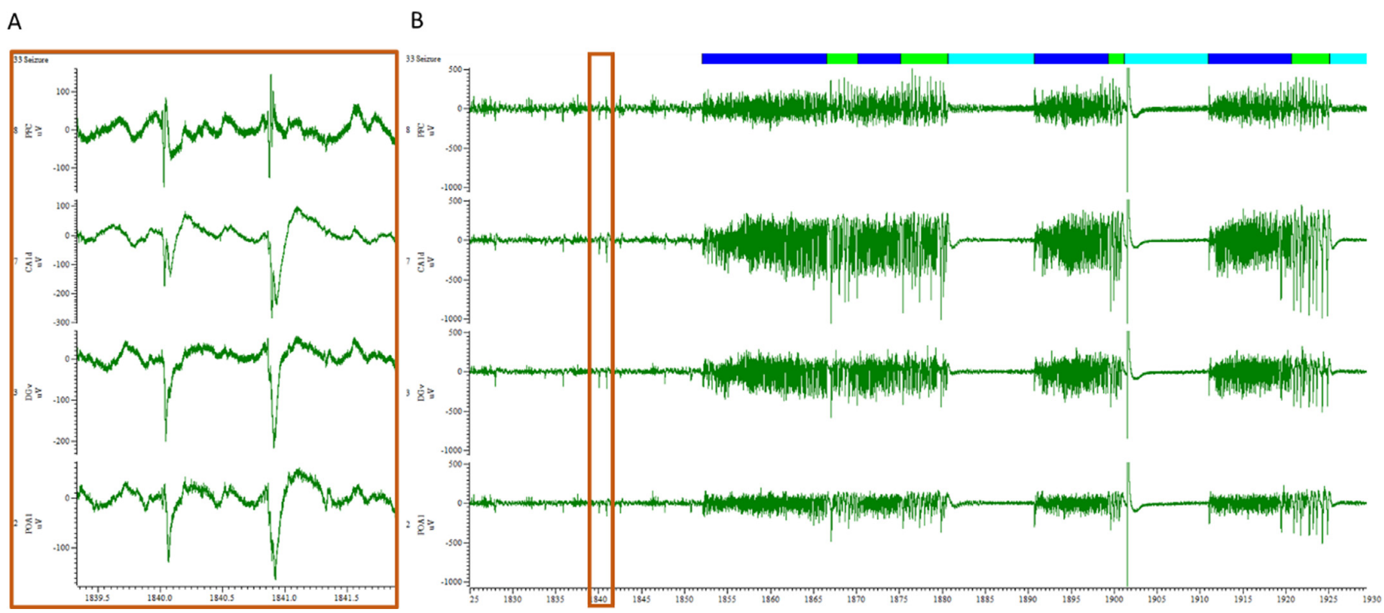


Fig. 2. (A) Example of two generalized IEDs recorded during the preictal phase of the recording shown in (B). (B) Example of a DS mouse recording showing each of the four stages considered for seizure detection: White, PreSeiz; Dark blue, Seiz1; Green, Seiz2; Light blue, InterSeiz. The orange rectangle defines the preictal segment zoomed in A.

ratio with IEDs. For each EEG channel, an individual marker channel was created containing both IED and non-IED events.

2.3. Model generation

2.3.1. Annotation for marker definition

The annotation process is implemented through a dedicated module that allows to select the files to analyze, channels of interest, and the temporal window for review and event processes (data_prep.m)

Depending on user preference two distinctive approaches are available. If the user wants to obtain IEDs candidates without directly initiating an annotation stage, the system calls a routine (event_marking_assist.m) that uses available models to detect IED candidates. Those IED candidates and twice of non-IEDs are presented to the user (Fig. 3, A) and confirms/rejects the nature of such events. If the user prefers to proceed without this assistance, an interactive interface is called (event_marking_nonassist.m) to manually annotate the presence and absence of IEDs (Fig. 3, B). Once finished, the annotations are stored as part of the session configuration.

Information about session configuration and experts markings can be exported as a Matlab's *.mat file for later usage. This functionality allows to combine the information from different experts (and annotation sessions) to be integrated in an iterative and cooperative multiuser/multisession scheme.

2.3.2. Characteristics estimation

Prior to the model training, key characteristics must be identified with both computational ease and clinical relevance in mind. Based on visual inspection, previous studies on animal models and clinical experts' experience, nine characteristics within time and frequency domains were selected (Fig. 4, D).

For time-domain characteristics (Fig. 4, A), five parameters have been selected: kurtosis, skewness, highest peaks (or lowest valleys) prominence and width, and dynamic range. To obtain frequency-based characteristics, first Gabor transform was applied to each marked IEDs on 400 ms windows and most relevant frequency ranges for IEDs waveform decomposition were identified (Fig. 4,

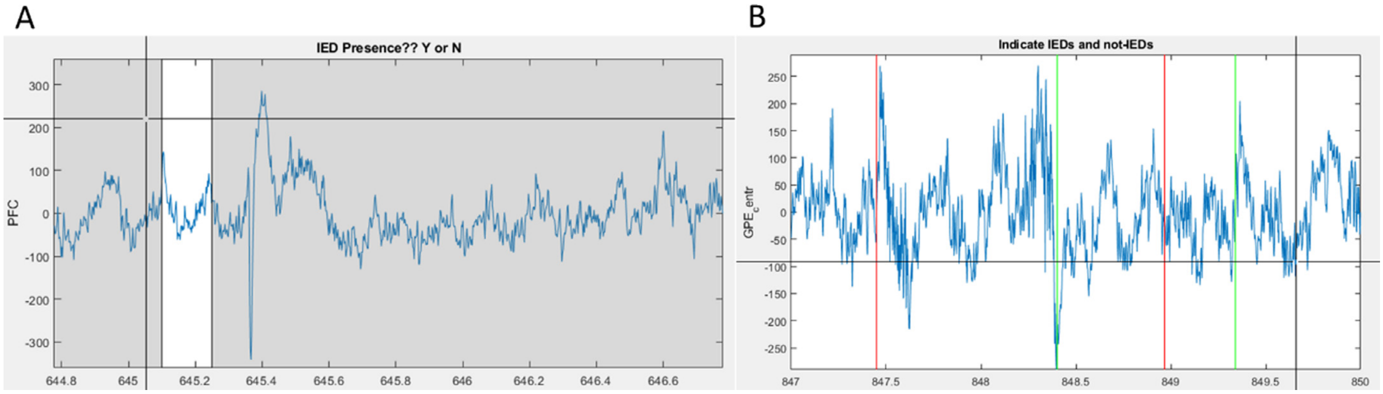


Fig. 3. Event marking interface: (A) Assisted annotation: event candidates (within the white window) are proposed and the user determines the presence or absence of such events. (B) Non-assisted annotation: a window with fragments of the raw signal is presented and the user uses the mouse to define approximately the timing of events of interest (green vertical lines) or the absence of such events (red vertical lines).

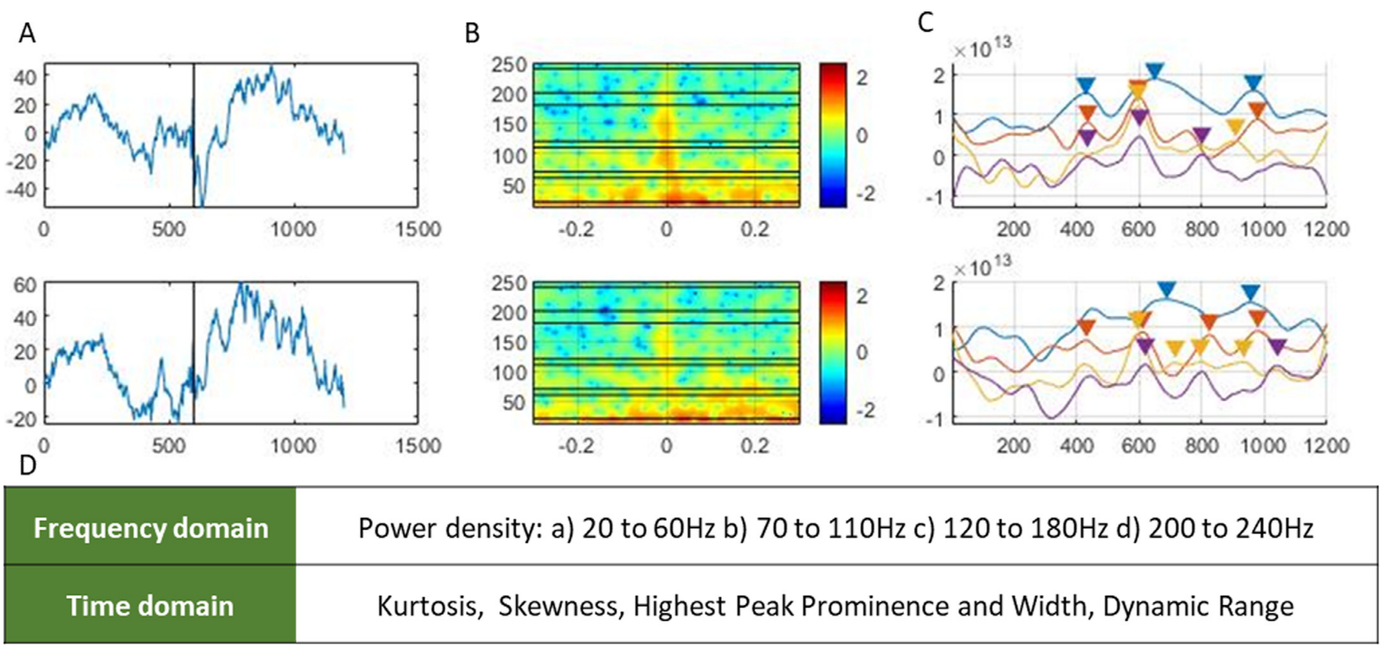


Fig. 4. (A) Data segments on two different channels show an IED on the middle; (B) Time frequency representation of the signal segment showing frequencies from 10 Hz to 250 Hz; (C) Envelope of the four frequency bands of interest extracted from the signal segments on (A); (D) Summary of the characteristics for each category.

B). Visual inspection of the logarithmic representation shows four main frequency bands of activity burst between 20 and 240 Hz (Fig. 4, B and C). In the case of human recordings, good performance was obtained by modifying the 20–60 Hz frequency range to 10–30 Hz. Then, the characteristics derived from these activations were calculated from the envelope of the normalized signal on each frequency band. All this processes are implemented in an independent and dedicated function (IED_charact.m).

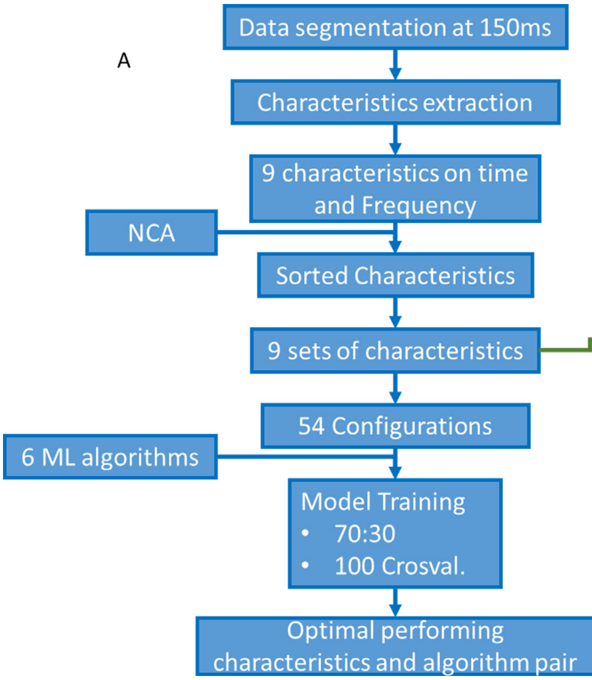
Prior to the characteristic estimation, signals were normalized (z-scored) with the mean and standard deviation estimated from the first five minutes of the recording. Different time windows for characteristics estimation were evaluated; for IED detection, results showed optimal performance for values of 150 ms (without overlap) whereas for seizures, windows of 2 s with 1 s of overlap gave the best separation.

2.3.3. Model selection and performance assessment

Multiple ML models have been developed and improved for a variety of applications. MATLAB offers a convenient environment for prototyping such solutions as it includes several of these models. For this application, six ML algorithms were selected and integrated on a dedicated function (model_train.m) that implements the screening process to establish the optimal combination of characteristics and ML model (Fig. 5A).

Markers from the annotated registers were used to select the data segments to estimate the characteristics that serve to train the ML model. In the absence of non-IED samples random timestamps (without IEDs) can be used. As non-IED fragments include a variety of activities that result in a greater variability than IED samples (i.e., there is a larger similarity within IED samples than within non-IED examples), we suggest generating datasets with 20% IEDs and 80% non-IEDs. Neighbor component analysis (NCA) is used over to sort the characteristics from the most to the least relevant, thus generating nine sets of characteristics. Each set is composed of a different number of parameters: from 1 for the single most relevant feature to 9 for the full set of nine features, where parameters are added according to their relevance in the NCA analysis (Fig. 5B).

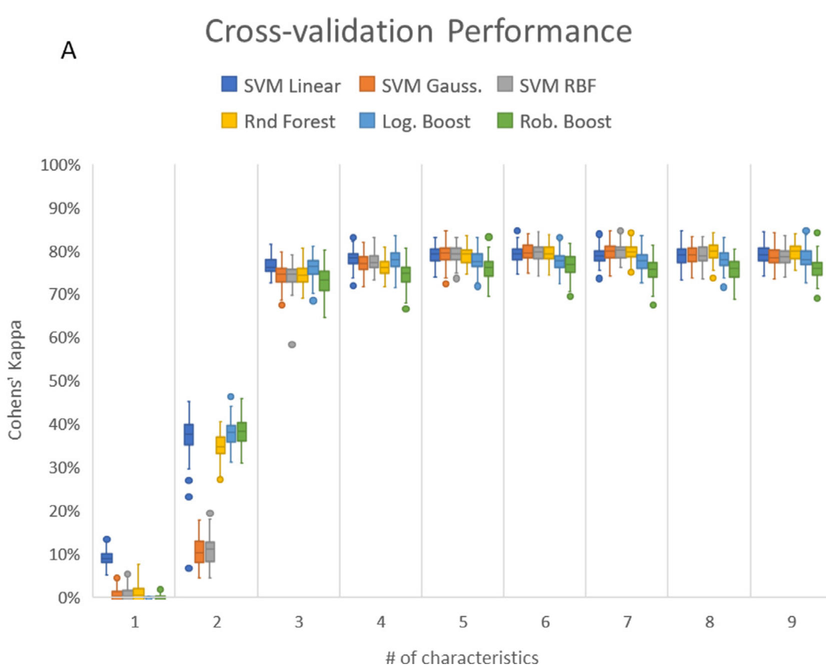
Supported vector machine (SVM) with Linear, Gaussian, and Radial Based Function (RBF) kernels, Random Forest, and Tree En-



B

	Set 1	Set 2	Set 3	Set 4	Set 5	Set 6	Set 7	Set 8	Set 9
Char. 1	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Char. 2	-	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Char. 3	-	-	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Char. 4	-	-	-	Yes	Yes	Yes	Yes	Yes	Yes
Char. 5	-	-	-	-	Yes	Yes	Yes	Yes	Yes
Char. 6	-	-	-	-	-	Yes	Yes	Yes	Yes
Char. 7	-	-	-	-	-	-	Yes	Yes	Yes
Char. 8	-	-	-	-	-	-	-	Yes	Yes
Char. 9	-	-	-	-	-	-	-	-	Yes

Fig. 5. (A) Workflow showing the process of training the model from data extraction to model selection; (B) Characteristics included on each set for training.



B

Optimal model testing performance						
ML type	# Charac.	Pre	TPR	TNR	Acc	Kpp
RBF	6	88.5%	85.5%	97.2%	94.9%	83.8%

C

	Re-training performance				
	Pre	TPR	TNR	Acc	Kpp
Build-in	80.9%	95.3%	95.3%	95.3%	84.7%
Retrain	87.3%	92.3%	97.2%	96.3%	87.5%

Fig. 6. ML model training summary for *dMice1_PreSeiz* mice dataset. (A) Boxplot showing the validation performance after the 100-fold cross-validation for each ML algorithm and the number of characteristics. (B) Indicating the selected model after the validation and the corresponding post-testing performance of the initial model. (C) Performance uplift after the initial model after the re-training.

semble with a Robust and Logistic Boost have been tested for the model training. This results in 54 model configurations by pairing characteristic set and algorithm. The performance of each configuration is tested in a data set according to a 70:30 train:test ratio with 100-fold cross-validation, where the training set is randomly divided according to a 70:30 train:validation ratio.

Optimal model configuration is selected by determining a set of characteristics and a machine learning model pair with the best mean Cohen's Kappa (Kpp) [28] (also considering the standard deviation). With the reaming 30% of the data, final testing is accomplished and performance values are extracted: accuracy (Acc), predictability (Pre), true positive ratio (TPR), true negative ratio (TNR) and Kpp.

Whereas the IED classification dataset is composed of two classes (IED/non-IED) the seizure data is composed of four different classes. This fact precludes the use of some of the ML algorithms. Accordingly, Robust and Logistic Boost for Tree Ensemble were replaced by Random Under-Sampling (RUS) Boost and Total Boost. Cohen's Kappa was interchanged by Scotts' Pi and global accuracy and Acc, Pre, TPR, TNR and F1 scores for each class were calculated.

	Re-training performance				
	Pre	TPR	TNR	Acc	Kpp
Build-in	77.3%	95.8%	91.9%	92.8%	80.8%
Retrain	82.2%	97.4%	94.0%	94.7%	85.7%

Fig. 7. Performance uplift after of the final model obtained on the Case of use 1 re-trained with data from dMice2.

Once finished, both performance values and ML model are automatically stored on a *.txt text file and a Matlab *.mat file respectively. This ML model file can be used to have a starting point for further analysis as a template for the assisted event marking (event_marking_assist.m), full-file event detection (fullfile_event_detect.m), model re-training with new events (model_retrain.m) or the exchange of the results.

3. Results

The proposed framework and their implementation takes into consideration principles of ease-of-use and rigor to cope with specific, *real-world* problems of biomedical research. To illustrate it, four different scenarios evoked from daily needs of life-science researchers and clinicians were explored.

3.1. Case of use 1: de-novo IED annotation in mice recordings

This case of use represents the most complete example of workflow use applied to a specific dataset. The user starts without any previous models or preliminary results, so it is necessary to perform an initial expert level annotation on a first file, create an initial model and use the remaining files to refine it.

It is illustrated by analyzing preictal data from the *dMice1* dataset (*dMice1_PreSeiz*). After expert's annotation, initial model training and cross-validation performance of each pair of characteristics/model configuration is carried out (Fig. 6, A). Mean Kpp rises rapidly as new characteristics are added, obtaining optimal performance with RDF and 6 characteristics. Testing performance reaches following values for Acc, Pre, TPR, TRN, Kpp: 88.5%, 88.5%, 97.2%, 94.9% and 83.8% (Fig. 6, B).

Next, the remaining six recordings from the *dMice1_PreSeiz* dataset are loaded. The model from the previous step is applied for an assisted annotation followed by re-training for further refinement. Two set of performance parameters are obtained from orig-

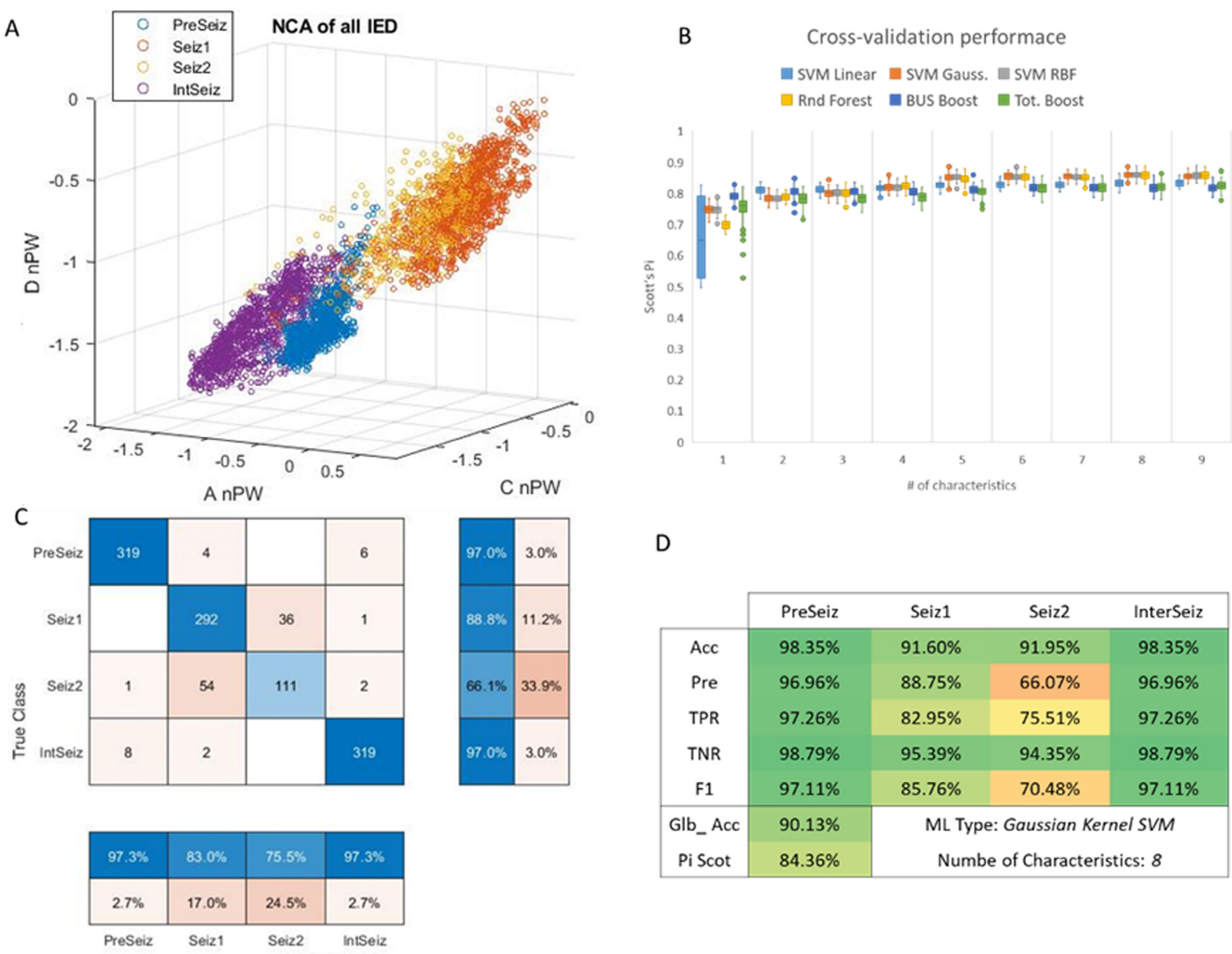
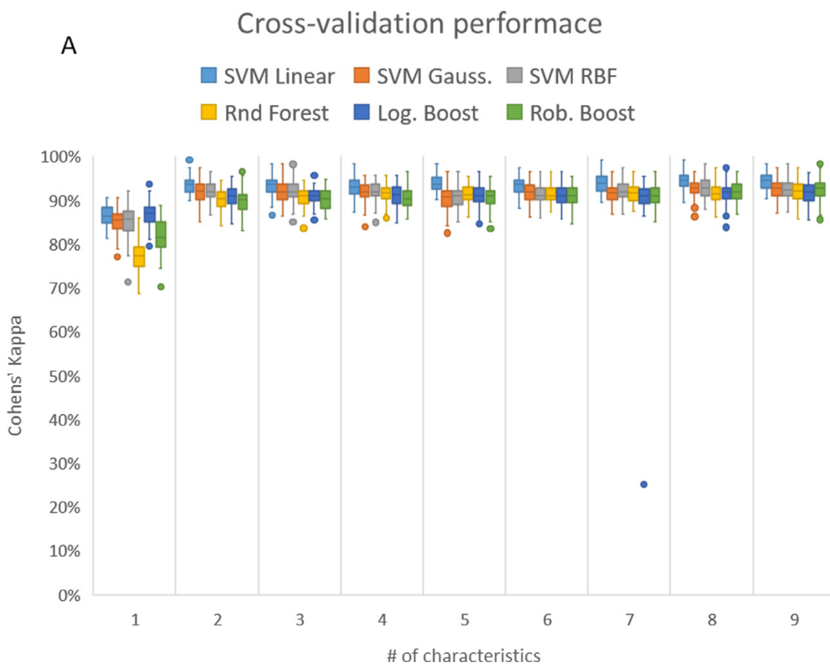


Fig. 8. ML model training summary for *dMice1* mice dataset. (A) scatter plot of the four classes represented with the 3 most relevant characteristics scored with NCA. (B) boxplot showing the validation performance after the 100-fold cross-validation for each ML algorithm and the number of characteristics. (C) Confusion matrix obtained after testing with the selected model. (D) Global and individual class performance obtained with the selected model.

A



B

Optimal model testing performance						
ML type	# Charac.	Pre	TPR	TNR	Acc	Kpp
SVMl	8	97.0%	92.2%	99.2%	97.8%	93.2%

C

	Re-training performance				
	Pre	TPR	TNR	Acc	Kpp
Build-in	97.4%	93.3%	98.9%	97.2%	93.2%
Retrain	94.9%	94.9%	97.8%	96.6%	92.7%

Fig. 9. ML model training summary for *dHuman* patients' dataset. (A) Boxplot showing the validation performance after the 100-fold cross-validation for each ML algorithm and the number of characteristics. (B) Selected model after the cross-validation and the corresponding testing performance. (D) Performance uplift of the initial model and after the re-training.

inal (*Build-in*) and updated (*Retrain*) model (Fig. 6, C). With these data and both models accessible, it is up to the user to continue with the refinement or to apply the selected model for detection.

3.2. Case of use 2: IED annotation in mice recording in a different set-up

The second case of use considers access to previous work, such as examples of events already annotated and a preexisting model (ex.: model and annotations from case of use 1). Nevertheless, it might happen that the dataset to analyze comes from a different experiment or set-up and/or includes a different number of channels or changes in the electrode configuration; *dMice2* dataset fulfills these conditions. In this case, events were marked by combining assisted annotation (model from case of use 1) and complementary non-assisted annotation.

Fig. 7 shows the initial performance obtained with case of use 1 model before (*Build-in*) and after re-training (*Retrain*). A noticeable uplift on the performance after re-training with custom annotations on this dataset is observed. Nevertheless, results show that even if the initial model does not fit perfectly with new data set, it could perform successfully. Alternatively, the retrained model can be generated and stored for later use.

3.3. Case of use 3: seizure annotation in mice recording

This third case of use aims to reflect the adaptability of the workflow when the detection of different type of events is attempted: now the interest is the annotation of seizures at their different stages from *dMice1* dataset. For such task, a change in the time window for characteristics estimation and usable models are required (see 2.3.2 Characteristics estimation). The new configuration is therefore modified for seizure detection and the whole *dMice1* dataset including pre-ictal, ictal and post-ictal periods is considered.

Fig. 8, A shows the output of the NCA analysis. The scatter plot illustrates the difference between the different classes. *Seiz1* and *Seiz2* classes that represent the initiation and consolidation of the seizures do not show a clear boundary between each other. Despite that, model cross-validation gave a satisfactory performance for almost all model configurations (Fig. 8, B). Results show that the optimal model is a Gaussian kernel SVM with 8 characteristics with satisfactory testing performance values represented on Fig. 8, C. *Seiz1* and *Seiz2* show the lowest values of performance even though they are satisfactory. Similarity between this two classes can be observed on the confusion matrix (Fig. 8, D) in concordance with the likeness of brain activity during these stages (actually, it is quite uncommon to differentiate among them); nevertheless, they are clearly differentiated from both *PreSeiz* and *InterSeiz* classes.

3.4. Case of use 4: IED annotation in human EEG recordings

Finally, a change of species is considered. Again, the same process of training, annotation and re-training was applied to *dHuman* dataset. This is a common scenario motivated by the interest of clinicians in using tools coming from the life-science/experimentalist developments (and vice-versa). As in case of use 3, this change of paradigm required the tuning of some of parameters for the characteristics estimation (see 2.3.2 Characteristics estimation).

Cross-validation during initial training shows remarkable performance (Fig. 9, A), selecting a linear SVM trained with 8 characteristics as the optimal model. Testing of this configuration shows Acc, Pre, TPR, TNR and Kpp of 97.0%, 92.2%, 99.2%, 97.8% and 93.2% respectively (Fig. 9, B). With such high-performance values, the performance did not change much after the re-training with the rest of the dataset (Fig. 9, C).

These results confirm furthermore the *transversality* (trans-species versatility) of the proposed workflow and thus demonstrate their capability to cope with multiple rodent datasets including different configurations, as well as dealing with data sets from clinical practice.

4. Conclusion and discussion

The adoption of new advances and tools from the fields of signal analysis and machine learning are often jeopardized by the complexity and difficulties to adapt such developments to the backgrounds, routines and daily needs of life-science researchers and clinicians [3,29]. However, their availability is essential in both clinical practice and basic research as they are key to assess (both on humans and animal models) the severity of the disease and the effects of therapeutic interventions.

In the field of epilepsy, there is an urgent need to process automatically large datasets of recordings obtained from experimental investigations and clinical tests. In this field, IEDs and seizure are key features to be identified. Due to their nature and low impact on patients' behavior, individual IEDs are only recordable on EEG registers. On contrast, seizures can be identified by multiple signal modalities with varied performance depending on the seizure and epilepsy type [15]. Among them, EEG is the gold standard for diagnosis, as brains behavior is altered regardless the type of seizures and epilepsy. In the literature, several approximations to ease IED and seizure annotations have been proposed [12–14,17–21,30]. Despite offering automatized detection systems, most of them are restricted to detect either interictal activities or seizures, tested on fixed configurations (same electrode configuration and recording settings) or restricted to a single species. User-friendliness for clinicians and basic researchers is not reflected and, in most cases, these solutions are static or require high levels of coding and data analysis skills for proper adaptation to different conditions.

In this context, we propose an interactive and iterative workflow capable of combining the knowledge of clinical and basic research experts with new methodologies to perform the automatic annotation of epileptic traits on electrophysiological recordings [31]. This workflow has been implemented in a stand-alone interface that integrates all the processes of annotation, cross-validation, test and optimization of ML models. It was conceived to accomplish such tasks in a research environment where clinicians, engineers and life-science researchers interact and carry out investigations involving both, animal models and studies in human beings. To test this pipeline, four different cases of use are presented: they range from the most preliminary scenario without previous models and just recordings to process to a more general situation in which previous knowledge can be used to ease the analysis of datasets from different configurations, setups or species.

Performance in all these exemplifying cases is in consonance with the figures obtained in other investigations. Metrics are satisfactory and comparable with some commercial systems [16] even if there is a modification in the recording set-up (different configuration or recording conditions), the event of interest changes (and focus is modified from studying interictal activities to detecting seizures) or the data comes from a completely different environment (experimental/clinical) or species (mouse/human).

Such results demonstrate that the capabilities of the proposed approach (and tool) represent a convenient workspace for both clinical and research environments, aiding experts with the workload of annotating large amounts of hours-long registers. In addition, the trans-species and trans-event versatility of the tool could help to better understand the relation between multiple pathological events and to characterize differences and similarities between human diseases and animal models. The possibility to mark events on individual channels permits to identify region-specific susceptibilities and will allow to detect and track the epileptic foci. This functionality could lead to faster and more exhaustive markings with less effort. These flexibilities are obtained by simple modifications within the implemented tool and likely could extend their use to other fields of research such as sleep.

The detection motor and the interactive framework together with the pre-trained models and the datasets described in this work are freely available through request to the authors. By doing so, we expect not only to provide life-science researchers and clinicians with a tool that will help them in their investigations, but also to motivate data analysts and machine learning experts to improve our development and to adopt our scope of providing rigorous but usable/feasible tools to be used in non-specialized environments.

Declaration of Competing Interest

Funding sources did not have any role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript. This does not alter the authors' adherence to all the journal policies on sharing data and materials.

Acknowledgments

Authors acknowledge financial support from CureDravet E-rare project from EU, Inocente-inocente Foundation and Instituto de Salud Carlos III "DTS19/00130″. (Co-funded by *European Regional Development Fund; "A way to make Europe"*). GB is a recipient of an ADA (Asociación de Amigos de la Universidad de Navarra) fellowship. We thank Dr. Ruben Hernandez-Alcoceba, Dr. Ana Ricobaraza, Dr. Rocio Sanchez-Carpintero, Ángel Ochoa-Martinez, Tomás Muñoz and Irene Gonzalez-Burgos for their valuable comments and support.

References

[1] C.E. Stafstrom, L. Carmant, Seizures and epilepsy: an overview for neuroscientists, cold spring harb, Perspect. Med. 5 (2015) a022426, doi:10.1101/CSHPERSPECT.A022426.

[2] O. Devinsky, A. Vezzani, T.J. O'Brien, N. Jette, I.E. Scheffer, M. de Curtis, P. Perucca, Epilepsy, Nat. Rev. Dis. Prim. 41 (2018) 1–24 20184, doi:10.1038/nrdp.2018.24.

[3] S. Noachtar, J. Rémi, The role of EEG in epilepsy: a critical review, Epilepsy Behav. 15 (2009) 22–33, doi:10.1016/j.yebeh.2009.02.035.

[4] J.N. Gelinas, D. Khodagholy, T. Thesen, O. Devinsky, G. Buzsáki, Interictal epileptiform discharges induce hippocampal-cortical coupling in temporal lobe epilepsy, Nat. Med. 22 (2016) 641–648, doi:10.1038/nm.4084.

[5] P. Dahal, N. Ghani, A. Flinker, P. Dugan, D. Friedman, W. Doyle, O. Devinsky, D. Khodagholy, J.N. Gelinas, Interictal epileptiform discharges shape large-scale intercortical communication, Brain 142 (2019) 3502–3513, doi:10.1093/brain/awz269.

[6] E.L. So, Interictal epileptiform discharges in persons without a history of seizures: what do they mean? J. Clin. Neurophysiol. 27 (2010) 229–238, doi:10.1097/WNP.0b013e3181ea42a4.

[7] W.O. Tatum, G. Rubboli, P.W. Kaplan, S.M. Mirsatari, K. Radhakrishnan, D. Gloss, L.O.K. Caboclo, F.W. Drislane, M. Koutroumanidis, D.L. Schomer, D. Kastelijn-Nolst Trenite, M.J. Cook, S. Beniczky, Clinical utility of EEG in diagnosing and monitoring epilepsy in adults, Clin. Neurophysiol. 129 (2018) 1056–1082, doi:10.1016/J.CLINPH.2018.01.019.

[8] M.A. Kural, L. Duez, V. sejer Hansen, P.G. Larsson, S. Rampp, R. Schulz, H. Tankisi, R. Wennberg, B.M. Bibby, M. Scherg, S. Beniczky, Criteria for defining interictal epileptiform discharges in EEG: a clinical validation study, Neurology 94 (2020) e2139–e2147, doi:10.1212/WNL.0000000000009439.

[9] S.B. Wilson, R.N. Harner, F.H. Duffy, B.R. Tharp, M.R. Nuwer, M.R. Sperling, Spike detection. I. Correlation and reliability of human experts, Electroencephalogr. Clin. Neurophysiol. 98 (1996) 186–198, doi:10.1016/0013-4694(95)00221-9.

[10] J. Jing, A. Herlopian, I. Karakis, M. Ng, J.J. Halford, A. Lam, D. Maus, F. Chan, M. Dolatshahi, C.F. Muniz, C. Chu, V. Sacca, J. Pathmanathan, W. Ge, H. Sun, J. Dauwels, A.J. Cole, D.B. Hoch, S.S. Cash, M.B. Westover, Interrater reliability of experts in identifying interictal epileptiform discharges in electroencephalograms, JAMA Neurol. 77 (2020) 49–57, doi:10.1001/jamaneurol.2019.3531.

[11] F. Moraes, D.A. Callegari, Automated detection of interictal spikes in EEG : a literature review, Clin. Neurophysiol. (2014) 1095–1103.

[12] J. Jing, H. Sun, J.A. Kim, A. Herlopian, I. Karakis, M. Ng, J.J. Halford, D. Maus, F. Chan, M. Dolatshahi, C. Muniz, C. Chu, V. Sacca, J. Pathmanathan, W. Ge, J. Dauwels, A. Lam, A.J. Cole, S.S. Cash, M.B. Westover, Development of expert-level automated detection of epileptiform discharges during electroencephalogram interpretation, JAMA Neurol. 77 (2020) 103–108, doi:10.1001/jamaneurol.2019.3485.

[13] M. El-Gohary, J. McNames, S. Elsas, User-guided interictal spike detection, in: Proceedings of the 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, IEEE Computer Society,

2008, pp. 821–824, doi:10.1109/iembs.2008.4649280. EMBS'08 - "Personalized Healthc. through Technol.
[14] S.S. Lodder, M.J.A.M. Van Putten, A self-adapting system for the automated detection of inter-ictal epileptiform discharges, PLoS ONE 9 (2014) e85180, doi:10.1371/journal.pone.0085180.
[15] C. Baumgartner, J.P. Koren, M. Rothmayer, Automatic computer-based detection of epileptic seizures, Front. Neurol. 9 (2018) 639, doi:10.3389/FNEUR.2018.00639/BIBTEX.
[16] J. Koren, S. Hafner, M. Feigl, C. Baumgartner, Systematic analysis and comparison of commercial seizure-detection software, Epilepsia 62 (2020) 426–438, doi:10.1111/EPI.16812.
[17] F. Fürbass, M.A. Kural, G. Gritsch, M. Hartmann, T. Kluge, S. Beniczky, An artificial intelligence-based EEG algorithm for detection of epileptiform EEG discharges: validation against the diagnostic gold standard, Clin. Neurophysiol. 131 (2020) 1174–1179, doi:10.1016/J.CLINPH.2020.02.032.
[18] J.A. Pfammatter, R.K. Maganti, M.V. Jones, An automated, machine learning–based detection algorithm for spike-wave discharges (SWDs) in a mouse model of absence epilepsy, Epilepsia Open 4 (2019) 110–122, doi:10.1002/epi4.12303.
[19] Q.M. Tieng, I. Kharatishvili, M. Chen, D.C. Reutens, Mouse EEG spike detection based on the adapted continuous wavelet transform, J. Neural Eng. 13 (2016), doi:10.1088/1741-2560/13/2/026018.
[20] R.A. Bergstrom, J.H. Choi, A. Manduca, H.S. Shin, G.A. Worrell, C.L. Howe, Automated identification of multiple seizure-related and interictal epileptiform event types in the EEG of mice, Sci. Rep. 3 (2013) e1483, doi:10.1038/srep01483.
[21] P.M. Casillas-Espinosa, A. Sargsyan, D. Melkonian, T.J. O'Brien, A universal automated tool for reliable detection of seizures in rodent models of acquired and genetic epilepsy, Epilepsia 60 (2019) 783–791, doi:10.1111/epi.14691.
[22] A. Ricobaraza, L. Mora-Jimenez, E. Puerta, R. Sanchez-Carpintero, A. Mingorance, J. Artieda, M.J. Nicolas, G. Besne, M. Bunuales, M. Gonzalez-Aparicio, N. Sola-Sevilla, M. Valencia, R. Hernandez-Alcoceba, Epilepsy and neuropsychiatric comorbidities in mice carrying a recurrent dravet syndrome SCN1A missense mutation, Sci. Rep. 9 (2019) 1–15, doi:10.1038/s41598-019-50627-w.
[23] C. Dravet, Les epilepsies graves de l'enfant, Vie Med. 8 (1978) 543–548.
[24] M. Wolff, C. Cassé-Perrot, C. Dravet, Severe myoclonic epilepsy of infants (dravet syndrome): natural history and neuropsychological findings, Epilepsia 47 (2006) 45–48 Suppl 2, doi:10.1111/J.1528-1167.2006.00688.X.
[25] M. Sakauchi, H. Oguni, I. Kato, M. Osawa, S. Hirose, S. Kaneko, Y. Takahashi, R. Takayama, T. Fujiwara, Retrospective multiinstitutional study of the prevalence of early death in dravet syndrome, Epilepsia 52 (2011) 1144–1149, doi:10.1111/j.1528-1167.2011.03053.x.
[26] C. Dravet, M. Bureau, B.D. Bernardina, R. Guerrini, Severe myoclonic epilepsy in infancy (dravet syndrome) 30 years later, Epilepsia 52 (2011) 1–2, doi:10.1111/J.1528-1167.2011.02993.X.
[27] J. Van Erum, D. Van Dam, P.P. De Deyn, PTZ-induced seizures in mice require a revised racine scale, Epilepsy Behav. 95 (2019) 51–55, doi:10.1016/j.yebeh.2019.02.029.
[28] M.L. McHugh, Interrater reliability: the kappa statistic, Biochem. Med. 22 (2012) 276–282, doi:10.11613/bm.2012.031.
[29] S. Beniczky, H. Aurlien, J.C. Brøgger, L.J. Hirsch, D.L. Schomer, E. Trinka, R.M. Pressler, R. Wennberg, G.H. Visser, M. Eisermann, B. Diehl, R.P. Lesser, P.W. Kaplan, S. Nguyen The Tich, J.W. Lee, A. Martins-da-Silva, H. Stefan, M. Neufeld, G. Rubboli, M. Fabricius, E. Gardella, D. Terney, P. Meritam, T. Eichele, E. Asano, F. Cox, W. van Emde Boas, R. Mameniskiene, P. Marusic, J. Zárubová, F.C. Schmitt, I. Rosén, A. Fuglsang-Frederiksen, A. Ikeda, D.B. MacDonald, K. Terada, Y. Ugawa, D. Zhou, S.T. Herman, Standardized computer-based organized reporting of EEG: SCORE – second version, Clin. Neurophysiol. 128 (2017) 2334–2346, doi:10.1016/J.CLINPH.2017.07.418.
[30] L. Billeci, A. Tonacci, D. Marino, L. Insana, G. Vatti, M. Varanini, A machine learning approach for epileptic seizure prediction and early intervention, in: Biosystems Biorobotics, Springer International Publishing, 2019, pp. 972–976, doi:10.1007/978-3-030-01845-0_195.
[31] J. Patel, The democratization of machine learning features, in: Proceedings of the IEEE 21st International Conference on Information Reuse and Integration for Data Science (IRI), 2020, pp. 136–141, doi:10.1109/IRI49571.2020.00027. 2020.