Improving automated diagnosis of epilepsy from EEGs beyond IEDs

Thangavel, Prasanth; Thomas, John; Sinha, Nishant; Peh, Wei Yan; Yuvaraj, Rajamanickam; Cash, Sydney S.; Chaudhari, Rima; Karia, Sagar; Dauwels, Justin; More Authors

DOI
10.1088/1741-2552/ac9c93

Publication date
2022

Document Version
Final published version

Published in
Journal of Neural Engineering

Improving automated diagnosis of epilepsy from EEGs beyond IEDs

Journal of Neural Engineering

RECEIVED
9 November 2021

REVISED
24 September 2022

ACCEPTED FOR PUBLICATION
21 October 2022

PUBLISHED
24 November 2022

PAPER

Improving automated diagnosis of epilepsy from EEGs beyond IEDs

Prasanth Thangavel[1,12], John Thomas[2,12], Nishant Sinha[3], Wei Yan Peh[1], Rajamanickam Yuvaraj[4], Sydney S Cash[5], Rima Chaudhari[6], Sagar Karia[7], Jin Jing[5], Rahul Rathakrishnan[8], Vinay Saini[9], Nilesh Shah[7], Rohit Srivastava[9], Yee-Leng Tan[10], Brandon Westover[5] and Justin Dauwels[1,11,*]

[1] Nanyang Technological University (NTU), Singapore
[2] Montreal Neurological Institute, McGill University, Montreal, Canada
[3] University of Pennsylvania, Pennsylvania, Philadelphia, United States of America
[4] National Institute of Education, Singapore
[5] Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts, United States of America
[6] Fortis Hospital Mulund, Mumbai, India
[7] Lokmanya Tilak Municipal General Hospital, Mumbai, India
[8] National University Hospital, Singapore
[9] Department of Biosciences and Bioengineering, IIT Bombay, Mumbai, India
[10] National Neuroscience Institute, Singapore
[11] TU Delft, Delft, The Netherlands
[12] Equal contribution.
[*] Author to whom any correspondence should be addressed.

E-mail: J.H.G.Dauwels@tudelft.nl

Keywords: deep learning, electroencephalogram (EEG), EEG classification, epilepsy diagnosis, interictal epileptiform discharges (IEDs), multi-center study, wavelets

Abstract

Objective. Clinical diagnosis of epilepsy relies partially on identifying interictal epileptiform discharges (IEDs) in scalp electroencephalograms (EEGs). This process is expert-biased, tedious, and can delay the diagnosis procedure. Beyond automatically detecting IEDs, there are far fewer studies on automated methods to differentiate epileptic EEGs (potentially without IEDs) from normal EEGs. In addition, the diagnosis of epilepsy based on a single EEG tends to be low. Consequently, there is a strong need for automated systems for EEG interpretation. Traditionally, epilepsy diagnosis relies heavily on IEDs. However, since not all epileptic EEGs exhibit IEDs, it is essential to explore IED-independent EEG measures for epilepsy diagnosis. The main objective is to develop an automated system for detecting epileptic EEGs, both with or without IEDs. In order to detect epileptic EEGs without IEDs, it is crucial to include EEG features in the algorithm that are not directly related to IEDs. *Approach.* In this study, we explore the background characteristics of interictal EEG for automated and more reliable diagnosis of epilepsy. Specifically, we investigate features based on univariate temporal measures (UTMs), spectral, wavelet, Stockwell, connectivity, and graph metrics of EEGs, besides patient-related information (age and vigilance state). The evaluation is performed on a sizeable cohort of routine scalp EEGs (685 epileptic EEGs and 1229 normal EEGs) from five centers across Singapore, USA, and India. *Main results.* In comparison with the current literature, we obtained an improved Leave-One-Subject-Out (LOSO) cross-validation (CV) area under the curve (AUC) of 0.871 (Balanced Accuracy (BAC) of 80.9%) with a combination of three features (IED rate, and Daubechies and Morlet wavelets) for the classification of EEGs with IEDs vs. normal EEGs. The IED-independent feature UTM achieved a LOSO CV AUC of 0.809 (BAC of 74.4%). The inclusion of IED-independent features also helps to improve the EEG-level classification of epileptic EEGs with and without IEDs vs. normal EEGs, achieving an AUC of 0.822 (BAC of 77.6%) compared to 0.688 (BAC of 59.6%) for classification

only based on the IED rate. Specifically, the addition of IED-independent features improved the BAC by 21% in detecting epileptic EEGs that do not contain IEDs. *Significance.* These results pave the way towards automated detection of epilepsy. We are one of the first to analyze epileptic EEGs without IEDs, thereby opening up an underexplored option in epilepsy diagnosis.

1. Introduction

Epilepsy affects about 70 million worldwide [1]. It is characterized by recurrent unprovoked seizures. An important part of clinical diagnosis of epilepsy is identifying interictal epileptiform discharges (IEDs) in scalp electroencephalograms (EEGs). The traditional epileptiform EEG markers are seizures or ictal events, IEDs, and high-frequency oscillations (HFOs) [2]. Seizure-based diagnosis is often impractical owing to the relative infrequency of seizures for most epileptic patients. It is common practice to diagnose epilepsy based on the occurrence of IEDs in the EEG, however, annotating IEDs is time-consuming and subjective [3, 4]. Moreover, the yield of epilepsy diagnosis in the first EEG could be as low as 29% [5]. IEDs are not apparent in EEGs of some epileptic patients. In addition, IEDs can be confused with eye blinks and other artifacts due to their morphological resemblance [6]. HFOs are potential alternative biomarkers of epilepsy, however, the EEGs need to be recorded at a high sampling rate (at least 500 Hz). Moreover, the effectiveness of HFOs is yet to be evaluated. Slowing can be a good indicator of underlying pathology, but it is not a definitive indicator of epilepsy.

It is evident from recent studies [7–10] that IEDs can effectively classify EEGs with epilepsy from normal EEGs. In our earlier study performed on a single dataset, we have achieved an area under the curve (AUC) of 0.847 for classifying epileptic EEGs with IEDs vs. normal EEGs. [10]. In our previous multicenter studies of time-domain [7] and frequency-domain [8] features, we have achieved a mean EEG classification Leave-One-Subject-Out (LOSO) cross-validation (CV) AUC of 0.812 (balanced accuracy (BAC) of 74.8%) and 0.856 (BAC of 79.5%), respectively, on a sizable dataset of more than 2000 EEGs from 5 different centers. To the best of our knowledge, previous studies do not consider the scenarios where epileptic EEGs have no or very few IEDs. In the literature, studies have shown that spectral [7, 11, 12], connectivity or functional networks [13–16], graph metrics [17–19], wavelet transforms [20–22], Stockwell transform (ST) [23–25] etc features can classify epileptic EEGs, abnormal EEGs and mental tasks on one or more datasets. These studies have achieved EEG classification accuracy in the range of 60%–90%. The majority of these studies are evaluated only on a small cohort (less than 50 subjects) collected from a single center and have explored only a few types of features. Furthermore, many studies are performed only on handpicked EEG segments instead of the entire EEGs, failing to emulate the real-world clinical setting.

In this paper, we consider the problem of classifying epileptic EEGs vs. normal EEGs. This classification problem is relevant for epilepsy diagnosis. Two different type of features can be extracted from EEGs to address this classification problem: *(1) IED-dependent features*: For these features to be extracted, IED waveforms in epileptic EEGs needs to be manually annotated by a neurologist first. From the annotated IEDs, features (number of IEDs per minute, peak of the IED, etc) can be extracted and applied for the classification algorithm, *(2) IED-independent features*: Features such as relative power, spectrum, entropy can be computed from the EEG, without the need for expert annotated IED segments. Though these features are affected by IEDs, they do not explicitly quantify IEDs. Specifically, we investigate and compare IED-independent features such as univariate temporal features, spectral features, wavelet transform, ST, connectivity, and graph metrics features that are suggested to help diagnose epilepsy from EEG in various studies in the literature [7, 11–25]. While performing this analysis, we determine the best feature sets and their combination to distinguish epileptic EEGs vs. normal EEGs. We employ a state-of-the-art classifier, the eXtreme Gradient Boosting (XGBoost) algorithm, an implementation of the scalable gradient-boosted decision trees [26] for classification. This algorithm has been shown in the literature to be effective in the classification of epileptic signals [27], processing massive datasets [28], cloud computing [29] and many other fields of research. Other factors behind the selection of the XGBoost algorithm include its capability to efficiently classify continuous data, ability to handle missing values, low processing time, robustness against noise, ability to select variables implicitly, and efficient learning from non-linear data [30, 31]. We perform LOSO CV on each dataset separately by training the classifier on all EEGs except one, and testing it on the left out EEG, and aggregate the results finally. We evaluate different features and systems primarily with LOSO as features like spectral could be dataset specific [7]. The system for detecting epileptic EEGs should perform at least as well as existing systems for epileptic EEGs with IEDs, which rely on IED detectors. Moreover, it should also perform well for epileptic EEGs without IEDs. This is a challenging problem, since existing epileptic EEG detectors fail to detect IED-free epileptic EEGs since they usually make use of IED detectors.

Though epilepsy affects people of all ages, the person's age has a significant impact on epilepsy [25]. Children and older individuals are at greater risk of developing epilepsy than young and middle-aged people [32]. Klimes *et al* have reported that the state of vigilance has an effect in localizing the epileptogenic zone in interictal intracranial EEG [33]. However, the majority of the studies fail to investigate the effect of age and vigilance state. Therefore, we also investigate how including the patient age and vigilance state as features affects the classification performance.

The first contribution in the study is an investigation of IED-independent EEG features for classifying epileptic EEGs with IEDs vs. normal EEGs. The proposed IED-independent system achieves a maximum LOSO CV AUC of 0.809 (BAC of 74.4%) by leveraging the univariate temporal measures (UTMs). The second contribution of the study is to investigate the added value of IED-independent EEG features when combined with the IED rate for classifying epileptic EEGs with IEDs vs. normal EEGs. Our proposed ensemble system obtains a LOSO CV AUC of 0.871 (BAC of 80.9%) with a combination of three features (IED rate, and Daubechies and Morlet wavelets). Our results show that the IED detector performs well in general, and IED-independent features could further help to detect IED EEGs. The third contribution of the study is the exploration of an ensemble system for detecting epileptic without IEDs vs. normal EEGs and epileptic with and without IEDs vs. normal EEGs. Although detecting epileptic EEGs without IEDs is important for clinical practice, since epileptic EEGs do not always have IEDs, this particular classification problem is rarely studied; the only study on this matter that we were able to find is [15]. To the best of our knowledge, we could be one of the very first to investigate automated algorithms for detecting epileptic EEGs without IEDs for diagnosis of epilepsy. Our results show that adding IED-independent features vastly improves the classification compared to the classifier based only on the IED rate, but still, the accuracy could be improved in the future most probably. This shows that IEDs are a crucial biomarker for epilepsy diagnosis and also that more research is required into IED-independent features, especially for classifying IED-free epileptic EEGs. The fourth contribution of the study is to investigate the effect of age and vigilance state of the patient. Our results show that including age and vigilance state improves the overall classification performance.

The paper is organized as follows. In section 2, we include the description of the EEG dataset, preprocessing steps, the proposed EEG-level features extraction methods, and the training and evaluation methodology. We present the results for individual feature sets and a combination of feature sets in section 3. In section 4, we discuss our results, including findings for epileptic EEGs without IEDs. In section 5, we offer concluding remarks with suggestions for future research.

2. Materials and methods

The whole framework of the study is depicted in figure 1.

2.1. EEG dataset

We analyze routine scalp EEG recordings from the following centers: Massachusetts General Hospital (MGH, Boston, USA), Temple University Hospital (TUH, USA) [34], National University Hospital (NUH, Singapore), National Neuroscience Institute (NNI, Singapore), Fortis Hospital Mulund (Mumbai, India), and Lokmanya Tilak Municipal General Hospital (LTMGH, Mumbai, India). For those patients, the EEG was recorded for initial diagnosis. Their first seizure might have happened long before or during their previous clinical assessments. The EEGs were recorded at various sampling frequencies according to the International 10–20 electrode placement scheme. The EEGs were annotated by experts at the individual centers independently in typical clinical settings. We derive the EEG labels from the clinical reports of the patient's diagnosis and EEG findings reported by neurologists from the respective centers. We categorize the EEGs into epileptic EEGs (EEGs from patients diagnosed with epilepsy as per the clinical report) and normal EEGs (non-epileptic EEGs with no marked abnormality in the clinical report). From the 'normal' EEGs, we have excluded the EEGs of patients that have been reported to have other neurological and psychological conditions (such as autism, schizophrenia, etc) besides epilepsy. We have discarded the EEGs with ictal events as they might skew the background characteristics of EEG. It has also been shown in [7] that EEGs with ictal events are easier to classify in comparison with EEGs with interictal events. The patient details of the six datasets are given in table 1.

The MGH dataset contains 18 164 IEDs that are cross-annotated by two experts, and all the annotated IEDs are included in the study. The NUH, NNI, Fortis, and LTMGH datasets are recorded and evaluated during routine clinical care at the respective institutions independently. The LTMGH data is recorded by locally manufactured equipment in a non-standard environment. The EEGs contained excessive power in the delta band, most likely due to the sweat artifacts. This makes the analysis of this dataset challenging. We utilize the TUH Epilepsy corpus [35] from the TUH database [34], the largest publicly available EEG corpus. To ensure the same mean EEG length, we choose EEGs with durations ranging from 5 to 60 min. All the epileptic EEGs summarized in table 1 contain EEGs with IEDs (patients diagnosed with epilepsy whose EEGs

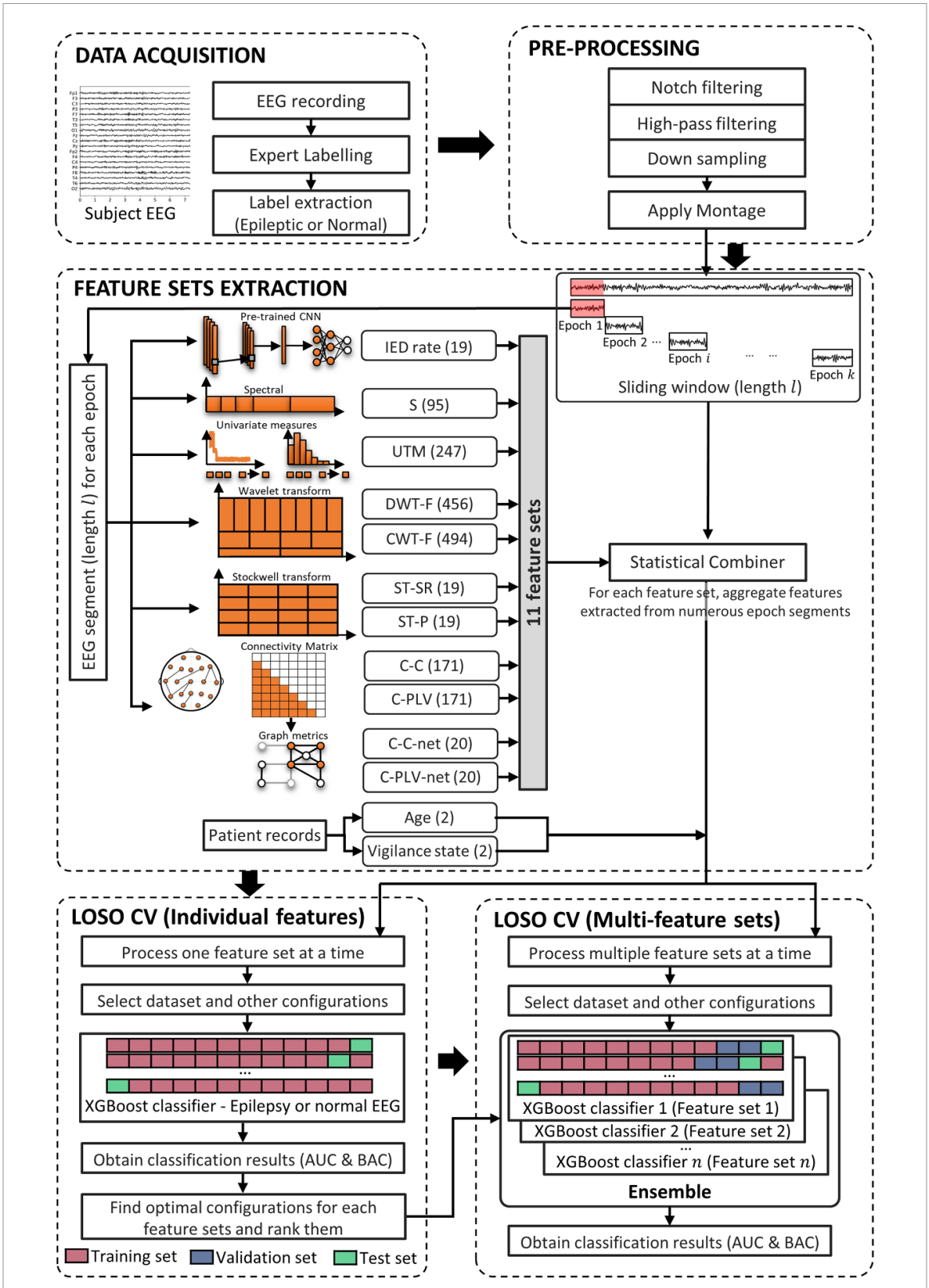

Figure 1. Overview of the EEG pipeline. The size of the feature set is reported in brackets for a 19-channel electrode configuration in CAR montage.

exhibit IED patterns as per the clinical report), with the exception of the TUH EEGs, which also comprise of epileptic EEGs without IEDs (patients diagnosed with epilepsy whose EEGs do not exhibit IED patterns as per the clinical report). Specifically, in the TUH dataset, there are 421 epileptic EEGs (from 69 patients diagnosed with epilepsy), of which 260 epileptic EEGs (42 patients) contain IEDs, and the remaining 161 epileptic EEGs (32 patients) are free of IEDs according to the clinical reports. Out of 69 epileptic patients, 5 patients have EEGs both with and without IEDs. To be consistent with the

Table 1. Patient details of the six datasets.

Dataset	Type	F_s (Hz)	Number of EEGs (Number of patients)	Age/Gender		
				Male	Female	Unknown
MGH	Epileptic	128	93 (84)	43 (35.2 ± 27.2)	41 (37.1 ± 28.2)	—
	normal		461 (461)	—	—	Adult EEGs
NUH	Epileptic	256	65 (65)	32 (50.3 ± 20.2)	33 (56.4 ± 19.7)	
	normal		99 (99)	60 (48.8 ± 17.9)	39 (50.9 ± 20.5)	—
NNI	Epileptic	200	119 (119)	55 (44.5 ± 19.0)	64 (47.0 ± 21.2)	
	normal		118 (118)	60 (44.0 ± 16.8)	58 (51.2 ± 18.4)	—
TUH	Epileptic	200, 256, 500	421 (69)	32 (50.3 ± 20.2)	33 (56.4 ± 19.7)	
	normal		44 (30)	60 (48.8 ± 17.9)	39 (50.9 ± 20.5)	—
Fortis	Epileptic	500	36 (36)	25 (37.0 ± 14.7)	10 (38.4 ± 17.4)	1 (25)
	normal		342 (342)	185 (48.5 ± 18.1)	147 (47.2 ± 17.3)	10 (41.0 ± 18.5)
LTMGH	Epileptic	256	44 (44)	26 (51.2 ± 24.3)	18 (46.3 ± 24.7)	
	normal		626 (626)	365 (41.0 ± 16.7)	261 (41.4 ± 19.0)	—
Total	Epileptic		778 (417)			
	normal		1690 (1676)			

F_s: sampling frequency, age is reported as mean ± standard deviation.
Information about the first seizure was not available in all the clinical reports.
Subjects reported with other neurological disorders are excluded from the analysis.

Table 2. Datasets investigated in this study and their corresponding institutional review board (IRB) details.

Dataset/ Institution	IRB protocol number	Date of approval
MGH	2013P001024/MGH	19 February 2016
NUH	2017/00452	22 June 2017
NNI	2017/00452	22 June 2017
TUH[a]	—	—
Fortis	IEC/2018/OAS/06	19 June 2018
LTMGH	IEC/113/17	31 January 2018

[a] Publicly availably Epilepsy EEG corpus.

other datasets, in this analysis of classifying epileptic vs. normal EEGs, we do not include epileptic EEGs without IEDs. We will, however, thoroughly investigate the epileptic EEGs without IEDs found in the TUH dataset in section 4. The datasets mainly contained adult EEGs. This research was authorized by the Review Boards of the corresponding centers (see table 2).

2.2. Pre-processing and montages

We adopt the following preprocessing pipeline: a notch filter (4th order, Butterworth) of 50/60 Hz to remove electrical artifacts, a high-pass filter (4th order) to eliminate direct current (DC) value, and baseline variations. For the current study, we down-sampled all the EEGs to 200 Hz for extracting IED-independent features, whereas we used a sampling rate of 128 Hz for extracting CNN-based IED feature (IED-dependent). In addition, we implemented artifact rejection based on noise statistics as described in [9].

EEGs tend to contain artifacts that are unrelated to cerebral activity. Some relevant sources of artifacts include muscle movement, physiological functioning, and even movement from the environment or the calibration of the electrodes. In order to limit the impact of artifacts, it is important to select a suitable electrode montage [36]. We investigate 4 different types of EEG montages, namely common average referential (CAR) [37, 38], Cz referential, Longitudinal Bipolar [36, 37], and Laplacian [38].

CAR is one of the reference-free techniques that is not affected by problems associated with an actual physical [38]. In CAR, the potential at each electrode is measured with respect to the average of all electrodes. CAR signals are computed as follows [37]:

$$V_i^{\mathrm{CAR}} = V_i^{\mathrm{ER}} - \frac{1}{n}\sum_{i=1}^{n} V_j^{\mathrm{ER}}, \tag{1}$$

where V_i^{ER} is the potential between ith electrode and the reference and n is the number of electrodes in the montage.

In the Cz montage, the Cz electrode is used as the reference. This montage is commonly used in applications that involve non-localized EEG abnormalities. However, it is less effective for localizing and identifying focal brain activity. The Cz is computed as given below:

$$V_i^{\mathrm{Cz}} = V_i^{\mathrm{ER}} - V^{\mathrm{Cz}}, \tag{2}$$

where V^{Cz} is the reference electrode.

Bipolar montages consist of chains of electrodes, each one connected to one or two neighboring electrodes. The bipolar longitudinal pattern is a widely-used montage method for its versatility [36]. It consists of a display in which each channel connects adjacent electrodes from anterior to posterior areas in two lines, essentially covering the parasagittal and temporal areas bilaterally. The midline electrodes are also linked in a chain fashion.

At last, in the Laplacian montage, the reference is computed by averaging across the nearest electrodes [37]:

$$V_i^{\text{Laplacian}} = V_i^{\text{ER}} - \frac{1}{n}\sum_{j\in Si}^{n} V_j^{\text{ER}}, \quad (3)$$

where *Si* is the set of surrounding electrodes of the *i*th electrode and *j* is a member of *Si*.

The Laplacian montage is less prone to defects or artifacts in some of the electrodes, since it does not compute the average from all electrodes. For instance, Shyam *et al* reported that Laplacian outperformed CAR in their study [38].

2.3. EEG-level features

We evaluate multiple features for EEG classification (see figure 1): IED rate [8], standard UTMs [39–41], spectral measures [7], wavelet transform [20, 22], ST [42], connectivity measures [43, 44], and graph metrics [17].

2.3.1. CNN IED-based EEG prediction

We apply a frequency-domain CNN-based IED detector for EEG classification similar to the design introduced by Thangavel *et al* in [8]. The CNN IED detector is trained only on the MGH dataset, and is implemented at the waveform level. Though the training is performed on 0.5 s non-overlapping EEG segments, we perform the testing with a sliding window of 75% overlap. For each single-channel EEG segment (0.5 s), the IED detector predicts a value between [0,1] with '1' indicating an IED. The outputs for 19 channels are combined by taking the maximum of the CNN outputs to produce the time-instant output. If there are multiple IED detections in the adjacent segments, they are considered as a single detection during post processing of the CNN outputs. Next, the IED rates (number of IEDs detected per minute) are computed for 100 thresholds (0.01, 0.02,…, 1) on the CNN output between [0,1], and the best threshold is selected by internal CV [8]. The resulting IED rates serve as features for EEG classification. This approach is the benchmark for the evaluation of the other features. Since the CNN IED detector is trained on the MGH dataset, we exclude MGH dataset from rest of the EEG classification analysis.

2.3.2. UTMs

We investigate standard time-domain EEG features to support the analysis [39]: mean, median, standard deviation (std), kurtosis, skewness, peak-to-peak V_{pp}, number of zero-crossings N_{zc}, number of peaks N_p, and other features such as nonlinear-energy-operators (envelope-derivative $NLEO_{ED}$ and Teager–Kaiser $NLEO_{TK}$) [40], signal energy (in the time domain E_t and frequency domain E_f) [41], and Shannon Entropy.

2.3.3. Spectral features (S)

We evaluate spectral features, specifically, relative power RP_f obtained from five EEG frequency bands: delta (δ, 1–4 Hz), theta (θ, 4–8 Hz), alpha (α, 8–13 Hz), beta (β, 13–30 Hz), and gamma (γ, >30 Hz) [7]. Relative power RP_f is defined as

$$RP_f = P_f/P_{\text{total}}, \quad (4)$$

where total power $P_{\text{total}} = P_\delta + P_\theta + P_\alpha + P_\beta + P_\gamma$ and f indicate different frequency bands ($f \in \{\delta, \theta, \alpha, \beta, \gamma\}$). This results in five feature values for each single-channel EEG segment.

2.3.4. Wavelet transform features

We evaluate the wavelet coefficients extracted using discrete wavelet transform (DWT) and continuous wavelet transform (CWT). We chose Daubechies (db4) [20] as the mother wavelet for DWT and extracted approximate and detailed coefficients for six levels. We select Morlet (morl) [22] as the mother wavelet for CWT and extract coefficients with scales above frequency of 2 Hz. We extract two features from the wavelet coefficients: mean (MSA) and standard deviation (SSA) of square of absolute values of CWT (CWT-F) and DWT (DWT-F) coefficients. These are defined as:

$$\text{MSA}_{\text{morl}} = \text{mean}(|\text{CWT}_{\text{matrix}}|), \quad (5)$$

$$\text{SSA}_{\text{morl}} = \text{std}(|\text{CWT}_{\text{matrix}}|), \quad (6)$$

$$\text{MSA}_{\text{db4}} = \text{mean}(|\text{DWT}_{\text{matrix}}|), \quad (7)$$

$$\text{SSA}_{\text{db4}} = \text{std}(|\text{DWT}_{\text{matrix}}|). \quad (8)$$

2.3.5. ST features

The ST is defined as a CWT with a specific mother wavelet $w(t,f)$ multiplied by a phase factor [42]:

$$S(\tau,f) = \exp(\text{i}2\pi f\tau)W(\tau,d), \quad (9)$$

where CWT of an input function $x(t)$ is defined as:

$$W(\tau,d) = \int_{-\infty}^{\infty} x(t)w(t-\tau,d)\,\text{dt}, \quad (10)$$

and the specific mother wavelet is defined as:

$$w(t,f) = \frac{f}{2\pi}\exp\left(-\frac{t^2f^2}{2}\right)\exp(-\text{i}2\pi ft)\,\text{dt}. \quad (11)$$

Here, the scale parameter d is the inverse of frequency f. We extract two features from the ST: Mean square root of standard deviations (ST-SR) [24], and skewness of sum of powers of S transform over epochs of $F_s/2$ (ST-P) [23]. These are defined as:

$$\text{ST}-\text{SR} = \text{mean}\left(\sqrt{\text{std}(\text{ST}_{\text{matrix}})}\right), \quad (12)$$

$$\text{ST}-\text{P} = \text{skewness}\left(\text{sum}\,|\text{ST}_{\text{matrix}}|\right). \quad (13)$$

2.3.6. Connectivity features

We evaluate the connectivity between the n channels of EEG. First, we compute the n^2 connectivity matrix between the channels and extract the lower triangular matrix features as features. In total, there are $(n^2 - n)/2$ features. We extract two connectivity measures from connectivity matrix, namely, maximum normalized cross-correlation (C-C) [43] and phase locking value (C-PLV) [44]. The maximum cross-correlation between two input signals x_n and y_n is computed as:

$$\hat{R}_{xy,\max}(m) = \frac{1}{\sqrt{\hat{R}_{xx}(0)\hat{R}_{yy}(0)}}\hat{R}_{xy}(m), \quad (14)$$

$$\hat{R}_{xy}(m) = \left\{ \begin{array}{ll} \sum_{n=0}^{N-m-1} x_{n+m} \times y_m^* & m \geqslant 0 \\ \hat{R}_{yx}^*(-m) & m < 0 \end{array} \right\}, \quad (15)$$

where $*$ indicates complex conjugation. The C-PLV of two signals x_n and y_n is given by:

$$\mathrm{C-PLV} = \frac{1}{N}\left|\sum_{n=1}^{N}\exp(\mathrm{i}[\psi_x(n) - \psi_y(n)])\right|, \quad (16)$$

where ψ_x and ψ_y are the phase values of the signals in radians computed by applying Hilbert transform.

2.3.7. Graph metric features

Derived from C-C and C-PLV feature sets, we compute 20 different graph metrics [17] using the MATLAB Brain Connectivity toolbox [18], namely C-C network (C-C-net) and C-PLV network (C-PLV-net) respectively. Specifically, we calculate several nodal features (degree, strength, assortativity, characteristic path length, local efficiency, eccentricity, betweenness centrality, eigenvector centrality, clustering coefficient, node coreness, participation coefficient, and diversity coefficient), edge features (assortativity coefficient, global efficiency, radius, diameter, transitivity, edge neighborhood overlap, and node pair degree), and an aggregate feature (matching index).

2.3.8. Patient specific features

We also investigate whether the age of the patient and vigilance state during the recording is relevant for classifying the EEGs. Both age and vigilance state of the patient are extracted directly from the clinical reports written by the neurologists. We divide the patient's age into four groups, and the vigilance state of the patients into three by encoding (see table 3).

Table 3. Encoding for age and vigilance state of the patient.

Feature	Encoding	Feature value
Age	00	$18 \leqslant$ age <30
	01	$30 \leqslant$ age <50
	10	$50 \leqslant$ age <70
	11	age >70
Vigilance state of patient	00	Awake
	10	Drowsy
	11	Intermittent sleep

2.4. EEG segment length

We evaluate the features for 8 different non-overlapping EEG segment lengths: 2 s, 5 s, 10 s, 20 s, 30 s, 60 s, 300 s, and entire EEG. For each l-second EEG segment length, we compute various aggregate statistics from the features calculated from numerous non-overlapping epoch segments of the entire EEG recording: mean, median, standard deviation (std), skewness, and kurtosis.

2.5. Training and evaluation

To set a benchmark, we followed the same CNN architecture and model selection pipeline as proposed in our recent study [7]. The IED detector is trained on the MGH dataset, and LOSO CV EEG classification (epileptic EEGs with IEDs vs. normal EEGs) is applied on each EEG dataset separately, excluding MGH dataset. We perform EEG classification across multiple centers in two stages. In stage one, we performed LOSO on each feature set separately and each EEG dataset individually (see figure 2). In this individual feature scenario, LOSO is performed as follows: For each dataset with the selected feature, we train the classifier on $N-1$ subjects and evaluate it on the Nth subject. We repeat this step N times to evaluate all the subjects. Then, we generate a receiver operating characteristics (ROC) curve from the test result and compute LOSO CV AUC and BAC. The same procedure is applied across different datasets, and we report the LOSO CV AUC on the individual datasets. We also average the LOSO results for the different datasets and report the mean AUC and BAC results. As we evaluate the EEG with four different montages and eight different segment lengths (7 fixed segment lengths and 1 that spans across the entire EEG), we use five different statistical measures to combine features (see table 4). As a result, for each feature set, there will be $4 \times (7 \times 5 + 1) = 144$ LOSO CV EEG classification results, excluding the classification result collected from the IED rate. Then, we rank the ten features based on the maximum LOSO CV AUC result obtained from 144 different combinations. This ranking information will be used in the next stage while combining multiple features to perform LOSO CV. In addition, we also have information about the best-performing combination for each of the 10 features.

In stage two, we evaluate LOSO (epileptic EEGs with IEDs vs. normal EEGs) by combining the IED rate with different combinations of IED-independent features (see figure 3). In this multi-feature scenario, LOSO is performed as follows: For each dataset with N subjects, we split the data as train, valid, and test set as follows: we leave the Nth subject for testing. From the remaining $N-1$ subjects, we use 50% of the data for training and 50% for validation. To find

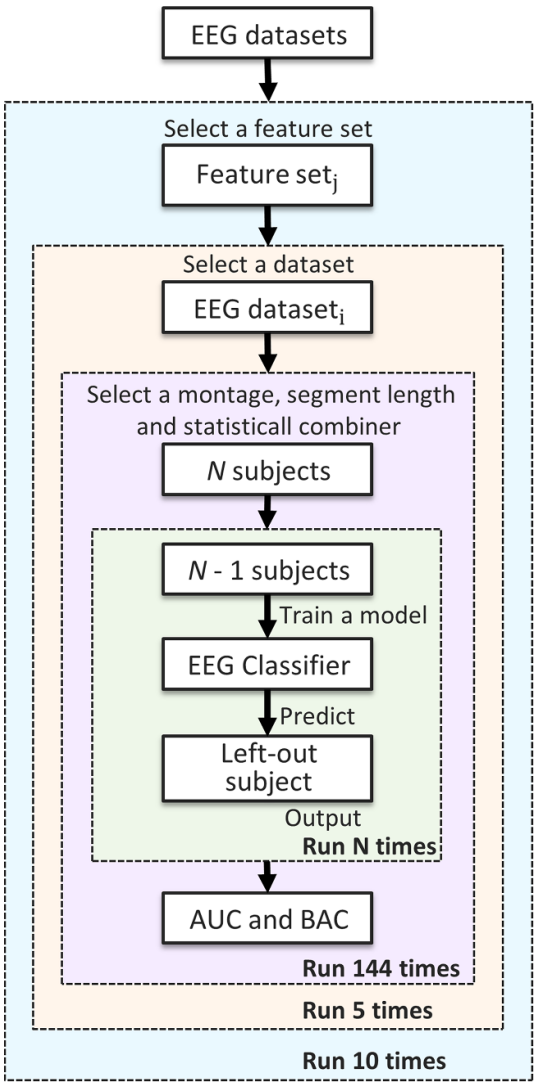

Figure 2. LOSO CV process on individual features. There are 10 different feature sets, and 5 different datasets. For each feature set and individual dataset, there are 144 combinations of montages, EEG segment lengths, and statistical combiners. Each LOSO CV is applied N times where N represents the number of subjects in EEG dataset$_i$.

Table 4. Different montages, segment lengths, statistical combiners and feature sets investigated in this study.

Montages	Segment lengths	Statistical combiners	Feature sets
			IED rate
			S
	2 s		C-C
	5 s	mean	C-PLV
CAR	10 s	median	ST-SR
Cz	20 s	std	ST-P
Bipolar	30 s	skewness	DWT-F
Laplacian	60 s	kurtosis	CWT-F
	300 s		C-C-net
	Entire EEG		C-PLV-net
			UTM

std: standard deviation.

the best combination of weights for multiple features, we define the weights $w_i \in [0, 1]$ for each feature set and evaluate the possible weights (with the condition $\sum_{i=1}^{n} w^i = 1$) in a grid search manner. Finally, the best combination of weights is applied on the test subject to predict the output:

$$\text{Output} = \sum_{i=1}^{n} w_i^{\text{opt}} \times o_i, \tag{17}$$

where w_i^{opt} is the optimal weight for output o_i from the classifier$_i$ and n represents the total number of feature sets.

We repeat this step N times to evaluate all the subjects. Then, we compute AUC and BAC from the ROC curve. The same procedure is applied across different datasets, and the mean LOSO results are reported. As there are 100+ possible combinations of montages, features, segment lengths, statistical combination measures, we propose a guided approach to evaluate the combination of feature sets. Based on the features ranked in stage one, we evaluate it as a combination of two, three, four feature sets, and so on. While combining two feature sets, we combine

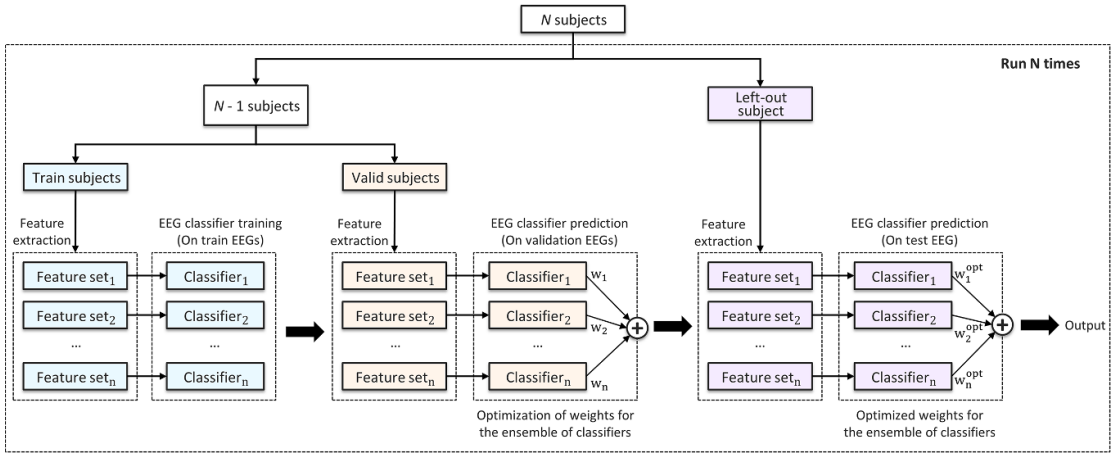

Figure 3. LOSO CV process on combination of features. For each dataset with N subjects, we leave out EEG(s) from the Nth subject for testing and use the remaining EEGs from $N-1$ subjects for training and validation. From the remaining $N-1$ subjects, we use half of them for training and the other half for selecting the best combination of weights w for multiple feature sets. For each feature set$_i$ extracted from EEGs of train subjects, an XGBoost classifier$_i$ is trained. w_i indicates weights applied to XGBoost classifier prediction output on feature set$_i$ of valid subjects in a grid search manner to find the optimal weight combination w_i^{opt}. With this set of optimal weights, predictions are made on the EEG(s) of test subject.

the IED rate with every IED-independent feature one by one and compute LOSO CV. For the three feature combinations, we combine the IED rate and top-ranked IED-independent feature with the remaining nine features one by one and compute LOSO CV. We repeat the same for the combination of four features up to ten features. Finally, we will combine the IED rate with all the ten features and perform LOSO CV. For each feature, we utilize the montage, EEG segment length, and statistical measure that best performed on that feature on the mean of five datasets in terms of AUC.

In both stages, the BAC is reported for 80% sensitivity in order to standardize the results. We applied the XGBoost classifier, as it often performs well and also is able to provide the relevance of each feature [27]. The sample_pos_weight XGBoost parameter controls the balance of positive and negative weights for handling class imbalance. The other XGBoost hyperparameters are set to default values. For the TUH dataset, as there are multiple EEG recordings from the same subject, we implement subject-level classification. To this end, we combine the features for a single subject by considering the maximum prediction value from multiple EEGs. We also consider the problem of detecting epileptic EEGs without IEDs for the TUH dataset.

3. Results

We evaluate the feature sets by applying LOSO CV on datasets containing epileptic EEGs with IEDs and normal EEGs from five centers. The inclusion of age, vigilance state, both age and vigilance state of the patient improved the overall mean BAC by 1.38%, 0.14%, and 1.46%, respectively, across the different combinations and sets of features. Therefore, in the following, we include the age and vigilance state of the patient in all feature sets.

The LOSO CV results for each individual set of features are summarized in figure 4 and table 5. The optimal choice of montages, segment lengths, and the aggregate statistics for each individual feature is also presented in table 5. The DWT-F achieved the best mean LOSO CV AUC of 0.825 and BAC of 74.1%, followed by CWT-F (AUC of 0.821, BAC of 73.6%), and UTM (AUC of 0.809, BAC of 74.4%). The ranking of the features based on AUC is as follows: DWT-F, CWT-F, UTM, ST-SR, S, C-PLV, C-C, C-C-net, C-PLV net, and ST-P features. However, these results are lower than the benchmark LOSO CV EEG classification results based on IED detection (AUC of 0.857, BAC of 79.5%) [8]. The choice of montage, EEG segment length, and statistical combiner varied across the type of individual feature. CAR and Laplacian montage yield the best AUC results, with CAR performing well for the top 5 ranked features. On the whole, shorter EEG segment lengths performed better than longer EEG segment lengths, except CWT. Also, while combining the results from EEG segments, the median measure was more reliable than other aggregate statistics, followed by the mean.

Next, we investigate the combination of the best performing features (see table 5). The mean results for LOSO CV for five datasets are summarized in table 6. Certain combinations of IED rate and IED-independent feature sets performed better than the baseline. The combination of IED rate, DWT-F, and CWT-F performed the best overall with an AUC of 0.871 and BAC of 80.9%. The LOSO CV results for five datasets for the best performing combination are also displayed in figure 4.

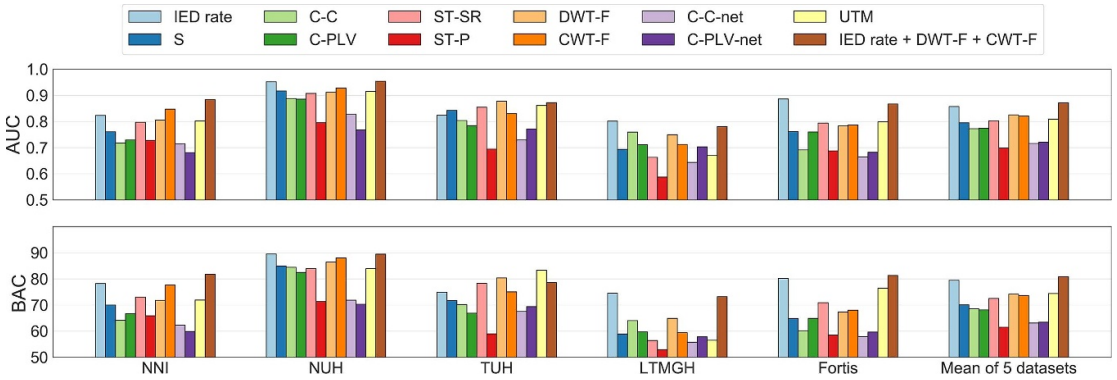

Figure 4. LOSO CV (AUC and BAC) for individual features and best combination of features for the different datasets.

Table 5. LOSO CV results for best performing combination in individual features.

Feature	Montage	Segment length	Combiner	AUC	BAC
IED rate	—	—	—	0.857 ± 0.06	79.5 ± 6.1
S	CAR	5	median	0.795 ± 0.06	70.1 ± 10.3
C-C	Laplacian	20	mean	0.772 ± 0.09	68.6 ± 9.8
C-PLV	Laplacian	5	mean	0.774 ± 0.10	68.1 ± 8.5
ST-SR	CAR	20	median	0.803 ± 0.11	72.5 ± 11.8
ST-P	Laplacian	20	median	0.698 ± 0.10	61.5 ± 9.5
DWT-F	CAR	30	median	0.825 ± 0.08	74.1 ± 10.2
CWT-F	CAR	300	median	0.821 ± 0.11	73.6 ± 12.3
C-C-net	Laplacian	10	std	0.716 ± 0.08	63.1 ± 7.9
C-PLV-net	Laplacian	2	mean	0.721 ± 0.06	63.4 ± 4.7
UTM	CAR	20	median	0.809 ± 0.11	74.4 ± 12.7

The results are reported as mean ± standard deviation for the different datasets.
BAC is reported for 80% sensitivity.

Table 6. LOSO CV results for combination of features.

Feature combination	LOSO CV	
	AUC	BAC
IED rate (*Baseline*)	0.857 ± 0.06	79.5 ± 6.1
IED rate + S	0.844 ± 0.11	79.6 ± 9.6
IED rate + C-PLV	0.773 ± 0.16	71.8 ± 14.9
IED rate + ST-SR	0.825 ± 0.10	76.6 ± 9.5
IED rate + ST-P	0.859 ± 0.06	78.9 ± 6.6
IED rate + DWT-F	0.834 ± 0.08	77.4 ± 7.9
IED rate + CWT-F	0.830 ± 0.10	78.1 ± 8.2
IED rate + C-CC-net	0.817 ± 0.11	73.6 ± 14.5
IED rate + C-PLV-net	0.831 ± 0.09	76.5 ± 10.3
IED rate + UTM	0.836 ± 0.09	76.3 ± 10.5
IED rate + DWT-F + CWT-F	0.871 ± 0.06	80.9 ± 5.9
IED rate + DWT-F + UTM	0.860 ± 0.07	79.5 ± 7.3
IED rate + DWT-F + S	0.867 ± 0.07	79.9 ± 7.3
IED rate + DWT-F + C-PLV-net	0.865 ± 0.06	80.2 ± 6.5
IED rate + DWT-F + CWT-F + UTM	0.879 ± 0.06	80.3 ± 5.9
IED rate + DWT-F + CWT-F + S	0.873 ± 0.06	80.7 ± 5.6
IED rate + DWT-F + CWT-F + C-C	0.864 ± 0.06	80.4 ± 5.5
IED rate + DWT-F + CWT-F + C-C-net	0.870 ± 0.05	80.3 ± 6.1
IED rate + DWT-F + CWT-F + C-PLV-net	0.876 ± 0.06	80.1 ± 5.8
IED rate + DWT-F + CWT-F + ST-P	0.862 ± 0.06	79.7 ± 5.8

The results are reported as mean ± standard deviation for the different datasets.
BAC is reported for 80% sensitivity.

4. Discussion

4.1. Classification of epileptic EEGs with IEDs vs. non-epileptic EEGs

We have developed automated methods for classifying epileptic EEGs with IEDs vs. normal EEGs and have tested it on datasets from five institutions. The system was evaluated in two stages of LOSO CV: individual feature sets and combination of feature sets (with and without IEDs). The benchmark LOSO CV EEG classification system based on IED detection has an AUC of 0.857 (BAC of 79.5%). The proposed system achieved a best mean LOSO CV AUC of 0.871 (BAC of 80.9%) by combining the IED rate with IED-independent features.

In general, the mean LOSO CV results across different EEG features improves by including the age and vigilance state of the patient. The improvement in classification results indicates that patient information such as age and vigilance state affect the pathogenesis of epilepsy (epileptogenic potential) as shown by earlier studies [25, 32, 33].

Among the IED-independent features, UTM features yields the best BAC result (AUC of 0.809, BAC of 74.4%), followed by features based on DWT-F (AUC of 0.825, BAC of 74.1%), CWT-F (AUC of 0.821, BAC of 73.6%), ST-SR, S, C-PLV, C-C, C-C-net, C-PLV-net, and ST-P. This overall trend is reflected in the LOSO CV AUC and BAC results consistently for all individual datasets. The BAC for the IED-independent features has a higher standard deviation across datasets compared to the baseline; the IED-independent features might be more dataset-specific, as they are potentially more sensitive to differences in the EEG machines. UTM performs the best among all considered EEG measures, probably because epileptic EEGs exhibit clear differences in time-domain due to epileptiform activity (IEDs and potentially beyond). Among the 13 individual features in UTM, peak-to-peak V_{pp} is ranked as the most important feature, followed by the number of zero-crossings N_{zc}. Spectral measures also lead to good classification results, which might be explained by the fact that IEDs perturb the spectrum; moreover, IEDs are sometimes followed by slow waves, which also leave their imprint on the EEG spectrum. As IEDs have a duration of less than one second, shorter segment lengths led to better results for spectral features. Slowing is another characteristic trait of EEG abnormality that is captured in spectrum [11]. This also explains the importance of including time-frequency-based wavelet features (CWT and DWT) for classifying epileptic vs. normal EEGs. As ST is a modified version of the Fourier Transform and CWT, it offers superior time resolution and strong frequency localization and resolution, and leads to good classification results. The connectivity and network-based features yielded the worst results as they are mainly relevant for high-density EEG with 128 or more electrodes recorded at a high sampling frequency [23]. In addition, connectivity for different types of epilepsy may have different feature patterns, thus rendering it unsuitable for general purpose epilepsy diagnosis. Network based diagnosis of specific types of epilepsy is a promising area of future research.

Though the BAC for IED-independent features is still lower than the benchmark BAC of 79.5% based on the IED rate, we achieved a superior LOSO CV AUC of 0.871 (BAC of 80.9%) by combining the IED rate with the DWT-F and CWT-F features. Concretely, we have gained a BAC improvement of 1.4% (specificity improvement of 2.8%) compared to the benchmark, corresponding to an additional 30 EEGs that are correctly classified. We have also performed Leave-One-Institution-Out (LOIO) CV in which we train the EEG classifier on subjects from $N-1$ institutions and test it on Nth institution [8]. We have achieved a mean LOIO CV AUC of 0.84 (BAC of 78.1%) with IED rate, DWT-F, and CWT-F, and the results are coincidentally identical to our previous study [8]. In contrast to LOSO CV, we did not achieve improvements for LOIO CV by adding IED-independent features to the IED rate for classification, since these IED-independent features may be dataset specific [7].

The LOSO CV results for the publicly available TUH Epilepsy Corpus [34] are reported in tables 7 and 8. These performance metrics may in the future serve as a benchmark for the classification of EEG signals for epilepsy diagnosis. The LTMGH data is recorded under non-standard conditions and contains excessive delta power; therefore, the wavelet and other IED independent features lead to weaker results for that dataset. When we exclude LTMGH from LOSO CV analysis, AUC (BAC) for the UTM features, the IED rate, and the IED rate + DWT-F + CWT-F is 0.844 (78.9%), 0.871 (80.7%), and 0.894 (82.8%), respectively.

In figure 5, we compared the average CWT (morl) coefficients of EEGs of all the datasets. Epileptic EEGs have high power in lower frequencies (from 1.66 Hz to 6.4 Hz which roughly translates to delta and theta frequency bands) than normal EEGs. For frequencies from 8.97 Hz to 24.72 Hz (which roughly translates to alpha and beta band), epileptic EEGs have lower power than normal EEGs. The overall power in LTMGH EEGs is noticeably higher than other datasets EEGs. We have also compared the relative power of EEGs of all the datasets in figure 6. The p-value (based on a Mann–Whitney U test applied on the delta, theta, alpha, and beta relative power) suggests that the relative power features discriminate well ($p \leqslant 0.05$) on all the five datasets, except for NUH theta and Fortis alpha relative power. From the box plot, we can observe that epileptic EEGs have stronger delta and theta relative, but lower alpha and beta power, similar to the findings from in figure 5. Higher delta and theta relative power, with lower alpha and beta

Table 7. LOSO CV results for TUH dataset.

		IED rate		IED rate + DWT-F + CWT-F		IED rate + DWT-F + ST-P	
	Epileptic vs. normal	AUC	BAC	AUC	BAC	AUC	BAC
Subject-level classification	Epileptic with IED vs. normal	0.824	74.9	0.872	78.7	0.863	80.2
	Epileptic without IED vs. normal	0.538	51.8	0.630	57.3	0.715	65.7
	Epileptic with and without IED vs. normal	0.694	58.8	0.790	71.7	0.767	67.7
EEG-level classification	Epileptic with IED vs. normal	0.839	75.5	0.935	85.5	0.917	82.1
	Epileptic without IED vs. normal	0.608	57.2	0.662	58.9	0.787	73.0
	Epileptic with and without IED vs. normal	0.688	59.6	0.822	77.6	0.790	75.2

BAC is reported for 80% sensitivity.

There are 421 epileptic EEGs (260 EEGs with IEDs and the remaining 161 EEGs without IEDs).

Table 8. Detailed analysis of LOSO CV results for TUH dataset (epileptic with and without IED vs. normal).

Classification type	Feature combination	TN	FP	TP_1	FN_1	TP_2	FN_2	BAC_1	BAC_2
Subject-level classification	IED rate	11	19	40	2	15	12	66.0	46.1
	IED rate + DWT-F + CWT-F	19	11	37	5	18	9	75.7	65.0
	IED rate + DWT-F + ST-P	17	13	37	5	18	9	72.4	61.7
EEG-level classification	IED rate	17	27	234	26	103	58	64.3	51.3
	IED rate + DWT-F + CWT-F	33	11	224	36	113	48	80.6	72.6
	IED rate + DWT-F + ST-P	31	13	236	24	101	60	80.6	66.6

BAC is reported for 80% sensitivity.

TN: True negatives, FP: False positives, TP: True positives, FN: False negatives.

TP_1 and FN_1 are reported for epileptic with IED. TP_2 and FN_2 are reported for epileptic without IED.

BAC_1 is reported for epileptic with IED vs. normal. BAC_2 is reported for epileptic without IED vs. normal.

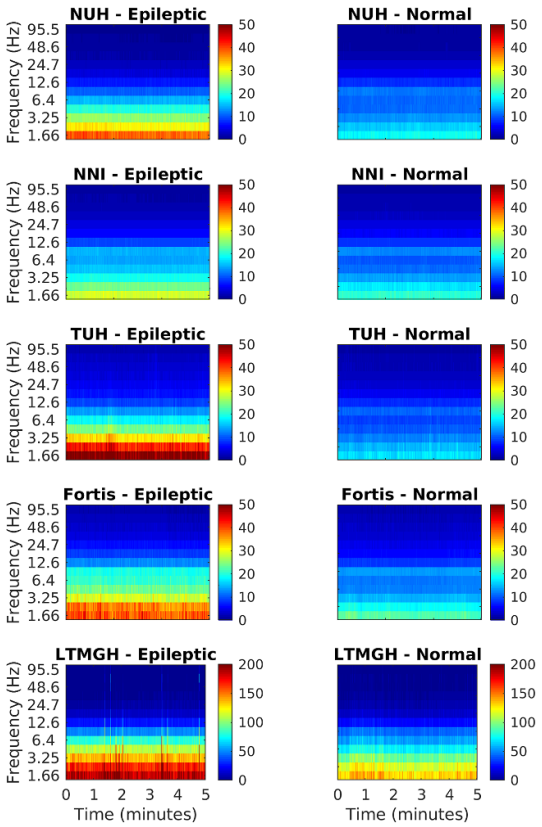

Figure 5. Average CWT coefficients across all epileptic EEGs vs. normal EEGs for each dataset. The color bar indicates the squared magnitude (power) of the CWT coefficients.

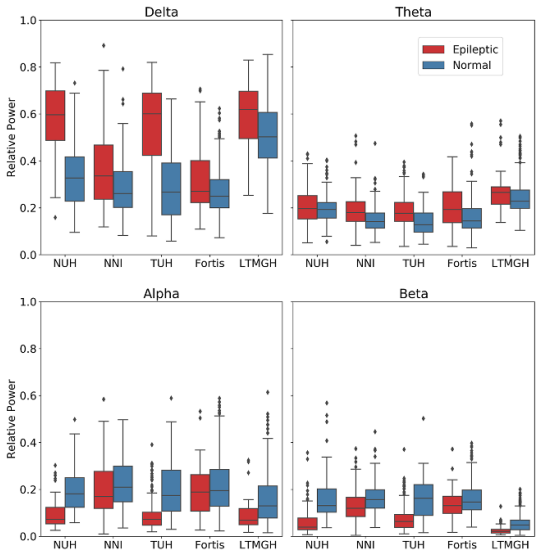

Figure 6. Relative power in the delta, theta, alpha, and beta band for five different datasets.

relative power indicates slowing in the EEG [11], therefore, slowing occurs in the epileptic EEGs from NUH, NNI, TUH, Fortis, and LTMGH. Thus, slowing could be one of the effects in the EEG that have inherently helped in our classification based on IED-independent features.

In our previous study [8], we compared our results with similar studies in the literature. Most of the epileptic EEG classification studies in the literature have one or more of the following drawbacks: evaluated on smaller datasets, labeled by experts from the same institution or similar training [10, 45], evaluated only on children with epileptic spasms [16, 46], handpicked artifact-free EEG segments [15], recorded intracranially [15, 47], and performed on seizures EEGs that are easier to identify than EEG between seizures or with IEDs. In this study, we have addressed all of these shortcomings and achieved superior LOSO CV results. The inter-rater agreement (IRA) reported for classifying epileptic EEGs is around 70%–80% [7]). The proposed approach performs within the IRA range (albeit for other datasets), and hence may reach the level of human experts. In the future, we will conduct IRA studies on the datasets considered in the current study, as validation of the proposed methods.

4.2. Classification of epileptic EEGs with and without IEDs vs. non-epileptic EEGs

In contrast to the other datasets, the TUH dataset not only contains epileptic EEGs with IEDs, but also IED-free epileptic EEGs. We investigated whether the latter can also be distinguished from normal EEGs. In figure 7, we display the normalized IED rate for normal EEGs, epileptic EEGs with and without IEDs, in addition to the output of the classifier with IED

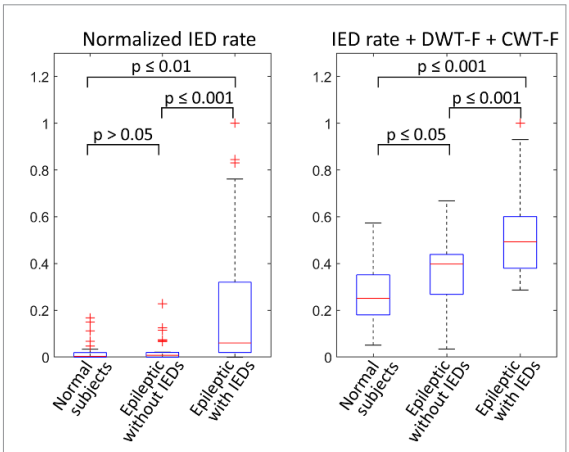

Figure 7. Normalized IED rates and output for the classifier with IED rate, db4, and morl as features, for normal EEGs and epileptic EEGs with and without IEDs.

rate + DWT-F + CWT-F as features. The p-value (based on a two-sample t-test applied to the normalized IED rates and classifier outputs) suggests that output based on the feature combination (IED rate combined with other IED-independent features) can better discriminate the three cases compared to the IED rate only. We summarize the LOSO results for the TUH dataset in table 7 and report the detailed analysis of the confusion matrix for classifying epileptic with and without IED vs. normal in table 8. For classifying epileptic EEGs without IEDs vs. normal EEGs (on the level of individual EEGs), the classifier based only on the IED rate achieved an AUC of 0.608 (BAC of 57.2%), whereas the feature combination of IED rate + DWT-F + ST-P achieved an AUC of 0.787 (BAC of 73%). For EEG-level classification of epileptic EEGs with and without IEDs vs. normal EEGs, the feature combination of IED rate + DWT-F

+ CWT-F achieved an AUC of 0.822 (BAC of 77.6%), while based only on the IED rate attained an AUC of 0.688 (BAC of 59.6%). More specifically, the former classifier attains a BAC of 80.6% for classifying epileptic EEGs with IEDs vs. normal EEGs, and a BAC of 72.6% for IED-free epileptic EEGs vs. normal EEGs (see table 8). For combining these three feature sets for classifying epileptic EEGs without IEDs vs. normal EEGs, the mean optimal weight configurations for $w^{opt}_{IEDrate}$, w^{opt}_{DWT-F}, and w^{opt}_{CWT-F} are found to be 0.55, 0.26, and 0.19, respectively. From the optimal weights, it can be observed that the IED rate contributed the highest, whereas DWT-F and CWT-F had noticeable contribution in classifying EEGs. Specifically, the addition of IED-independent features improved the BAC by 21% (BAC of 51.3% from IED rate vs. 72.6% from combination of IED rate with IED-independent features) in detecting epileptic EEGs that do not contain IEDs. These promising results suggest that it might indeed be possible to distinguish algorithmically IED-free epileptic EEGs from normal EEGs by combinations of appropriate feature sets.

Verhoeven *et al* performed automated diagnosis of temporal lobe epilepsy in the absence of IEDs. The study was conducted on a single dataset of 40 epileptic and 35 controls and achieved an EEG classification accuracy of 90.7%. The study was performed on sixty epochs of 1s segments, free of IEDs and artifacts handpicked from each epileptic subject [15], thus rendering the results unreliable as these IED-free segments are extracted from epileptic EEGs with IEDs, and the evaluation was not conducted on the entire EEG. In summary, we have developed a new system that performs as good as (or even slightly better than) existing IED detectors in detecting epileptic EEGs with IEDs, but also performs much better on epileptic EEGs without IEDs compared to the state-of-the-art systems. As far as we know, this study might be the first to showcase the classification of entire EEGs from epileptic patients without IEDs vs. normal EEGs with reasonable accuracy, where traditional IED detectors perform poorly.

The study comes with few limitations. Firstly, it would be interesting to explore deep learning classifiers as an alternative for XGBoost. This topic goes beyond the scope of this study, and we leave it for future research. Secondly, though cross-institutional assessment is carried out, each EEG has only been reviewed by one neurologist (different for each institution); the results would be more reliable if the EEGs were independently reviewed by multiple neurologists. Thirdly, investigating statistical differences in individual features within each feature set and its contribution in epilepsy diagnosis is important that it merits its own separate study. Though the IED detector is robust against artifacts, as observed by Peh *et al* [48, 49], a separate artifact rejection module may help to improve the overall performance of the system. We intend to explore this avenue in future studies.

5. Conclusion

We have developed an automated, efficient, and generalized epileptic EEG classification system that relies on both IED-dependent and IED-independent EEG measures. The proposed system was cross-validated on data from five different institutions and attained a mean LOSO CV AUC of 0.871 (BAC of 80.9%) and AUC of 0.809 (BAC of 74.4%) with and without IED features, respectively for classification of epileptic EEGs with IEDs vs. normal EEGs. The system was also evaluated on detection of epileptic EEGs with and without IEDs and achieved promising results, but more research into IED-independent features is required. Specifically, the addition of IED-independent features improved the BAC by 21% in detecting epileptic EEGs that do not contain IEDs. In summary, we have confirmed that adding these non-IED based features does not diminish the classification performance, but instead slightly improves the accuracy of classifying epileptic EEGs with IEDs vs. normal EEGs. More importantly, we have also shown that these non-IED based features when combined with CNN-based IED detector effectively aids in detecting epileptic EEgs without IEDs from normal EEGs. A critical advantage of the IED-independent features is that there is no requirement to annotate or observe IEDs in EEG to make a diagnosis. Alternatively, these IED-independent features could be applied along with IED-based EEG metrics to improve reliability, especially for detecting epileptic EEGs without any IEDs. The proposed EEG classification system may thus be a valuable tool for neurologists to review epileptic EEGs efficiently.

Data availability statement

The data generated and/or analysed during the current study are not publicly available for legal/ethical reasons but are available from the corresponding author on reasonable request.

Acknowledgments

The data collection at MGH was carried out under the guidance of Dr Cash, Dr Jing, and Dr Westover. The NNI and NUH dataset collection was performed with the support of the National Health Innovation Centre (NHIC) Grant (NHIC-I2D-1608138), under the guidance of Dr Rahul Rathakrishnan and Dr Yee-Leng Tan, respectively. This research also had support from the Ministry of Education (MoE), Singapore (Academic Research Funding TIER 1-2019-T1-001-116 RG16/19).

Appendix. Glossary of Terms, Abbreviations, and Acronyms

Table A1. Glossary of Terms, Abbreviations, and Acronyms.

Name	Description
AUC	Area under the curve
BAC	Balanced accuracy
CAR	Common average referential
CNN	Convolutional neural network
CV	Cross-validation
CWT	Continuous wavelet transform
C-C	Connectivity—Cross correlation
C-PLV	Connectivity—Phase lock value
db4	Daubechies wavelet
DWT	Discrete wavelet transform
EEG	Electroencephalogram
Epileptic EEG	EEG from patient diagnosed with epilepsy as per the clinical report
Epileptic EEG with IEDs	Patient diagnosed with epilepsy whose EEGs exhibit IED patterns
Epileptic EEG without IEDs	Patient diagnosed with epilepsy whose EEG do not exhibit IED patterns
FN	False negatives
FP	False positives
IED	Interictal epileptiform discharge
IED-dependent features	IED rates (Number of IEDs detected per minute by CNN IED detector)
IED-independent features	Computed from the EEG, without expert annotated IED segments
LOIO	Leave-One-Institution-Out
LOSO	Leave-One-Subject-Out
LTMGH	Lokmanya Tilak Municipal General Hospital
MGH	Massachusetts General Hospital
morl	Morlet wavelet
NUH	National University Hospital
NNI	National Neuroscience Institute
Non-epileptic EEG	Normal EEG with no marked abnormality in the clinical report
ROC	Receiver operating characteristics
ST	Stockwell transform
TP	True positives
TUH	Temple University Hospital
UTM	Univariate temporal measures
XGBoost	eXtreme Gradient Boosting

ORCID iDs

Prasanth Thangavel https://orcid.org/0000-0001-9023-297X
John Thomas https://orcid.org/0000-0003-0144-3746
Nishant Sinha https://orcid.org/0000-0002-2090-4889
Rajamanickam Yuvaraj https://orcid.org/0000-0003-4526-0749
Rahul Rathakrishnan https://orcid.org/0000-0003-2078-1759
Rohit Srivastava https://orcid.org/0000-0002-3937-5139

References

[1] Thijs R D, Surges R, O'Brien T J and Sander J W 2019 Epilepsy in adults *Lancet* **393** 689–701
[2] Frauscher B and Gotman J 2019 Sleep, oscillations, interictal discharges and seizures in human focal epilepsy *Neurobiol. Dis.* **127** 545–53
[3] Wei B, Zhao X, Shi L, Xu L, Liu T and Zhang J 2021 A deep learning framework with multi-perspective fusion for interictal epileptiform discharges detection in scalp electroencephalogram *J. Neural Eng.* **18** 0460b3
[4] Geng D, Alkhachroum A, Melo Bicchi M A, Jagid J R, Cajigas I and Chen Z S 2021 Deep learning for robust detection of interictal epileptiform discharges *J. Neural Eng.* **18** 056015
[5] Pillai J and Sperling M R 2006 Interictal EEG and the diagnosis of epilepsy *Epilepsia* **47** 14–22
[6] Khosropanah P, Ramli A R, Abbasi M R, Marhaban M H and Ahmedov A 2020 A hybrid unsupervised approach toward EEG epileptic spikes detection *Neural Comput. Appl.* **32** 2521–32
[7] Thomas J *et al* 2021 Automated adult epilepsy diagnostic tool based on interictal scalp electroencephalogram characteristics: a six-center study *Int. J. Neural Syst.* **31** 2050074
[8] Thangavel P *et al* 2021 Time–frequency decomposition of scalp electroencephalograms improves deep learning based epilepsy diagnosis *Int. J. Neural Syst.* **31** 2150032

[9] Thomas J, Jin J, Thangavel P, Bagheri E, Yuvaraj R, Dauwels J, Rathakrishnan R, Halford J J, Cash S S and Westover B 2020 Automated detection of interictal epileptiform discharges from scalp electroencephalograms by convolutional neural networks *Int. J. Neural Syst.* **30** 2050030

[10] Jing J *et al* 2019 Development of expert-level automated detection of epileptiform discharges during electroencephalogram interpretation *JAMA Neurol.* **77** 103–8

[11] Peh W Y *et al* 2020 Five-institution study of automated classification of pathological slowing from adult scalp electroencephalograms (arXiv:2009.13554)

[12] Larsson Pål G and Kostov H 2005 Lower frequency variability in the alpha activity in EEG among patients with epilepsy *Clin. Neurophysiol.* **116** 2701–6

[13] Schmidt H, Woldman W, Goodfellow M, Chowdhury F A, Koutroumanidis M, Jewell S, Richardson M P and Terry J R 2016 A computational biomarker of idiopathic generalized epilepsy from resting state EEG *Epilepsia* **57** e200–4

[14] Chowdhury F A, Woldman W, FitzGerald T H B, Elwes R D C, Nashef L, Terry J R and Richardson M P 2014 Revealing a brain network endophenotype in families with idiopathic generalised epilepsy *PLoS One* **9** e110136

[15] Verhoeven T *et al* 2018 Automated diagnosis of temporal lobe epilepsy in the absence of interictal spikes *NeuroImage Clin.* **17** 10–15

[16] Smith R J, Hu D K, Shrey D W, Rajaraman R, Hussain S A and Lopour B A 2021 Computational characteristics of interictal EEG as objective markers of epileptic spasms *Epilepsy Res.* **176** 106704

[17] Fagerholm E D, Hellyer P J, Scott G, Leech R and Sharp D J 2015 Disconnection of network hubs and cognitive impairment after traumatic brain injury *Brain* **138** 1696–709

[18] Rubinov M and Sporns O 2010 Complex network measures of brain connectivity: uses and interpretations *Neuroimage* **52** 1059–69

[19] West C, Woldman W, Oak K, McLean B and Shankar R 2021 A review of network and computer analysis of epileptiform discharge free EEG to characterize and detect epilepsy *Clin. EEG Neurosci.* **53** 74–78

[20] Indiradevi K P, Elias E, Sathidevi P S, Dinesh Nayak S and Radhakrishnan K 2008 A multi-level wavelet approach for automatic detection of epileptic spikes in the electroencephalogram *Comput. Biol. Med.* **38** 805–16

[21] Gajic D, Djurovic Z, Gligorijevic J, Gennaro S Di and Savic-Gajic I 2015 Detection of epileptiform activity in EEG signals based on time–frequency and non-linear analysis *Front. Comput. Neurosci.* **9** 38

[22] Türk Omer and Özerdem M Sç 2019 Epilepsy detection by using scalogram based convolutional neural network from EEG signals *Brain Sci.* **9** 115

[23] Mooij A H, Frauscher B, Gotman J and Huiskamp G J M 2020 A skew-based method for identifying intracranial EEG channels with epileptic activity without detecting spikes, ripples, or fast ripples *Clin. Neurophysiol.* **131** 183–92

[24] Hariharan M, Vijean V, Sindhu R, Divakar P, Saidatul A and Yaacob S 2014 Classification of mental tasks using Stockwell transform *Comput. Electr. Eng.* **40** 1741–9

[25] Beghi E and Giussani G 2018 Aging and the epidemiology of epilepsy *Neuroepidemiology* **51** 216–23

[26] Chen T and Guestrin C 2016 XGBoost: a scalable tree boosting system *Proc. 22nd ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining* pp 785–94

[27] Torlay L, Perrone-Bertolotti M, Thomas E and Baciu M 2017 Machine learning–XGBoost analysis of language networks to classify patients with epilepsy *Brain Inform.* **4** 159–69

[28] Mitchell R and Frank E 2017 Accelerating the XGBoost algorithm using GPU computing *PeerJ Comput. Sci.* **3** e127

[29] Chen Z, Jiang F, Cheng Y, Gu X, Liu W and Peng J 2018 XGBoost classifier for DDoS attack detection and analysis in SDN-based cloud *2018 IEEE Int. Conf. on Big Data and Smart Computing (BIGCOMP)* (IEEE) pp 251–6

[30] Friedman J H 2017 *The Elements of Statistical Learning: Data Mining, Inference and Prediction* (Springer Open) (https://doi.org/https://doi.org/10.1007/978-0-387-84858-7)

[31] Murphy K P 2012 *Machine Learning: A Probabilistic Perspective* (Cambridge, MA: MIT Press)

[32] Holmes G L 2012 Consequences of epilepsy through the ages: when is the die cast? Introduction *Epilepsy Curr.* **12** 4–6

[33] Klimes P, Cimbalnik J, Brazdil M, Hall J, Dubeau Fçois, Gotman J and Frauscher B 2019 NREM sleep is the state of vigilance that best identifies the epileptogenic zone in the interictal electroencephalogram *Epilepsia* **60** 2404–15

[34] Obeid I and Picone J 2016 The temple university hospital EEG data corpus *Front. Neurosci.* **10** 196

[35] Veloso L, McHugh J, von Weltin E, Lopez S, Obeid I and Picone J 2017 Big data resources for EEGS: enabling deep learning research *2017 IEEE Signal Processing in Medicine and Biology Symp. (SPMB)* (IEEE) pp 1–3

[36] Britton J W, Frey L C, Hopp J L, Korb P, Koubeissi M Z, Lievens W E, Pestana-Knight E M and Louis St E K 2016 *Electroencephalography (EEG): An Introductory Text and Atlas of Normal and Abnormal Findings in Adults, Children and Infants* (Chicago, IL: American Epilepsy Society)

[37] McFarland D J, McCane L M, David S V and Wolpaw J R 1997 Spatial filter selection for EEG-based communication *Electroencephalogr. Clin. Neurophysiol.* **103** 386–94

[38] Syam S H F, Lakany H, Ahmad R B and Conway B A 2017 Comparing common average referencing to Laplacian referencing in detecting imagination and intention of movement for brain computer interface *MATEC Web of Conf.* vol 140

[39] Saab K, Dunnmon J, Ré C, Rubin D and Lee-Messer C 2020 Weak supervision as an efficient approach for automated seizure detection in electroencephalography *npj Digit. Med.* **3** 1–12

[40] O'Toole J M, Temko A and Stevenson N 2014 Assessing instantaneous energy in the EEG: a non-negative, frequency-weighted energy operator *2014 36th Annual Int. Conf. IEEE Engineering in Medicine and Biology Society* (IEEE) pp 3288–91

[41] Harati A, Golmohammadi M, Lopez S, Obeid I and Picone J 2015 Improved EEG event classification using differential energy *2015 IEEE Signal Processing in Medicine and Biology Symp. (SPMB)* (IEEE) pp 1–4

[42] Stockwell R G, Mansinha L and Lowe R P 1996 Localization of the complex spectrum: the S transform *IEEE Trans. Signal Process.* **44** 998–1001

[43] Stoica P and Moses R L 2005 *Spectral Analysis of Signals* (Upper Saddle River, NJ: Pearson Prentice Hall)

[44] Lachaux J-P, Rodriguez E, Martinerie J and Varela F J 1999 Measuring phase synchrony in brain signals *Hum. Brain Mapp.* **8** 194–208

[45] Roy S, Kiral-Kornek I and Harrer S 2019 ChronoNet: a deep recurrent neural network for abnormal EEG identification *Conf. on Artificial Intelligence in Medicine in Europe* (Springer) pp 47–56

[46] Lin L-C, Ouyang C-S, Wu R-C, Yang R-C and Chiang C-T 2020 Alternative diagnosis of epilepsy in children without epileptiform discharges using deep convolutional neural networks *Int. J. Neural Syst.* **30** 1850060

[47] Antoniades A, Spyrou L, Martin-Lopez D, Valentin A, Alarcon G, Sanei S and Took C C 2017 Detection of interictal discharges with convolutional neural networks using discrete ordered multichannel intracranial EEG *IEEE Trans. Neural Syst. Rehabil. Eng.* **25** 2285–94

[48] Peh W Y, Yao Y and Dauwels J 2022 Transformer convolutional neural networks for automated artifact detection in scalp EEG *2022 44th Annual Int. Conf. IEEE Engineering in Medicine & Biology Society (EMBC)* (IEEE) pp 3599–602

[49] Peh W Y, Yao Y and Dauwels J 2022 Transformer convolutional neural networks for automated artifact detection in scalp EEG (arXiv:2208.02405)