



Research

Cite this article: Ali E, Angelova M, Karmakar C. 2024 Epileptic seizure detection using CHB-MIT dataset: The overlooked perspectives. *R. Soc. Open Sci.* **11**: 230601.
<https://doi.org/10.1098/rsos.230601>

Received: 19 May 2023

Accepted: 28 March 2024

Subject Category:

Computer science and artificial intelligence

Subject Areas:

artificial intelligence, medical computing, biomedical engineering

Keywords:

epilepsy, seizure, machine learning, seizure event detection, cross-subject analysis, health informatics

Authors for correspondence:

Emran Ali

e-mails: emran.ali@research.deakin.edu.au;

aliem@deakin.edu.au

Maia Angelova

e-mails: maia.a@deakin.edu.au;

m.angelova@aston.ac.uk

Chandan Karmakar

e-mail: karmakar@deakin.edu.au

Electronic supplementary material is available online at <https://doi.org/10.6084/m9.figshare.c.7184012>.

Epileptic seizure detection using CHB-MIT dataset: The overlooked perspectives

Emran Ali¹, Maia Angelova^{1,2,3} and Chandan Karmakar¹

¹School of Information Technology, Deakin University, Melbourne Burwood Campus, Melbourne, Victoria 3125, Australia

²Aston Digital Futures Institute, EPS, Aston University, Birmingham, UK

³Institute of Biophysics and Biomedical Engineering, Bulgarian Academy of Sciences, Sofia, Bulgaria

EA, 0000-0003-3095-5265; MA, 0000-0002-0931-0916; CK, 0000-0003-1814-0856

Epilepsy is a life-threatening neurological condition. Manual detection of epileptic seizures (ES) is laborious and burdensome. Machine learning techniques applied to electroencephalography (EEG) signals are widely used for automatic seizure detection. Some key factors are worth considering for the real-world applicability of such systems: (i) continuous EEG data typically has a higher class imbalance; (ii) higher variability across subjects is present in physiological signals such as EEG; and (iii) seizure event detection is more practical than random segment detection. Most prior studies failed to address these crucial factors altogether for seizure detection. In this study, we intend to investigate a generalized cross-subject seizure event detection system using the continuous EEG signals from the CHB-MIT dataset that considers all these overlooked aspects. A 5-second non-overlapping window is used to extract 92 features from 22 EEG channels; however, the most significant 32 features from each channel are used in experimentation. Seizure classification is done using a Random Forest (RF) classifier for segment detection, followed by a post-processing method used for event detection. Adopting all the above-mentioned essential aspects, the proposed event detection system achieved 72.63% and 75.34% sensitivity for subject-wise 5-fold and leave-one-out analyses, respectively. This study presents the real-world scenario for ES event detectors and furthers the understanding of such detection systems.

1. Introduction

Epilepsy is one of the most prevalent cranial conditions in human beings that affects millions of people around the world

[1]. It brings down the quality of life for those affected, regardless of age, race or geographic location. A transitory incidence of signs or symptoms associated with abnormally excessive and synchronized neuronal activity in the brain is known as an epileptic seizure (ES) or generally a seizure [2,3]. Although 75% of epilepsy cannot be prevented [4], with adequate medical treatment and diagnosis, many epilepsy patients can be cured [5–7]. An appropriate diagnosis is required to identify the seizure's type, nature and location for effective treatment of ES. Manual detection of ES from very long and continuous data is distressful and infeasible. Thus, an automated ES detection system is used independently or as a clinical support system. Various diagnostic options are available for identifying ES occurrence, including blood tests, physiological examination and different signal analysis processes. The signal modalities used for ES detection are accelerometry (ACM) [8–10], ECG [11–13], EEG [14–18], MRI [19,20] and so forth. Continuous electroencephalography (cEEG) is considered the gold standard for seizure detection among all available modalities [15,16,21]. The scalp electroencephalography (scalp EEG) signal is preferred over intracranial EEG for diagnostic and general purposes as being a non-invasive process [22,23]. One of the most extensive continuous scalp EEG datasets is the CHB-MIT dataset, and numerous studies have been conducted using this dataset [24].

Diagnosis of ES using EEG is simply classifying or separating the seizure portion of the EEG signal from the normal (non-seizure) one. This classification can be done using threshold-, machine learning (ML)- or deep learning (DL)-based methods. Among these methods, ML methods are preferred concerning the model's generalizability, feature explainability and detection latency [25,26]. They use different explainable (human perceivable) features that define the seizure event and differentiate from the normal signal [27–31]. In medical diagnosis, only classifying seizure from non-seizure data is not enough; rather, a proper explanation of the classification process is required. The explanations include which features are responsible for separating the seizure data from non-seizure, and based on what physical characteristics those features can do so. These explanations help medical experts map the feature with corresponding biological or chemical factors for proper treatment planning.

One of the best ways to get the actual scenario for seizure detection using data-driven methods is to test the method's generalizability power. To do so it is required to train the model with sufficient data and test the trained model using the rest of the unseen data. However, continuous EEG data is typically imbalanced and has fewer (seizure) events than the normal signals. Most of the studies found in the literature that used large public datasets (especially the CHB-MIT dataset) have not considered all the data for testing the performance of their method. These types of experiments on partial data may temporarily show significantly higher performance. Still, they will struggle to perform well once they encounter the actual scenario (the imbalanced nature of continuous data). All the studies that considered different datasets such as the Bonn dataset considered all the data for observation [32–38]. However, it is a minimal dataset, too small to test the generalizability of a model. Other studies that used private data showed the performance scores on the entire dataset, but due to availability issues, they cannot be tested [14,39,40]. Most of the studies that were done on the CHB-MIT dataset considered only the records that have seizures in them [41–43]. The majority of the previous studies discarded the records with no seizure events in them, even though those records are part of the continuous EEG signal. Some studies even considered a portion of the data or excluded some based on their choice for analysis [44–48]. When the data are balanced for testing or portions of data are left out of consideration in the analytical process, this can greatly impact the performance score. Different data alteration processes are often used in many studies, such as filtering, augmentation, scaling and related preprocessing [49–52]. This type of alteration may change the original property of the signal [53]. Data in its entirety is demanded and also legitimately evaluated to test a model's capability.

Similar to any other physiological signals, EEG signals are highly nonlinear and non-stationary [54]. The inter-subject variability in EEG signal is much higher than the intra-subject changes [55,56]. Thus, the random choice of the signal data to be included in training and test sets for classification (for random fold-based splitting) may lead to a data leakage problem. In this case, the data segments from the same subjects may appear in both the training and test sets. This situation gives the model a better idea about the data pattern available in the test set before testing, and it might produce better results that are not realistic. This data leakage situation leads to the need for cross-subject analysis so that data leakage can be prevented [51,57–60]. Most of the previous studies ignored this factor and did not consider cross-subject analysis [18,42,61–63]. Some subject-specific or patient-specific systems were investigated in earlier studies, where the training and test were done on the same-specific subject [24,57,64–66]. However, this approach has a critical pitfall: the system needs to be trained on the subject before it can be used for that same subject. Additionally, a large amount of data collected from other

patients cannot be used in this system design. Thus, it hinders the system from being robust and generalized for all the patients.

Different approaches can be used to detect the seizure occurrence from the EEG signal. The most common approach for ES classification methods is segment classification [37,67,68]. This approach divides the continuous raw EEG signals into small chunks called segments. This process of dividing the signal into small pieces is called segmentation. The long seizure events are also divided and distributed into several segments during segmentation. The ML model is trained on some (e.g. 70–80% of total segments) randomly selected segments. Then, the trained model is tested on separate random segments (e.g. 30–20% of total segments) that have not been presented to the model during training. The performance of the classifier model is evaluated based on how many test segments are correctly identified or not [27,37,59,60,63,68–70]. Only considering the segment detection does not always guarantee that the correctly detected segments are from the same event. So, the question remains: is the model able to detect seizure events correctly rather than just detecting random segments? The other approach to classification is event classification. In this approach, the event with its original length can be separated and detected. The entire length of seizure or non-seizure can be separated and used for analysis to classify seizure or non-seizure. Another alternative to the entire event separation for event detection is to combine segment results. The detected segments can be sorted in their original sequence to check if all the or majority of the segments in the entire event are being detected correctly. Event detection is a more realistic consideration for real-world application [41,42,61,62] because it specifies if a model is capable of detecting seizure events, not just random segments. Unfortunately, most of the previous studies did not focus on this important factor but rather focussed only on random seizure segment detection. Several previous studies worked on event detection that includes entire event separation and detection (as mentioned above) and seizure onset (the initial portion(s) of the seizure event) detection [41–43]. Only one of these studies tackled data leakage and random segment detection problems by observing the event detection for the cross-subject setup [41].

All the factors mentioned above are associated with the real-world situation of a seizure detection system. These must be noticed when developing an actual ES detection system. Although some studies in the literature considered one or two factors, almost all of them lack considering all these factors altogether. To address all the unnoticed issues, in this study, we investigated the detection of ES events using multiple channels of continuous EEG signal from the most commonly used and publicly available scalp EEG dataset, the CHB-MIT dataset. However, to detect the seizure events using the ML model, the cross-subject classification strategy was employed and considered all the data from the dataset unchanged during testing. The main aim of this study is to investigate the efficacy of the cross-subject seizure event detection system using the ML model on all of the original data from the CHB-MIT dataset, which reveals the actual seizure detection scenario. To the best of our knowledge, this is one of the first studies that considers all the above-mentioned factors for ES detection. Below are a number of objectives that support this study's main aim, which are the main novel contributions of our study:

- Detection of ES using ensemble ML models with explainable features.
- Observing the model's efficacy using the real-world data imbalance scenario. To do that, the unchanged data from the entire dataset is used for testing.
- Observing the cross-subject (subject-wise, SW) analysis of ES to ensure the inter-subject variability is intact and there is no data leakage between the training and test sets.
- Implementing the original goal of a seizure detection system, that is, to detect ES events instead of detecting random seizure segments.
- Combining all of these criteria to observe the efficacy of a seizure detection system.

The rest of the article is organized by starting with the data and methodological description in §2. It includes the dataset description, feature extraction and selection process. This section describes the classification method (model), parameters used in the classification, post-processing method and performance metrics. The obtained result is described in §3, and the comparison with the most relevant articles with detailed discussion is presented in §4. Finally, the report is concluded with the findings in §6 followed by the limitations of this study and the future direction of this study in §5.

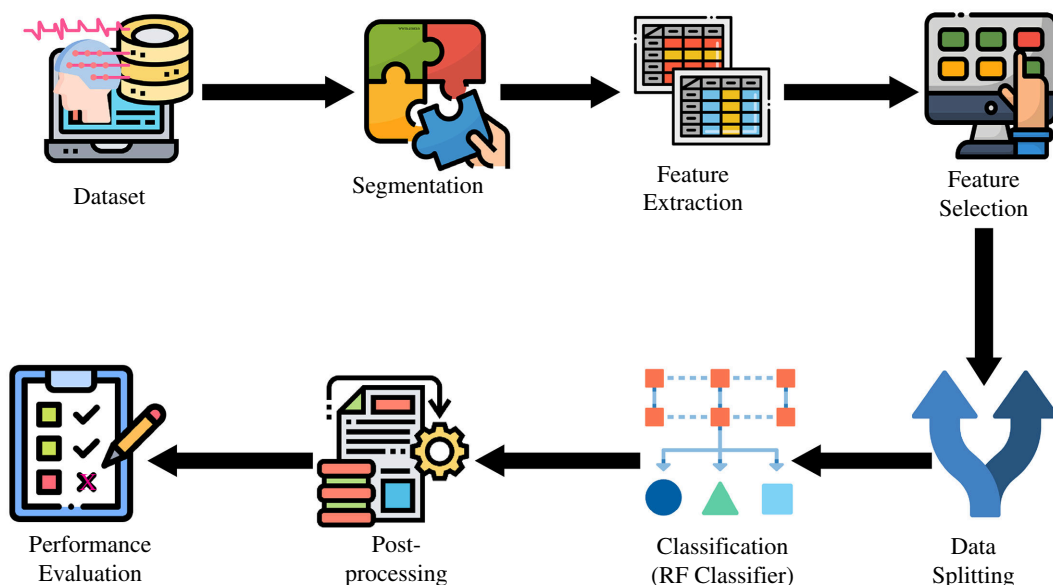


Figure 1. Block diagram of the event detection system demonstrating the steps followed for this study. Starting with raw EEG signal data collection from the dataset, the raw signal is cut into segments and different features are extracted from each segment of each channel. Feature selection is the next to select important features only. After properly splitting in the appropriate ratio for training and testing, data are sent to the RF classifier to be classified as seizure or non-seizure. Post-processing for event detection is done with the classified results and evaluated with proper performance measurement metrics.

2. Material and methods

2.1. Methodology

The methodology that we followed is described using the block diagram presented in figure 1. It includes data collection, segmentation and feature extraction, feature selection, data splitting and model training, testing and result generation, result post-processing and performance evaluation. The appropriate data from the CHB-MIT dataset are considered for the analyses of this study. The data are divided channel-wise and cut into segments for each of the channels. Different features are extracted from each segment of the raw EEG signal. Feature selection is applied to select more meaningful features for seizure detection. After properly splitting the data SW into appropriate ratios for training and testing, they are sent to the random forest (RF) classifier to be classified as seizure or non-seizure. Data splitting and classification that includes training, validation and testing of the RF model are repeated for k times SW, where $k = 5$ is for 5-fold analysis and $k = 24$ is for leave-one-out (LOO) analysis. Although the LOO analysis is our main target, we expanded the experiment to 5-fold to include more variations in the test set and to train the model with less data to observe the performance variations. To investigate the results in more detail, the segment-based results are also explored and compared before proceeding with event detection. This will also facilitate the comparison of the effect of subjective data leakage, which is one of the factors we considered in this study. Post-processing for event detection is done from the classified results and evaluated with proper metrics. Each of these steps is described in detail in the following sections. The entire analytical process (all the steps described) is done with the most popular programming language, Python. The flowchart in the electronic supplementary material presents a more detailed view of how the analysis was done using different steps in figure 1 of appendix B.

2.2. Dataset

The CHB-MIT scalp EEG dataset from Physionet (<https://physionet.org/>) was first introduced by Shoeb *et al.* [24]. It is one of the scalp EEG datasets frequently used for seizure detection or prediction investigations [48,71,72]. The data are obtained from 23 paediatric patients or 24 subjects (since subject 21 was obtained 1.5 years after subject 01 from the same patient). The male patients are in the age

Table 1. Summary of total available data in the dataset and the portion of data that is considered and observed for this study. Due to the selection of channel groups to make the experiments consistent with all the subjects, the total amount of seizure and non-seizure-related data is reduced. The column on the right-hand side represents the amount of data that is observed for this study.

criteria	total available	observed
subjects	24	24
records	686	655
seizure records	141	136
seizures events	198	175
signal frequency	256 Hz	256 Hz
channel groups	5	1
channels	18–32	22
signal types	EEG, ECG, VNS, –	EEG
seizure event samples	11 693 s	9925 s
seizure to non-seizure ratio	—	1:344
data balancing	—	imbalanced test
extracted features	—	92 (per channel)

Notes: The dash (—) represents that the data are unavailable or not represented.

group of 3 to 22 years, whereas female patients are in the age group of 1.5 to 19 years. A minimum of two types of seizures were recorded in each of the subjects among a total of three different categories, including partial (SP), complex partial (CP) and generalized tonic-clonic (GTC) seizures. Seizures in the hippocampus (in the deep temporal) area were recorded for eight subjects, and another 11 had seizures in the neocortical brain location. A total of 24 subjects contain 686 records with more than 950 hours (h) of recordings. The recording was done continuously using 256 Hz, 16-bit resolution for 1–4 h for each record. A total of 22–36 EEG channels were used to record the data from the subjects. However, in many cases, the record contains other signals along with the EEG signal, such as ECG, vagal nerve stimulus (VNS) and other dummy signals. Of all records, 141 have at least one seizure event (called seizure records), totalling 198 seizure events of lengths starting from 7 to 753 second (s). There are a total of 545 normal recordings (non-seizure records) for all the subjects, with no seizure events at all. The dataset has an approximate total of 11 693 s and 3 417 165 s of seizure and non-seizure events, respectively.

The EEG signals were recorded using the different combinations of electrodes (or channels), which can be termed channel groups. Five different channel groups were found in this dataset, each with 22–36 unique EEG channels. The electronic supplementary material presents more details about the channel in table 1 of electronic supplementary material, appendix C. One of the five groups is considered for analysis in this study. Although we did not intend to select channels for this study, we chose this particular group because it was used for most of the records in the dataset. This particular channel group has been used by 655 records (out of 686), which contain 175 seizure events (out of 198). There are 22 unique channels in this selected group and those are: FP1-F7, F7-T7, T7-P7, P7-O1, FP1-F3, F3-C3, C3-P3, P3-O1, FP2-F4, F4-C4, C4-P4, P4-O2, FP2-F8, F8-T8, T8-P8, P8-O2, FZ-CZ, CZ-PZ, P7-T7, T7-FT9, FT9-FT10 and FT10-T8. Signals from each of the 22 individual channels were segmented with a 5 s non-overlapping window. A total of 1985 seizure segments with 5 s duration and 6 83 433 non-seizure segments were obtained from the CHB-MIT dataset. A brief summary of the data used in this study is presented in table 1. However, a more detailed version of the data description can be found in table 2 of appendix C in the electronic supplementary material.

This dataset has a higher class imbalance as the non-seizure segments are more than 344 times larger in number than the seizure segments. The data were kept imbalanced to observe the real-world seizure detection scenario from continuous EEG, as in general, data will have more non-seizure duration than the seizure event duration. The signals from the dataset were used without any changes in the properties of the signal. No other preprocessing (e.g. filtering, artefact removal, etc.) was applied to the signal, as these may alter its original characteristics, as previously described. This alteration also may lead to changes in the actual properties of the seizure portion of the signal.

Detected	0	0	1	1	0	0	0	0	1	1	0	0	0	0	1	0	0	0	0	1	1	1	0	0	1	1	1	1	0	1	1	1	1	0	0	0	0	
Generated	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0

Figure 2. The process of event generation from the results from all the segments. The detected portion is the actual segment detection by the ML model and the generated one is the new/modified results after applying post-processing for the corresponding segments. After combining the results of the segments sequentially, the labels of the seizure segments (with 1) in a continuous non-seizure event are replaced with non-seizure labels if the duration is 10 (ten) s or less (as the first two changes indicated using the shaded areas). Similarly, the labels for non-seizure segments (with 0) in a continuous seizure event are replaced with seizure labels if the duration is 10 (ten) s or less (as the last two changes indicated using the shaded areas).

2.3. Segmentation and feature extraction

This study used a 5 s non-overlapping window to segment the signal from the channels. The reason for choosing a 5 s window is that segments that are too small or too large might lead to more inaccurate classification. Too small segments might not have enough effective patterns that the extracted feature can capture, while too long segments might dilute the actual pattern of the data. In addition, several studies have achieved good results with this length [60,73,74]. Although the effect of different segment lengths could be observed, that is kept out of the scope of this study as future work.

Different linear and nonlinear features extracted from the EEG signal have been found to be useful for ES detection [75–79]. Thus, from each 5 s scalp EEG segment, different linear and nonlinear features from the time and frequency domains were extracted, which are listed in table 3 of appendix C in the electronic supplementary material. Entropy profiling is one of the important features that can describe heart-rate variability very well [80]. Some secondary features from the sample entropy profile were used in this study as the time-domain nonlinear features. In addition, several statistical features from five commonly used frequency sub-bands of EEG (alpha, beta, delta, theta and gamma) were extracted. Similarly, spectral power-based nonlinear features were extracted from the raw EEG signal. All features used in this study are listed in table 3 of appendix C, and their details are provided in appendix A of the electronic supplementary material.

In summary, 92 features were extracted from each 5 s scalp EEG segment. These features were extracted using a library developed by our team that is available on our GitHub repository.¹ All these hand-crafted features were then stored along with the original label (if the segment is seizure or non-seizure) for each segment. When preparing the dataset, seizure segments were labelled as 1, and non-seizure segments were labelled as 0. Some segments contained a portion of both the seizure and non-seizure EEG. The labels for those segments were decided as seizures if they contained 60% (3 s in our case) or more seizure portions, and non-seizure otherwise.

2.4. Feature selection

Since this is a multi-channel study and 92 features were extracted for each channel, it resulted in a total of (92 features × 22 channels) = 2024 features for each epoch/segment, which are huge in number and might cause the curse of dimensionality. Thus, feature selection was used to remove the least effective features. The feature selection was done on sample data using four different feature selection methods, namely, chi-squared, correlation (with target), decision tree (DT) classification importance, and RF (permutation importance-based). The sample data comprised the segments taken from one channel of randomly selected five subjects. The feature selection process was repeated for randomly chosen five different channels, and each feature's average importance score was calculated at the end. To select features, the cumulative average importance scores (on a scale of 1.0) of all feature selection methods were compared as shown in electronic supplementary material, figure 2 where a threshold of 0.9 was used (empirically) as a cut-off point. The minimum number of features required to achieve 0.9 cumulative score by the four feature selection methods is listed in table 4 of appendix C in the electronic supplementary material. Based on the cumulative average importance scores and the minimum number of features that achieved the cut-off point, the chi-squared feature selection method was chosen for this study. Thus, 34 features out of 92 were selected for each channel, and those are listed in table 5 of appendix C in the electronic supplementary material. Finally, from 22 channels, a total of (34 features × 22 channels) = 748 features were then used for the classification of seizures.

¹https://github.com/WWM-EMRAN/DIHC_FeatureManager

2.5. Data splitting and training data balancing

Data splitting was done based on the demand of the experiments used in this study. Two sets of cross-subject experiments were conducted in this study to observe the event detection performances. One used SW 5-fold cross-validation, and the other used the LOO cross-validation mechanism (random samples (RS) are not distributed in different folds). All 24 subjects were randomly grouped into five groups for the first case. The grouping was done this way so that each group has some random subjects that are unavailable in any of the other four groups, similar to other previous studies [51,60]. So, the classifier model was trained using only four groups and tested with the left-out group for seizure segment detection. For all five groups, each was tested based on the training using the remaining four groups of subjects. For the LOO mechanism, similar to some related studies, each of the 24 subjects was considered the test set, and the model was trained on the rest of the 23 subjects other than the only test subject in a repeated fashion [41,51,81]. Both mechanisms ensured the training-test process, in which the model was not tested with the data of any subjects it was trained on.

Once the training-test splitting was done, we applied class balancing on the training dataset. As described earlier about the imbalance of classes in the CHB-MIT dataset, the total duration of the non-seizure portion is higher than the seizure portion. If we keep the training set highly imbalanced, the model will most likely be biased towards the majority class (non-seizure). For this reason, the model was trained on the balanced data using the under-sampling mechanism, but the test data was kept as it was. We have used under-sampling since over-sampling or augmentation may lead to generating seizure patterns that are not actual seizures.

2.6. Classification

RF classifier is an ensemble ML classifier that achieved greater performance scores compared with other ML methods on average as found in previous studies [47,60,62,63]. The ML models are more explainable compared with the DL models; for that reason, the ML model, specifically the RF model, was chosen as an initial model for this study. Investigating the performance of other similar methods would be very interesting to observe and would be a good finding itself. To focus on the main goal of this study, this criterion is left out for future work. However, different ML models were observed (using PyCaret Python library²) on the sample dataset explained in §2.4 to select the final model for this experiment. Table 6 of appendix C in the electronic supplementary material lists the performance scores of all the models for that observation. Even though other models were not tested in this study, the overall performance scores for the sample data shown in this table helped with model selection. Based on these performance scores, the RF model was chosen for this study over other ML models. The RF classifier generated labels (predicted on test data) for each segment for the corresponding features were later used for event analysis.

The RF classifier can be equipped and tuned with many different parameters to perfect the classifier so that we can find a more robust and generalized model. Of all the possible parameters that can be used, in this study, the '*n_estimators*' parameter was used and tuned for the model. The '*n_estimators*' is the number of trees that can be used to actively participate in the selection criteria of the voting score of RF, or simply the number of trees used in the forest. The model was checked and tuned for the '*n_estimators*' parameter with the values—15, 21, 30, 50 and 75. To optimize these parameters using Grid-search and choose the best model with optimized parameters, a subject-level random 30% of the training data was used for validation purposes.

2.7. Post-processing

As obvious, every subject contains many long-term records, and all the records for the same subject can be considered continuous (as the dataset describes). For event detection, all the records from the same subject are sorted based on their temporal order. This way, a large chunk of a logical unit of signal from every single subject is achieved and used for event observation. The segment-based result was kept as it is, based on the counts of the segments that were properly classified. Unlike seizure segment detection, simply counting the number of segments is not enough for performance evaluation for event detection. For event detection, checking the continuity of the detected segments,

²<https://github.com/pycaret/pycaret>

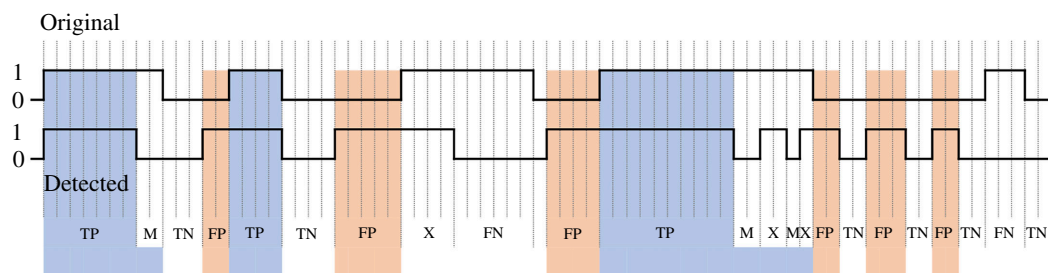


Figure 3. The process of event matching in post-processing by emphasizing event detection and missing events or unnecessary event detection. The digital waveform at the top represents the original events with 1 as seizure and 0 as non-seizure. Similarly, the waveform at the bottom represents the detected events by the segment-based systems (after event generation process) with a similar annotation for seizure and non-seizure events. The shaded green region shows the correct event detection or TP and the shaded orange region represents the false detection or FP. For the region, *M*, the FN is disregarded since the TP of both sides covers at least 70% (the event threshold) of the original seizure event. On the other hand, the region *X* is not considered TP since the total detected seizure duration does not cover the event threshold in the original event.

along with the start and end of the event, is also important. Since the goal of this study is to observe the performance of event detection, the performance was measured after some post-processing for event evaluation. However, some segment-based performances are also presented in this study. Due to the segmentation process, individual seizure event was distributed to several segments. For event detection, the segment-based results of the models were combined sequentially in the same order as the original data to generate a combined result, similar to some previous studies [63]. For event calculation, the seizure events that are less than or equal to 10 s in length in the combined result were ignored (and replaced with 0). Seizures can also be detected from shorter segments such as 2 or 5 s segments [41,60,82]. However, for better seizure detection, a 10 s latency has been found to be comparatively more effective in many previous studies [18,83]. For that reason, the segment length chosen for our study is 5 s as described in earlier section as well. Thus, to make an event, at least two such segments with a total length of 10 s are selected as the threshold for post-processing. Any non-seizure event between two seizure events was ignored (and replaced with 1), and those two events were considered continuous when the gap between them was less than or equal to 10 s. The event generation process from the segment results is shown in figure 2. The original model-generated example output (labelled as *Detected* in the figure) contains seven discrete events, but after post-processing (the changes marked with blue shadow) the three actual events are found (2nd row labelled as *Generated*). Then, the post-processed (generated) seizure labels were compared with the true target labels.

To compare, if any detected event (model result) covers 70% of the duration of the corresponding original seizure event (true data label), then it is considered that the model detected that event correctly. If there is one long true event (seizure sequence) and the model detects more than one small event within that period, then only one of the events is counted (preferably the largest one). This is done to ensure that there is no more than the actual number of events detected by the model and increases the sensitivity (*sen*) score incorrectly. On the other hand, the events detected in the non-seizure zones were kept as they were so that the actual misclassification was counted. After post-processing, different performance metrics were used to measure the performance of event-based seizure detection. In brief, a seizure event is considered to be detected if the original data contains an event (stream of continuous seizure segments) and the model detects at least 75% of that event duration after post-processing. On the other hand, a false detection is considered whenever the model detects an event (stream of continuous seizure segments) as a seizure, but that portion was non-seizure in the original data.

The event matching criteria with different situations are presented in figure 3. Here the *Original* label indicates the true seizure labels from the dataset and *Detected* refers to the *Generated* labels from the previous step. Small mistakes (false detection) are closely observed and punished to make the seizure detector more accurate and realistic. On the other hand, less flexibility is given to the detection of the correct events, which are bound to be at least a 70. This mechanism is similar to the overlap (OVLP) method explained in the previous study but not the same [64,84,85]. However, instead of the event-matching principle of ‘any’ part of the original event being detected by the ML model, we followed this 70% overlap principle to make

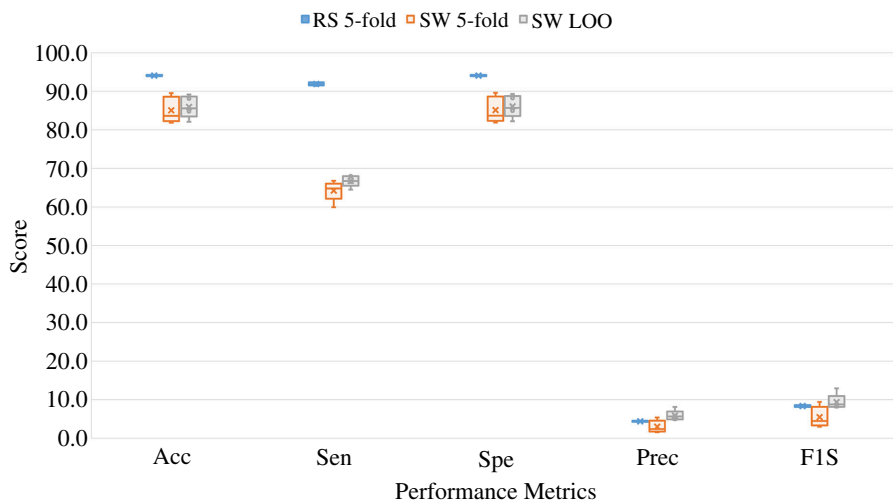


Figure 4. The segment-based scores in terms of Sen scores for RS 5-fold, SW 5-fold and SW LOO analyses. The Sen score is presented as a percentage result. This box plot shows the Sen scores on average with the distribution of the results for five runs for the analyses.

the system more reliable. From figure 3, if the detected segments cover at least 70% of the original event, it is considered true positive (TP). The regions marked with *M* are discarded because of the above condition, even though they are false negatives (FNs). The regions marked with *X* are also discarded, even though they are partially TPs. On the other hand, if the system detects a seizure but there is no seizure in the original signal, it is considered a false positive (FP). For example, in this figure, let us consider it is for a 2-h-long data. Since the system detected 3 out of 5 seizure events correctly, thus the (sen) will be 60%. On the other hand, it misclassified the portion of the seizure event six times in 2 h; thus, the false detection rate (FDR) is 3/h.

2.8. Performance metrics

The performance of the study was measured using the following metrics. For ES analysis, the seizure event is considered a positive (1), and the non-seizure is negative (0).

- Sensitivity: The TP rate corresponds to the proportion of positive data points that are correctly considered positive with respect to all the positive data points. This is usually called Sen or rec (rec).

$$\text{Sen} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

- False detection rate: The FDR is the hourly rate of misclassification of seizure event. It is the measure of the total misclassified seizure events divided by the total duration in hours.

$$\text{FDR} = \frac{\text{FP}}{\text{total_hours_observed}}$$

where,

- *TP*: The cases in which the model predicted positive and the actual output was also positive.
- *True negative (TN)*: The cases in which the model predicted negative and the actual output was negative.
- *FP*: The cases in which the model predicted positive and the actual output was negative.
- *FN*: The cases in which the model predicted negative and the actual output was positive.

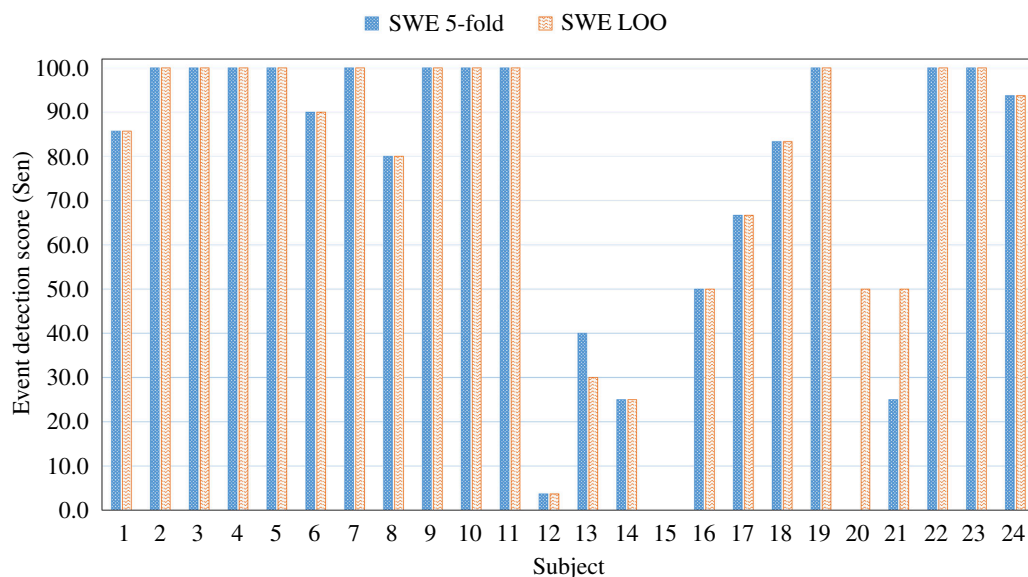


Figure 5. The event-based scores in terms of Sen scores for the SW cross-subject event detection (SWE) 5-fold and SWE LOO analyses for the individual subjects. The Sen score is presented as a percentage result.

Since the event-based analysis differs from segment detection and there is a high imbalance in the data classes, traditional metrics used for segment detection are inappropriate for this case. The more practical way to evaluate event-based performance is to use the Sen and FDR. These two metrics provide information on the number of seizure events that are correctly classified and the number that are misclassified per hour. In this case, the Sen score reports the seizure event detection performance, and the FDR score informs the efficacy of handling highly imbalanced non-seizure events. It is to be noted that the performances were calculated based on event count. The key information here is how many original events were properly detected and how many were missed.

3. Results

This section presents both the segment-based and event-based results. The first part of this section will discuss the segment-based results. The event-based results derived from the segment-based results will be discussed in detail in the later part. Keeping event detection in mind as our main target, segment detection performances were also observed. Seizure segment detection was done using different approaches: RS 5-fold, SW (cross-subject) 5-fold and SW LOO. In the first approach, the training and test samples were arbitrarily taken from different subjects for an 80:20 ratio for training and testing. On the other hand, for the latter two approaches, training and test data were divided SW as explained in §2.5. The experiment was run five times for all these approaches to get segment-based results. The segment-based comparative results for these three approaches are presented in figure 4. However, table 7 of appendix C in the electronic supplementary material shows more comprehensive detailed results for RS 5-fold. The graph of the segment-based results clearly shows that the RS 5-fold results show consistency across the runs, whereas the other two approaches have slight variations. For accuracy (Acc), Sen and specificity (Spe) scores, SW experiments performed significantly less than the RS 5-fold. On the other hand, a slightly different result has been found for SW LOO results. An increase in precision (Prec) has been found in the SW LOO approach, resulting in an improved *F1S*. This exception is expected to occur because comparatively large data (23 out of 24 subjects) have been used for training. This may also happen because of largely imbalanced classes in the test data. However, according to some earlier studies, *F1S* is not always a great choice for performance measurement when there is a higher imbalance in class data [86]. Thus, we stick to the two performance metrics defined in the earlier §2.8. These are the most common metrics used in earlier studies; thus, they can be used to compare our study with those. Since the variation in experimental results (of 5 runs) for both SW approaches is insignificant, we have taken one experimental result from each of these two SW approaches for event-based analysis. As explained in the earlier section, two types of cross-subject analytical approaches were observed for event detection in this study. The first is the

Table 2. The event count and performance score in terms of Sen and FDR of event-based investigation for SW both 5-fold and LOO analyses. Event count is presented with respect to the total number of actual seizures and non-seizures. The Sz indicates the seizure count and Nsz is the non-seizure count for the corresponding investigation. The performance scores of event-based investigation for both 5-fold and LOO analyses based on SW data splitting. The performance scores are presented with the Sen score on a scale of percentages and FDR score on a scale of missed event counts per hour. Rows with increasing intensities of olive colour indicate the lower performances for the different subjects (for the corresponding experiments). The overall values for the event count are calculated by adding the event values, and the performance scores are calculated by averaging the score values.

subj	event count				performance score					
	total		5-fold		LOO		5-fold		LOO	
	Sz	Nsz	TP	TN	TP	TN	Sen (%)	FDR (h)	Sen (%)	FDR (/h)
1	7	8	6	5	6	4	85.71	0.37	85.71	1.06
2	3	4	3	1	3	1	100.00	1.56	100.00	2.70
3	7	8	7	2	7	4	100.00	0.61	100.00	1.10
4	4	5	4	1	4	1	100.00	2.55	100.00	3.97
5	5	6	5	0	5	0	100.00	14.56	100.00	15.03
6	10	11	9	1	9	0	90.00	14.82	90.00	14.15
7	3	4	3	0	3	0	100.00	2.64	100.00	2.36
8	5	6	4	1	4	1	80.00	16.10	80.00	10.80
9	4	5	4	0	4	0	100.00	11.30	100.00	12.51
10	7	8	7	0	7	0	100.00	9.12	100.00	8.20
11	3	4	3	1	3	1	100.00	1.15	100.00	1.10
12	27	28	1	23	1	20	3.70	2.80	3.70	3.63
13	10	11	4	0	3	2	40.00	15.27	30.00	9.82
14	8	9	2	0	2	0	25.00	12.81	25.00	3.81
15	20	21	0	16	0	21	0.00	0.59	0.00	0.00
16	2	3	1	0	1	0	50.00	9.89	50.00	10.10
17	3	4	2	2	2	2	66.67	1.90	66.67	2.20
18	6	7	5	5	5	5	83.33	0.20	83.33	0.38
19	3	4	3	0	3	1	100.00	0.52	100.00	0.31
20	8	9	0	7	4	6	0.00	2.61	50.00	3.26
21	4	5	1	3	2	2	25.00	0.12	50.00	0.34
22	3	4	3	0	3	0	100.00	1.32	100.00	2.84
23	7	8	7	3	7	3	100.00	0.49	100.00	0.38
24	16	17	15	8	15	6	93.75	4.32	93.75	4.93
overall	175	199	99	79	103	80	72.63	5.32	75.34	4.79

5-fold cross-subject analysis, and the other is the SW LOO cross-subject analysis. For each analysis, the observation was done from an event perspective, and the performance score was presented for the individual subjects. The event-based performance scores in terms of Sen and FDR are presented in figure 5. The corresponding detailed numeric results are tabulated in table 2. The SW 5-fold and LOO results were shown in both of the presentations. As can be observed, both of the analyses exhibit almost similar results for each subject. Events were appropriately detected similarly, indicating that the data leakage was prevented well for both of these analyses. Both the SW 5-fold and LOO analyses detected the seizure events in a similar fashion as described below. For 18 subjects, both analyses produced the same Sen scores with slightly different FDR scores. Both achieved 100% Sen scores for 11 out of 24 subjects. For the other 5 subjects, both cross-subject analyses achieved more than 80% of the Sen score. The SW LOO testing produced comparatively better output for all subjects except for subjects 13 and 21 than the 5-fold analysis.

The 5-fold analysis failed to detect seizure events for subject 21. However, both analyses failed to detect any seizure event for subject 15. A similar scenario is found for both analyses in which only one seizure event was properly detected for subject 12 with a lower Sen score. The exact seizure and non-seizure counts in event-based analysis with the corresponding detection results are also presented in [table 2](#).

With these per-subject results, this study achieved the average Sen scores of 72.63% and 75.34% for SWE 5-fold and LOO analyses, respectively. The standard deviations of the results are 36.58 and 32.86 for the Sen score for the respective analyses. Average FDR scores of 5.32/h and 4.79/h were generated for SWE 5-fold and LOO analyses, respectively. Subject numbers 5, 6, 8, 9, 13, 14 and 16 produced higher FDR scores in both analyses. The best performance was reported for subject 23 with 100% Sen and the lowest FDR score of 0.49/h for the 5-fold analysis. Conversely, the best performance for SWE analysis was for subject 19, with 100% Sen and an FDR score of 0.31/h. Overall, the results indicate that the Sen scores are comparatively lower for subjects 12–15, and the model also struggled with subjects 20 and 21. Other than this, the result shows a moderately fair overall performance for this type of experimental setting that considers all the critical parameters or factors.

4. Discussion

The results show that the proposed method is capable of detecting ES with great efficiency considering the real-world situation. The results have shown the perfect Sen scores for most subjects except for a few (2–7) subjects out of 24. A tremendous performance improvement was observed in some studies if these subjects were excluded [44,45,47]. It is very challenging to achieve outstanding performance in this study setting for the following reasons, which coincide with the main contributions of this study:

- The dataset is highly imbalanced in terms of classes, and all data were considered for testing. This is related to the large data imbalance and entire data consideration. This also reflects the scenario of the continuous EEG data. (Contribution 2)
- The higher inter-subject variability of the signal is there in the EEG signal. This is linked to the subjective data leakage problem and the need for cross-subject validation. (Contribution 3)
- The event-based investigation has to deal with continuous segment detection, and there are fewer seizure events in long continuous scalp EEG. This is coupled with real event detection rather than segments. (Contribution 4)

Since these problems are directly associated with this study's contributions and important factors, it is better to discuss the results from these perspectives. Most DL methods can usually achieve better performance; however, the key insights of the seizure detection process remain a mystery [37,42,82]. In ML-based classification tasks, the test data is usually expected to be unrevealed to the model during training. However, the test data is expected to have some patterns similar to the training data. The highly fluctuating subjective results (Sen and FDR scores) echo the concept of inter-subject variability of the EEG signal, as proven by previous studies. However, the method applied in this study captures most of the variations of the seizure signals. Thus, near-perfect Sen scores have been achieved for most of the subjects. The less-performing subjects might have some rare seizure patterns that were never encountered during training using other subjects. These subjects need further intensive investigation in the future. Since this study perfectly tackles the subjective data leakage problem, the model learns no rare patterns for a single and specific subject. Another notable factor is that there is not only data imbalance in this dataset but also seizure event imbalance across subjects. Some subjects have fewer seizures compared with other subjects. Conventionally, when these subjects are in the training set, the model has less opportunity to learn the seizure pattern, which leads to lower Sen scores and more misclassification.

On the other hand, the higher FDR rate is due to such a high imbalance in data because the model has limited accurate knowledge about the pattern of seizure and non-seizure events. In addition, there are very long-duration non-seizure portions in the test sets, all the patterns of which never appeared in the training set due to data balancing in training sets. The model could not learn these completely unknown patterns. Hence, the model became confused and misclassified more, resulting in a large number of FDR scores. This type of phenomenon was also observed in another study for the CHB-MIT dataset [87]. Another factor is that during post-processing of the data for event detection, the minimum gaps between two seizures or two non-seizures were kept to 10 s, as this was found to be the standard

length in previous studies. The model might have detected some short-duration seizure events apart from each other for long-duration non-seizure portions and was discarded in the post-processing phase. In the same single non-seizure portion, the post-processing method might have detected two or more false seizures for some short duration in many situations. All these reasons can also largely increase the FDR.

Some recent and highly related studies are available for seizure detection based on the CHB-MIT dataset. Those studies are compared with the proposed study, and the comparison is presented in [table 3](#). In this part, we only focus on how efficiently the models can detect seizure events and how often they miss the original events. It is not possible to compare this study directly with previous studies due to the differences in the experimental parameter choice. The parameters include data, the nature of the analysis, the target of the research and other criteria, which are directly related to the factors we discussed in §1. The comparison table lists all the studies that considered those different factors, criteria or parameters for analysis. The studies in the table are listed based on feature explainability, whether the complete dataset was tested, whether the cross-subject versus fold-based analysis or subject-specific analysis was done, and whether event-based or segment-based analysis was undertaken.

The comparison table shows that only a few of them investigated the event-based seizure classification or seizure onset detection [17,18,41,42,62,64]. Some event detection-based studies suffered greatly from achieving higher performance scores, especially the Sen scores [62,64]. It is clear from the table that the overall performances of the models that did cross-subject analysis are not very high compared with the other studies that had considered different analysis approaches [41,57–59,81]. Significant performance degradation can be noticed in cross-subject analysis in many studies [41,57,58]. A performance drop is expected when considering both criteria (cross-subject and event-based) in analysis. Only one study was found in the literature that dealt with both of these issues [41].

Additionally, more criteria for considering the robustness of the model (using the entire dataset) put the model under more challenges, and the model struggles significantly to score better. The data imbalance problem, which reflects the real-world scenario of continuous EEG data, pushes the model into harder situations, where it faces unknown patterns not learned during its training phase. This is one of the main reasons why many studies have not explored this side. To test the model's performance, almost all the studies considered part of the data. The data selection for the studies was done using one of several methods, such as partial subject selection, considering only seizure records, filtering, under-sampling, modification or reduction in data channels, etc.

The results of this study and the summary of the literature ([table 3](#)) show it is more challenging to implement an ES detection system in a real-world scenario if all the factors are considered. The segment-based results shown in §3 also align with [table 3](#). They show the difficulty of achieving better performances (Sen) when the cross-subject analysis is undertaken for a generalized seizure detection system. The summary table also shows the trend of significant performance reduction when event detection is done. Combining these criteria (cross-subject analysis and event detection) with considering the entire dataset for testing adds additional complexity to the ML model. Thus, when adding more real-world factors to the models, the ML models generally tend to achieve relatively less seizure detection performance (Sen). This finding opens up a broader scope of exploration for other researchers to tackle these challenges and develop a more realistic solution. However, proper modelling of the problem might help researchers develop a better-performing and generalized solution.

5. Limitations and future direction

This study focuses mainly on different real-world factors that need to be reflected in the analysis based on recent related studies. The analytical implementation was simplified only to observe the impact of those factors. However, there is scope to extend the observation further to investigate the effects of different parameters. Various preprocessing methods and window lengths can be observed for seizure detection using a similar setup. Different features might be able to capture more meaningful information from different data lengths of the original signal, and extracting features from fixed-length data can sometimes be less efficient. Thus, independent observation of feature extraction with variable data length can be investigated in seizure detection. In-depth statistical analysis to reveal the underlying relationship can also be useful in seizure detection. Other different ML models can also be compared with test the generalizability of a model for this experimental setting. Different hand-crafted features,

Table 3. Summary of the performance scores and the factors considered during ES detection on CHB-MIT dataset. This table includes different ML and DL methods individually or combined with other methods such as preprocessing and post-processing. In this comparison of the entire ES detection process, some criteria are considered, such as the explainability of features, changes in data (considered fewer data, under-sampling, training over-sampling, channel alteration, filtering, etc.), cross-subject testing (to avoid data leakage) and event (or onset or segment) detection. Finally, the performance scores are presented in the last columns in terms of Sen and FDR where applicable. The tick mark represents that the criterion is applied/available, and the cross mark represents that the criterion is not applied or available. NR represents that the corresponding data are not reported. The numbers associated with the column headings indicate the corresponding contribution that this study covers.

study	feature explainability (1)	entire dataset (2)	cross-subject analysis (3)	event-based detection (4)	performance scores	
					Sen (%)	FDR (/h)
Abdelhameed <i>et al.</i> [82]	×	×	×	×	98.72	×
Gao <i>et al.</i> [37]	×	×	×	×	98.72	×
Amiri <i>et al.</i> [88]	✓	×	×	×	98.44	×
Sun <i>et al.</i> [45]	✓	×	×	×	96.79	×
Shen <i>et al.</i> [44]	✓	×	×	×	96.15	×
Jiang <i>et al.</i> [68]	✓	×	×	×	98.71	×
Zeng <i>et al.</i> [63]	✓	×	×	×	96.98	×
Bhattacharyya <i>et al.</i> [47]	✓	×	×	×	97.1	×
Wang <i>et al.</i> [42]	×	×	×	✓	99.31	0.2
Li <i>et al.</i> [43]	✓	×	×	✓	98.47	0.63
Vidyaratne <i>et al.</i> [18]	✓	×	×	✓	97.00	0.10
Boonyakitanont <i>et al.</i> [62]	✓	NR	×	✓	76.54	0.09
Pale <i>et al.</i> [64]	✓	NR	×	✓	34.10	0.05
Raghu <i>et al.</i> [40]	✓	NR	✓	NR	97.28	0.57
Wei <i>et al.</i> [41]	×	×	✓	✓	90.57	×
Hossain <i>et al.</i> [51]	×	×	✓	×	90.00	×
Thodorof <i>et al.</i> (2016) [81]	×	×	✓	×	85.00	0.80
Zhou <i>et al.</i> [59]	✓	×	✓	×	84.67	×
Wu <i>et al.</i> [60]	✓	×	✓	×	82.98	0.57
Zhao <i>et al.</i> [58]	×	×	✓	×	77.42	×
Wei <i>et al.</i> [41]	×	×	✓	×	72.11	×
Jana <i>et al.</i> [57]	×	×	✓	NR	55.63	×
proposed study (SWE 5-fold)	✓	✓	✓	✓	72.63	5.32
proposed study (SWE L00)	✓	✓	✓	✓	75.34	4.79

especially those from the wavelet domain, can also be applied and tested to investigate their efficacy in seizure event detection. On the other hand, the DL models can also be observed for this purpose if the feature explainability is not essential. The post-processing method can also be improved by implementing different thresholds for event gaps and various techniques for seizure event estimation. Additional knowledge from post-processing might help produce other fundamental findings that could reduce dependence on post-processing in the future. Since this is a multi-channel analysis, it would be interesting to observe the same for individual channels or some selected channels. Since evenly balanced data was used to train the model, that might not be enough to learn more different non-seizure patterns. It would be great to observe the same result for slightly imbalanced data. This experiment is limited to CHB-MIT dataset; observing the event detection efficacy across the different datasets would be great. This study was limited to a specific scope for simplicity and focused only on the criteria defined in §1. Some of these unexplored areas are our target to observe in the near future.

6. Conclusion

In this study, we investigated the effect of different important analytical factors altogether in ES detection using the CHB-MIT dataset. The factors include (i) considering the entire dataset for testing; (ii) subjective data leakage prevention using cross-subject and developing a generalized model; (iii) detecting ES events; and (iv) using explainable features. To the best of our knowledge, this study considers all these four crucial factors for the first time in ES detection. A total of 92 hand-crafted features from 22 scalp EEG channels were extracted. Using chi-squared, 32 selected features from each channel were used to detect seizure events by the RF model. The analyses are categorized mainly into segment- and event-based observations. Segment-based results show that subjective variation and similarity play a vital role in seizure detection. Hence, the RS 5-fold approach achieved a Sen score of 91.88% for seizure detection, which is significantly higher than the cross-subject performances. On the other hand, the other two SW cross-subject analysis approaches, namely, 5-fold and LOO observations, were able to achieve 64.24 and 66.75%, respectively. Cross-subject segment-based results were used to calculate event-based results using a post-processing method. For the event-based analyses, the cross-subject 5-fold and LOO approaches performed similarly with the Sen scores of 72.63 and 75.34%, respectively. Only a few subjects were found for which the model performed less efficiently, leading to the FDR of 5.32/h and 4.79/h for the respective analysis. Both cross-subject experiments (5-fold and LOO) perform very similarly, with the LOO obtaining a little better performance because of the model trained on more data.

Considering all the factors mentioned above that simulate a real-world environment, the overall performance obtained in this study is more realistic and fair. The experimental results and the trend of the literature show that developing a generalized seizure event detection system with the above-mentioned necessary criteria is more challenging. The findings of this study may help guide real-world application-oriented research a step further that can be used to develop devices for actual clinical applications.

Ethics. This work did not require ethical approval from a human subject or animal welfare committee.

Data accessibility. Used data are publicly available on Physionet: <https://doi.org/10.13026/C2K01R> [89,90], under license: Open Data Commons Attribution License v1.0: <https://www.physionet.org/content/chbmit/view-license/1.0.0/>. According to the License terms in section 3.0 'Rights granted' subsections—a, b, d and e; the following rights are granted to the users: Extraction and re-utilization; creation of derivative databases; creation of temporary or permanent reproductions; distribution, communication, display, lending, making available or performance to the public by any means and in any form, in whole or in part.

Electronic supplementary material is available online at [91].

Declaration of AI use. We have not used AI-assisted technologies in creating this article.

Authors' contributions. E.A.: conceptualization, data curation, formal analysis, investigation, methodology, resources, software, visualization, writing—original draft; M.N.A.: supervision, writing—review and editing; C.K.: conceptualization, supervision, validation, writing—review and editing.

All authors gave final approval for publication and agreed to be held accountable for the work performed therein.

Conflict of interest declaration. We declare we have no competing interests.

Funding. No funding has been received for the research and publication of this article.

Acknowledgements. We acknowledge the authors of 'CHB-MIT' dataset, Shoeb et al. [24] for the data and making them available for public usage. We also acknowledge the creators and facilitators of the icons used in the diagrams of this article: <https://www.flaticon.com>.

References

1. Giourou E, Stavropoulou-Deli A, Giannakopoulou A, Kostopoulos GK, Koutroumanidis M. 2015 Introduction to epilepsy and related brain disorders. In *Cyberphysical systems for epilepsy and related brain disorders* (eds NS Voros, CP Antonopoulos), pp. 11–38. Cham: Springer International Publishing. (doi:10.1007/978-3-319-20049-1)
2. Fisher RS *et al.* 2014 ILAE official report: a practical clinical definition of epilepsy. *Epilepsia* **55**, 475–482. (doi:10.1111/epi.12550)
3. Falco-Walter JJ, Scheffer IE, Fisher RS. 2018 The new definition and classification of seizures and epilepsy. *Epilepsy Res.* **139**, 73–79. (doi:10.1016/j.eplepsyres.2017.11.015)
4. French JA. 2007 Refractory epilepsy: clinical overview. *Epilepsia* **48**, 3–7. (doi:10.1111/j.1528-1167.2007.00992.x)
5. Shorvon S, Perucca E, Engel J. 2009 *The treatment of epilepsy*. 3rd ed. Chichester, UK: Wiley.

6. England MJ, Liverman CT, Schultz AM, Strawbridge LM. 2012 Summary: a reprint from epilepsy across the spectrum: promoting health and understanding. *Epilepsy Curr.* **12**, 245–253. (doi:10.5698/1535-7511-12.6.245)
7. Sen A, Jette N, Husain M, Sander JW. 2020 Epilepsy in older people. *Lancet* **395**, 735–748. (doi:10.1016/S0140-6736(19)33064-8)
8. Kusmakar S, Karmakar C, Yan B, Obrien T, Muthuganapathy R, Palaniswami M. 2019 Automated detection of convulsive seizures using a wearable accelerometer device. *IEEE Trans. Biomed. Eng.* **66**, 421–432. (doi:10.1109/TBME.2018.2845865)
9. Poh MZ, Loddenkemper T, Reinsberger C, Swenson NC, Goyal S, Sabtala MC, Madsen JR, Picard RW. 2012 Convulsive seizure detection using a wrist-worn electrodermal activity and accelerometry biosensor. *Epilepsia* **53**, e93–7. (doi:10.1111/j.1528-1167.2012.03444.x)
10. Massé F, Bussel MV, Serteyn A, Arends J, Penders J. 2013 Miniaturized wireless ECG monitor for real-time detection of epileptic seizures. *ACM Trans. Embed. Comput. Syst.* **12**, 1–21. (doi:10.1145/2485984.2485990)
11. Vandecasteele K, De Cooman T, Gu Y, Cleeren E, Claes K, Paesschen WV, Huffel SV, Hunyadi B. 2017 Automated epileptic seizure detection based on wearable ECG and PPG in a hospital environment. *Sensors (Basel)* **17**, 2338. (doi:10.3390/s17102338)
12. De Cooman T, Varon C, Hunyadi B, Van Paesschen W, Lagae L, Van Huffel S. 2017 Online automated seizure detection in temporal lobe epilepsy patients using single-lead ECG. *Int. J. Neural Syst.* **27**, 1750022. (doi:10.1142/S0129065717500228)
13. Jahanbekam A, Baumann J, Nass RD, Bauckhage C, Hill H, Elger CE, Surges R. 2021 Performance of ECG-based seizure detection algorithms strongly depends on training and test conditions. *Epilepsia Open* **6**, 597–606. (doi:10.1002/epi4.12520)
14. Karpov OE, Grubov VV, Maksimenko VA, Kurkin SA, Smirnov NM, Utyashev NP, Andrikov DA, Shusharina NN, Hramov AE. 2022 Extreme value theory inspires explainable machine learning approach for seizure detection. *Sci. Rep.* **12**, 11474. (doi:10.1038/s41598-022-15675-9)
15. Wu J, Zhou T, Li T. 2020 Detecting epileptic seizures in EEG signals with complementary ensemble empirical mode decomposition and extreme gradient boosting. *Entropy* **22**, 140. (doi:10.3390/e22020140)
16. Shoeibi A, Ghassemi N, Alizadehsani R, Rouhani M, Hosseini-Nejad H, Khosravi A, Panahiazar M, Nahavandi S. 2021 A comprehensive comparison of handcrafted features and convolutional autoencoders for epileptic seizures detection in EEG signals. *Expert Syst. Appl.* **163**, 113788. (doi:10.1016/j.eswa.2020.113788)
17. Li X, Han Z, Zhao T, Gao J. 2021 Online model for indoor temperature control based on building thermal process of air conditioning system. *J. Building Eng.* **39**, 102270. (doi:10.1016/j.job.2021.102270)
18. Vidyaratne LS, Iftekharruddin KM. 2017 Real-time epileptic seizure detection using EEG. *IEEE Trans. Neural Syst. Rehabil. Eng.* **25**, 2146–2156. (doi:10.1109/TNSRE.2017.2697920)
19. King MA, Newton MR, Jackson GD, Fitt GJ, Mitchell LA, Silvapulle MJ, Berkovic SF. 1998 Epileptology of the first-seizure presentation: a clinical, electroencephalographic, and magnetic resonance imaging study of 300 consecutive patients. *The Lancet* **352**, 1007–1011. (doi:10.1016/S0140-6736(98)03543-0)
20. Salmenpera TM, Duncan JS. 2005 Imaging in epilepsy. *J. Neurol. Neurosurg. Psychiatry* **76**, iii2–iii10. (doi:10.1136/jnnp.2005.075135)
21. Subasi A, Ismail Gursoy M. 2010 EEG signal classification using PCA, ICA, LDA and support vector machines. *Expert Syst. Appl.* **37**, 8659–8666. (doi:10.1016/j.eswa.2010.06.065)
22. Sivaraju A *et al.* 2022 Factors predicting outcome after intracranial EEG evaluation in patients with medically refractory epilepsy. *Neurology* **99**, e1–e10. (doi:10.1212/WNL.0000000000200569)
23. Flanary J *et al.* 2023 Reliability of visual review of intracranial electroencephalogram in identifying the seizure onset zone: a systematic review and implications for the accuracy of automated methods. *Epilepsia* **64**, 6–16. (doi:10.1111/epi.17446)
24. Shoeb AH. 2003 Patient-specific seizure onset detection [Bachelor of engineering thesis]. Massachusetts Institute of Technology.
25. Guleria P, Sood M. 2023 Explainable AI and machine learning: performance evaluation and explainability of classifiers on educational data mining inspired career counseling. *Educ. Inf. Technol. (Dordr)* **28**, 1081–1116. (doi:10.1007/s10639-022-11221-2)
26. Czerwinska U. 2022 Interpretability of machine learning models: how can one explain machine learning models? In *Applied data science in tourism* (ed. R Egger), pp. 275–303. Cham: Springer International Publishing, (Tourism on the Verge). See https://link.springer.com/10.1007/978-3-030-88389-8_14.
27. Sharma M, Shah S, Achuth PV. 2019 A novel approach for epilepsy detection using time–frequency localized bi-orthogonal wavelet filter. *J. Mech. Med. Biol.* **19**, 1940007. (doi:10.1142/S0219519419400074)
28. Mirowski PW, LeCun Y, Madhavan D, Kuzniecky R. 2008 Comparing SVM and Convolutional networks for epileptic seizure prediction from intracranial EEG. in: 2008 IEEE workshop on machine learning for signal processing. pp. 244–249 Cancun, Mexico: IEEE. See <http://ieeexplore.ieee.org/document/4685487/>
29. Sackellares JC, Shiao DS, Principe JC, Yang MCK, Dance LK, Suharitdamrong W, Chaovalitwongse W, Pardalos PM, Iasemidis LD. 2006 Predictability analysis for an automated seizure prediction algorithm. *J. Clin. Neurophysiol.* **23**, 509–520. (doi:10.1097/00004691-200612000-00003)
30. Yuan Q, Zhou W, Liu Y, Wang J. 2012 Epileptic seizure detection with linear and nonlinear features. *Epilepsy Behav.* **24**, 415–421. (doi:10.1016/j.yebeh.2012.05.009)
31. Kim T, Nguyen P, Pham N, Bui N, Truong H, Ha S, Vu T. 2020 Epileptic seizure detection and experimental treatment: a review. *Front. Neurol.* **11**, 701. (doi:10.3389/fneur.2020.00701)
32. Dastgoshadeh M, Rabiéi Z. 2022 Detection of epileptic seizures through EEG signals using entropy features and ensemble learning. *Front. Hum. Neurosci.* **16**, 1084061. (doi:10.3389/fnhum.2022.1084061)
33. Ilias L, Askounis D, Psarras J. 2023 Multimodal detection of epilepsy with deep neural networks. *Expert Syst. Appl.* **213**, 119010. (doi:10.1016/j.eswa.2022.119010)

34. Tran LV, Tran HM, Le TM, Huynh TTM, Tran HT, Dao SVT. 2022 Application of machine learning in epileptic seizure detection. *Diagnostics (Basel)* **12**, 2879. (doi:10.3390/diagnostics12112879)
35. Zhang T, Han Z, Chen X, Chen W. 2021 Subbands and cumulative sum of subbands based nonlinear features enhance the performance of epileptic seizure detection. *Biomed. Signal Process. Control* **69**, 102827. (doi:10.1016/j.bspc.2021.102827)
36. Aghababaei MH, Azemi G, O'Toole JM. 2021 Detection of epileptic seizures from compressively sensed EEG signals for wireless body area networks. *Expert Syst. Appl.* **172**, 114630. (doi:10.1016/j.eswa.2021.114630)
37. Gao X, Yan X, Gao P, Gao X, Zhang S. 2020 Automatic detection of epileptic seizure based on approximate entropy, recurrence quantification analysis and convolutional neural networks. *Artif. Intell. Med.* **102**, 101711. (doi:10.1016/j.artmed.2019.101711)
38. D.k. T, B.g. P, Xiong F. 2019 Epileptic seizure detection and prediction using stacked bidirectional long short term memory. *Patt. Recogn. Lett.* **128**, 529–535. (doi:10.1016/j.patrec.2019.10.034)
39. Wang Y et al. 2022 SEEG-Net: an explainable and deep learning-based cross-subject pathological activity detection method for drug-resistant epilepsy. *Comput. Biol. Med.* **148**, 105703. (doi:10.1016/j.combiomed.2022.105703)
40. Raghu S, Sriraam N, Vasudeva Rao S, Hegde AS, Kubben PL. 2020 Automated detection of epileptic seizures using successive decomposition index and support vector machine classifier in long-term EEG. *Neural Comput. Applic.* **32**, 8965–8984. (doi:10.1007/s00521-019-04389-1)
41. Wei Z, Zou J, Zhang J, Xu J. 2019 Automatic epileptic EEG detection using convolutional neural network with improvements in time-domain. *Biomed. Signal Process. Control* **53**, 101551. (doi:10.1016/j.bspc.2019.04.028)
42. Wang X, Wang X, Liu W, Chang Z, Kärkkäinen T, Cong F. 2021 One dimensional convolutional neural networks for seizure onset detection using long-term scalp and intracranial EEG. *Neurocomputing* **459**, 212–222. (doi:10.1016/j.neucom.2021.06.048)
43. Li G, Jung JJ. 2021 Seizure detection from multi-channel EEG using entropy-based dynamic graph embedding. *Artif. Intell. Med.* **122**, 102201. (doi:10.1016/j.artmed.2021.102201)
44. Shen M, Wen P, Song B, Li Y. 2022 An EEG based real-time epilepsy seizure detection approach using discrete wavelet transform and machine learning methods. *Biomed. Signal Process. Control* **77**, 103820. (doi:10.1016/j.bspc.2022.103820)
45. Sun Y, Chen X. 2022 Automatic detection of epilepsy based on entropy feature fusion and convolutional neural network. *Oxid. Med. Cell. Longev.* **2022**, 1322826. (doi:10.1155/2022/1322826)
46. Jiang Y, Chen W, You Y. 2020 Scattering transform-based features for the automatic seizure detection. *Biocyber. Biomed. Eng.* **40**, 77–89. (doi:10.1016/j.bbe.2019.11.002)
47. Bhattacharyya A, Pachori RB. 2017 A multivariate approach for patient-specific EEG seizure detection using empirical wavelet transform. *IEEE Trans. Biomed. Eng.* **64**, 2003–2015. (doi:10.1109/TBME.2017.2650259)
48. Myers MH, Padmanabha A, Hossain G, de Jongh Curry AL, Blaha CD. 2016 Seizure prediction and detection via phase and amplitude lock values. *Front. Hum. Neurosci.* **10**, 80. (doi:10.3389/fnhum.2016.00080)
49. Liu X, Wang J, Shang J, Liu J, Dai L, Yuan S. 2022 Epileptic seizure detection based on variational mode decomposition and deep forest using EEG Signals. *Brain Sci.* **12**, 1275. (doi:10.3390/brainsci12101275)
50. Emami A, Kunii N, Matsuo T, Shinozaki T, Kawai K, Takahashi H. 2019 Seizure detection by convolutional neural network-based analysis of scalp electroencephalography plot images. *Neuroimage. Clin.* **22**, 101684. (doi:10.1016/j.nicl.2019.101684)
51. Hossain MS, Amin SU, Alsulaiman M, Muhammad G. 2019 Applying deep learning for epilepsy seizure detection and brain mapping visualization. *ACM Trans. Multimedia Comput. Commun. Appl.* **15**, 1–17. (doi:10.1145/3241056)
52. Alotaiby TN, El-Samie FEA, Alshebeili SA, Alotaibi FM, Aljibrin K, Alrshod SR, Alkhanin IM, Alrajhi N. 2016 A common spatial pattern approach for scalp EEG seizure detection. in: 2016 5th international conference on electronic devices, systems and applications (ICEDSA). pp. 1–5. Ras Al Khaimah: IEEE. See <http://ieeexplore.ieee.org/document/7818564>.
53. Robbins KA, Touryan J, Mullen T, Kothe C, Bigdely-Shamlo N. 2020 How sensitive are EEG results to preprocessing methods: a benchmarking study. *IEEE Trans. Neural Syst. Rehabil. Eng.* **28**, 1081–1090. (doi:10.1109/TNSRE.2020.2980223)
54. Natarajan K, Acharya U R, Alias F, Tiboleng T, Puthusserypady SK. 2004 Nonlinear analysis of EEG signals at different mental states. *Biomed. Eng. Online* **3**, 7. (doi:10.1186/1475-925X-3-7)
55. Nasiri S, Clifford GD. 2021 Generalizable seizure detection model using generating transferable adversarial features. *IEEE Signal Process. Lett.* **28**, 568–572. (doi:10.1109/LSP.2021.3060967)
56. Bolagh SNG, Clifford GD. 2017 Subject selection on a riemannian manifold for unsupervised cross-subject seizure detection. arXiv. See <https://arxiv.org/abs/1712.00465>.
57. Jana GC, Swami K, Agrawal A. 2023 Capsule neural network based approach for subject specific and cross-subjects seizure detection from EEG signals. *Multimed. Tools Appl.* **82**, 35221–35252. (doi:10.1007/s11042-023-14995-w)
58. Zhao Y, Zhang G, Zhang Y, Xiao T, Wang Z, Xu F, Zheng Y. 2022 Multi-view cross-subject seizure detection with information bottleneck attribution. *J. Neural Eng.* **19**, 046011. (doi:10.1088/1741-2552/ac7d0d)
59. Zhou J, Liu L, Leng Y, Yang Y, Gao B, Jiang Z, Nie W, Yuan Q. 2022 Both cross-patient and patient-specific seizure detection based on self-organizing fuzzy logic. *Int. J. Neural Syst.* **32**, 2250017. (doi:10.1142/S0129065722500174)
60. Wu D, Wang Z, Jiang L, Dong F, Wu X, Wang S, Ding Y. 2019 Automatic epileptic seizures joint detection algorithm based on improved multi-domain feature of cEEG and spike feature of aEEG. *IEEE Access* **7**, 41551–41564. (doi:10.1109/ACCESS.2019.2904949)
61. Li X, Han Z, Zhao T, Zhang J, Xue D. 2021 Modeling for indoor temperature prediction based on time-delay and Elman neural network in air conditioning system. *J. Building Eng.* **33**, 101854. (doi:10.1016/j.job.2020.101854)

62. Boonyakitanont P, Lek-uthai A, Chomtho K, Songsiri J. 2020 A review of feature extraction and performance evaluation in epileptic seizure detection using EEG. *Biomed. Signal Process. Control* **57**, 101702. (doi:10.1016/j.bspc.2019.101702)
63. Zeng J, Tan X dan, Zhan CA. 2021 Automatic detection of epileptic seizure events using the time-frequency features and machine learning. *Biomed. Signal Process. Control* **69**, 102916. (doi:10.1016/j.bspc.2021.102916)
64. Pale U, Teijeiro T, Atienza D. 2023 Importance of methodological choices in data manipulation for validating epileptic seizure detection models. In: 2023 45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). pp. 1–7 Sydney, Australia: IEEE. See <https://ieeexplore.ieee.org/document/10340493>
65. Gerster M, Taher H, Škoch A, Hlinka J, Guye M, Bartolomei F, Jirsa V, Zakharova A, Olmi S. 2021 Patient-Specific network connectivity combined with a next generation neural mass model to test clinical hypothesis of seizure propagation. *Front. Syst. Neurosci.* **15**, 675272. (doi:10.3389/fnsys.2021.675272)
66. Zanetti R, Pale U, Teijeiro T, Atienza D. 2022 Approximate zero-crossing: a new interpretable, highly discriminative and low-complexity feature for EEG and iEEG seizure detection. *J. Neural Eng.* **19**, 066018. (doi:10.1088/1741-2552/aca1e4)
67. Prasanna J, Subathra MSP, Mohammed MA, Damaševičius R, Sairamya NJ, George ST. 2021 Automated epileptic seizure detection in pediatric subjects of CHB-MIT EEG database—a survey. *J. Pers. Med.* **11**, 1028. (doi:10.3390/jpm11101028)
68. Jiang Y, Chen W, Li M, Zhang T, You Y. 2021 Synchroextracting chirplet transform-based epileptic seizures detection using EEG. *Biomed. Signal Process. Control* **68**, 102699. (doi:10.1016/j.bspc.2021.102699)
69. Yuan J *et al.* 2020 A novel model for improving the prediction accuracy of the new heating station. *Energy Build.* **229**, 110521. (doi:10.1016/j.enbuild.2020.110521)
70. Aung ST, Wongsawat Y. 2020 Modified-distribution entropy as the features for the detection of epileptic seizures. *Front. Physiol.* **11**, 607. (doi:10.3389/fphys.2020.00607)
71. Mormann F, Andrzejak RG, Elger CE, Lehnertz K. 2007 Seizure prediction: the long and winding road. *Brain* **130**, 314–333. (doi:10.1093/brain/awl241)
72. Paul Y. 2018 Various epileptic seizure detection techniques using biomedical signals: a review. *Brain Inform.* **5**, 6. (doi:10.1186/s40708-018-0084-z)
73. Kusmakar S, Karmakar CK, Yan B, O'Brien TJ, Muthuganapathy R, Palaniswami M. 2017 Detection of generalized tonic-clonic seizures using short length accelerometry signal. In 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Seogwipo, pp. 4566–4569. Seogwipo: IEEE. <https://ieeexplore.ieee.org/document/8037872/>.
74. Alzaha A, Elgammal M, Mohammed H, Mostafa H. 2019 In *Optimal EEG Window Size for Neural Seizure Detection*. In: 2019 8th International Conference on Modern Circuits and Systems Technologies (MOCASST), pp. 1–4. Thessaloniki, Greece: IEEE. <https://ieeexplore.ieee.org/document/8741714>.
75. Mormann F, Kreuz T, Rieke C, Andrzejak RG, Kraskov A, David P, Elger CE, Lehnertz K. 2005 On the predictability of epileptic seizures. *Clin. Neurophysiol.* **116**, 569–587. (doi:10.1016/j.clinph.2004.08.025)
76. Ramgopal S *et al.* 2014 Seizure detection, seizure prediction, and closed-loop warning systems in epilepsy. *Epilepsy Behav.* **37**, 291–307. (doi:10.1016/j.yebeh.2014.06.023)
77. McSharry PE, He T, Smith LA, Tarassenko L. 2002 Linear and non-linear methods for automatic seizure detection in scalp electro-encephalogram recordings. *Med. Biol. Eng. Comput.* **40**, 447–461. (doi:10.1007/BF02345078)
78. Wang L, Xue W, Li Y, Luo M, Huang J, Cui W, Huang C. 2017 Automatic epileptic seizure detection in EEG signals using multi-domain feature extraction and nonlinear analysis. *Entropy* **19**, 222. (doi:10.3390/e19060222)
79. Acharya UR, Sree SV, Alvin APC, Yanti R, Suri JS. 2012 Application of non-linear and wavelet based features for the automated identification of epileptic eeg signals. *Int. J. Neural Syst.* **22**, 1250002. (doi:10.1142/S0129065712500025)
80. Udhayakumar RK, Karmakar C, Palaniswami M. 2018 Understanding irregularity characteristics of short-term HRV signals using sample entropy profile. *IEEE Trans. Biomed. Eng.* **65**, 2569–2579. (doi:10.1109/TBME.2018.2808271)
81. Thodoroff P, Pineau J, Lim A. 2016 Learning robust features using deep learning for automatic seizure detection. arXiv. (doi:10.48550/arXiv.1608.00220)
82. Abdelhameed A, Bayoumi M. 2021 A deep learning approach for automatic seizure detection in children with epilepsy. *Front. Comput. Neurosci.* **15**, 650050. (doi:10.3389/fncom.2021.650050)
83. Shueb A, Guttag J. 2010 Application of machine learning to epileptic seizure detection. *Int. Conf. Mach. Learn* <https://icml.cc/Conferences/2010/papers/493.pdf>
84. Temko A, Thomas E, Marnane W, Lightbody G, Boylan GB. 2011 Performance assessment for EEG-based neonatal seizure detectors. *Clin. Neurophysiol.* **122**, 474–482. (doi:10.1016/j.clinph.2010.06.035)
85. Obeid I, Selesnick I, Picone J (eds). 2021 *Biomedical signal processing: innovation and applications*. Cham: Springer International Publishing. (doi:10.1007/978-3-030-67494-6) See <https://link.springer.com/10.1007/978-3-030-67494-6>.
86. Chicco D, Jurman G. 2020 The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genomics* **21**, 6. (doi:10.1186/s12864-019-6413-7)
87. Zarei A, Asl BM. 2021 Automatic seizure detection using orthogonal matching pursuit, discrete wavelet transform, and entropy based features of EEG signals. *Comput. Biol. Med.* **131**, 104250. (doi:10.1016/j.combiomed.2021.104250)

88. Amiri M, Aghaeinia H, Amindavar HR. 2023 Automatic epileptic seizure detection in EEG signals using sparse common spatial pattern and adaptive short-time fourier transform-based synchrosqueezing transform. *Biomed. Signal Process. Control* **79**, 104022. (doi:10.1016/j.bspc.2022.104022)
89. Gutttag J. 2010 CHB-MIT Scalp EEG Database (version 1.0.0). *PhysioNet* See <https://doi.org/10.13026/C2K01R>.
90. Goldberger A, Amaral L, Glass L, Hausdorff J, Ivanov PC, Mark R, Stanley HE. 2000 PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. *Circulation* **101**, e215–e220.
91. Ali E, Angelova MN, Karmakar C. 2024 Supplementary material from: Epileptic seizure detection using CHB-MIT dataset: The overlooked perspectives. Figshare (doi:10.6084/m9.figshare.c.7184012)