Time-Frequency Decomposition of Scalp Electroencephalograms Improves Deep Learning-Based Epilepsy Diagnosis

Thangavel, Prasanth; Thomas, John; Peh, Wei Yan; Jing, Jin; Yuvaraj, Rajamanickam; Cash, Sydney S.; Chaudhari, Rima; Saini, Vinay; Dauwels, Justin; More Authors

DOI
10.1142/S0129065721500325

Publication date
2021

Document Version
Final published version

Published in
International Journal of Neural Systems

International Journal of Neural Systems, Vol. 31, No. 8 (2021) 2150032 (16 pages)

DOI: 10.1142/S0129065721500325

Time–Frequency Decomposition of Scalp Electroencephalograms Improves Deep Learning-Based Epilepsy Diagnosis

Prasanth Thangavel*, John Thomas*, Wei Yan Peh*,
Jin Jing†, Rajamanickam Yuvaraj*,‡,
Sydney S. Cash†, Rima Chaudhari§, Sagar Karia¶,
Rahul Rathakrishnan‖, Vinay Saini**, Nilesh Shah¶,
Rohit Srivastava**, Yee-Leng Tan††,
Brandon Westover† and Justin Dauwels*,‡‡,§§

Nanyang Technological University, Singapore

†Massachusetts General Hospital and Harvard Medical School, USA

‡National Institute of Education, Singapore

§Fortis Hospital Mulund, Mumbai, India

¶Lokmanya Tilak Municipal General Hospital, India

‖National University Hospital, Singapore

**Department of Biosciences and Bioengineering
IIT Bombay, India

††National Neuroscience Institute, Singapore

‡‡Delft University of Technology, Netherlands

§§j.h.g.dauwels@tudelft.nl

Received 7 May 2021
Accepted 8 May 2021
Published Online 16 July 2021

Epilepsy diagnosis based on Interictal Epileptiform Discharges (IEDs) in scalp electroencephalograms (EEGs) is laborious and often subjective. Therefore, it is necessary to build an effective IED detector and an automatic method to classify IED-free versus IED EEGs. In this study, we evaluate features that may provide reliable IED detection and EEG classification. Specifically, we investigate the IED detector based on convolutional neural network (ConvNet) with different input features (temporal, spectral, and wavelet features). We explore different ConvNet architectures and types, including 1D (one-dimensional) ConvNet, 2D (two-dimensional) ConvNet, and noise injection at various layers. We evaluate the EEG classification performance on five independent datasets. The 1D ConvNet with preprocessed full-frequency EEG signal and frequency bands (delta, theta, alpha, beta) with Gaussian additive noise at the output layer achieved the best IED detection results with a false detection rate of 0.23/min at 90% sensitivity. The EEG classification system obtained a mean EEG classification Leave-One-Institution-Out (LOIO) cross-validation (CV) balanced accuracy (BAC) of 78.1% (area under the curve (AUC) of 0.839) and Leave-One-Subject-Out (LOSO) CV BAC of 79.5% (AUC of 0.856). Since the proposed classification system only takes a few seconds to analyze a 30-min routine EEG, it may help in reducing the human effort required for epilepsy diagnosis.

Keywords: Deep learning; convolutional neural networks; EEG classification; interictal epileptiform discharges; multiple features; noise injection.

§§Corresponding author.

1. Introduction

Epilepsy is a neurological disorder associated with recurrent unprovoked seizures. It is the fourth frequent neurological condition, affecting 70 million people worldwide.[1] Epilepsy is mainly diagnosed based on the presence of Interictal Epileptiform Discharges (IEDs) in electroencephalogram (EEG).[2,3] Although there are other features linked to epilepsy[3] such as seizures (ictal events), high-frequency oscillations (HFOs), periodic lateralized epileptiform discharges (PLEDs), IEDs are still considered as the main markers of epilepsy in routine clinical setting. Visual interpretation of routine and prolonged scalp EEG recordings by trained clinicians is laborious and error prone.[4] Thus, the development of efficient techniques for automatic IED detection is highly beneficial for effective diagnosis, treatment planning and drug management for epilepsy.[2,5] In the literature, a wide variety of techniques has been evaluated for IED detection, namely template matching, mimetic analysis, parametric modeling, spectral methods, and, since recently, also deep learning methods.[6] However, only a few studies have evaluated EEG classification based on IED detection.[7–10]

Recently, deep convolutional neural networks (ConvNets) have attained superior performance for IED detection.[6–11] Features learned by deep learning techniques have often proven to be more robust than hand-engineered features. Roy *et al.*[12] have reported that 40% of the reviewed studies applied ConvNets. In the literature, studies have been performed based on various types of data inputs, including one-dimensional (1D) preprocessed temporal EEG, Fast Fourier Transform (FFT) coefficients, two-dimensional (2D) preprocessed temporal multichannel EEGs, and 2D spectral and wavelet coefficients. Studies have also shown that the detection of IEDs by EEG frequency sub-bands could improve the efficiency of IED detection[7,13,14] as the sub-bands provide additional details on the underlying neuronal activities. Likewise, wavelet transform is proven to be valuable in IED detection,[13,15] and seizure analysis.[16] In wavelet analysis, the signal is convolved by a function identified as the mother wavelet, and the transform is calculated for various segments of the signal in the temporal and frequency domain, providing different time scales at various frequencies. Discrete Wavelet Transform (DWT) enables multiresolution signal processing, offering localization in both frequency and temporal domains.[17] Daubechies 4 (db4) wavelets show a strong correlation with IEDs among standard wavelet families.[15] Therefore, we included orthogonal wavelets from db4 to db10 from the Daubechies family for our analysis. CWT is better for the time–frequency analysis of EEG signals.[18] It allows us to split and examine instantaneous changes of time series signal at arbitrary wavelet scales and frequencies of interest.[19] We have selected Mexican hat and Morlet wavelets as mother wavelet due to their capability in exploiting nonstationary epileptic IEDs and suitability for spectral analysis.[18,20]

Training a deep ConvNet with a relatively small dataset can force the model to overfit on training samples, failing to generalize on the new or unseen dataset. Small datasets may also pose a more challenging mapping problem for models to learn, given their sparse sampling of data points in the high-dimensional feature space. One approach to making the high-dimensional feature space simpler to learn is to infuse noise to input data during network training.[21,22] The addition of Gaussian noise during the training of a model has a regularization effect that improves the generalizability of the neural network.[23] Therefore, we added Gaussian noise during the training of the IED detector.

We have reviewed the latest studies related to EEG-level classification in Table 1. In two of our preliminary studies,[8,9] we achieved a cross-validated area under the curve (AUC) of 0.87 (IED detection AUC of 0.935)[9] and 0.847,[8] respectively, by applying ConvNets for IED-based epileptic EEG classification. Roy *et al.* have achieved an accuracy of 86.6% with Recurrent Neural Network (RNN)-based ChronoNet[24] on the Temple University Hospital (TUH) EEG data corpus.[26] Antoniades *et al.* proposed CNN-based[27] and ensemble deep learning-based[28] IED detection on joint scalp-intracranial EEG data and achieved a mean accuracy of 89% and 68%, respectively. Spyrou *et al.*[29] proposed a combined tensor factorization and machine learning approach to analyze epileptic EEG data and reported a mean LOSO specificity of 60% (sensitivity of 70%). Lin *et al.*[30] presented a ConvNet-based classifier to distinguish EEG between children with epilepsy without epileptiform discharges and controls

Table 1. Recent ConvNet and deep learning studies for EEG-level classification.

Reference	Algorithm/ input shape	Classification type	Performance metrics		
			Evaluation criteria	EEG-level	
				AUC	BAC
Thomas *et al.*[9] (2018)	1D ConvNet with SVM 1×64	IED versus non-IED EEG	4-fold CV	0.870	83.86%
Jing *et al.*[8] (2019)	SpikeNet (2D ConvNet) 37×128	IED versus non-IED EEG	10-fold CV	0.847	—
Roy *et al.*[24] (2019)	ChronoNet (RNN) $22 \times 15{,}000$	Normal versus abnormal EEG	Fixed fold	—	86.57%
Peh *et al.*[25] (2021)	1D ConvNet 1×150	Slowing versus slow-free EEG	LOSO CV	0.851	81.8%
			LOIO CV	0.845	82.0%
Thomas *et al.*[10] (2021)	1D ConvNet with Ensemble 1×64	IED versus non-IED EEG	LOSO CV	0.826	76.10%
			LOIO CV	0.812	74.80%

Notes: Input shape (to ConvNet or RNN): Number of Channels $\times$ Sample Points.

and achieved a mean accuracy of 80%. These studies[8,9,24,27–30] were evaluated only on datasets from a single center. The joint scalp-intracranial EEG-based IED detection studies[27–29] were evaluated on 18 subjects only, and the results are inferior to our IED detection results. In another study, we achieved an IED false detection rate per minute (FDR/min) of 0.92 at 90% sensitivity.[6] In our recent multi-center study, we achieved a Leave-One-Institution-Out (LOIO) cross-validation (CV) Balanced Accuracy (BAC) of 75.5% and Leave-One-Subject-Out (LOSO) CV of 74.8% with ConvNet only system. We also achieved LOIO BAC of 76.1% and LOSO BAC of 69.3% with an ensemble of three components: ConvNet for detecting IEDs, Template Matching for detecting IEDs, and Spectral features for classifying EEGs.[10] In another multi-center study, Peh *et al.*[25] proposed an EEG classifier based on slowing and reported an LOIO BAC 82.0% and LOSO BAC of 81.8%. Marleen *et al.*[31] achieved an AUC of 0.94 with 2D ConvNet for IED detection. In these five studies,[6,8,10,25,30] the ConvNet detector is investigated with only the preprocessed EEG signal. In our preliminary study, we proposed a 1D ConvNet that exploited frequency sub-bands as features for the detection of IEDs, and we tested it on a dataset containing 554 EEGs.[7] We attain a mean five-fold IED detection AUC of 0.988 with a false detection rate per minute (FDR/min) of 0.230 at 90% sensitivity. None of the aforementioned studies have investigated the effect of noise during training.

To address these shortcomings, we systematically analyzed the ConvNet-based IED detector system on a large dataset (554 subjects, 18,164 labeled IEDs) recorded at Massachusetts General Hospital, (MGH)[6] with various input features, namely preprocessed EEG signal, frequency sub-bands, wavelet transforms, and multiple ConvNet variants, such as 1D ConvNet, 2D ConvNet, and multi-channel (MC) ConvNet. We also investigated the effect of injecting additive Gaussian noise at different ConvNet layers during training. Furthermore, to diagnose epilepsy, the clinicians may be only concerned with identifying a few highly probable epileptic spikes. There are only limited studies that attempt to classify normal EEGs (IED free EEGs without any anomaly) versus epileptic EEGs (EEGs with IEDs).[10] In these contexts, we also perform the EEG-level classification on five different institutions (using IED detection features derived from ConvNet) for classifying normal EEGs from epileptic EEGs. Based on what we know, this study could likely be the one among the very few to conduct a cross-institutional assessment. The proposed IED detector achieves a mean FDR/min of 0.23 for a sensitivity of 90% with 1D ConvNet. The EEG classifier achieves a best mean LOIO BAC of 78.1% (specificity of 76.2%, and AUC of 0.839) and LOSO BAC of 79.5% (specificity of 79.0%, and AUC of 0.856).

The remainder of the paper is divided into four sections. We describe the clinical scalp EEGs recorded from multiple centers in Sec. 2.1, and the preprocessing and methodology followed to extract the features in Sec. 2.2. In Secs. 2.3 and 2.4, we explain the design and evaluation of the proposed IED detection system, respectively. In Sec. 2.4, we also describe the EEG-level classification pipeline. In Sec. 3, we present the experimental results, while in

Sec. 4, we discuss the results, strengths and limitations of the proposed study. In Sec. 5, we present our conclusions and ideas for future study.

2. Methods

2.1. *Clinical scalp EEG datasets*

In this study, we consider routine clinical scalp EEGs from six centers spanning different countries: MGH (USA), National University Hospital (NUH, Singapore), National Neuroscience Institute (NNI, Singapore), TUH (USA),[26] Fortis Hospital Mulund (India), and Lokmanya Tilak Municipal General Hospital (LTMGH, India). The description about different datasets are presented in Table 2. Based on the clinical report, the EEGs are classified into epileptic EEGs (recorded from epileptic subjects containing IEDs) and normal EEGs (showing no abnormalities, recorded from subjects who do not have epilepsy). None of the EEG reports for the datasets considered in this study mentions nonepileptic IEDs. To retain the uniform mean EEG recording length across all centers, we restricted this study to EEGs with a recording length of 5–60 min. The EEG recordings were performed according to the standard 10–20 international system and were captured with different EEG equipment at various sampling rates. EEG recordings consisted of predominantly adult EEGs (age $>$ 20 years). We obtained ethical approval from the review boards of the respective centers.

2.2. *Data preprocessing and IED extraction*

We followed standard preprocessing techniques to extract different IED level features from the EEG recordings. At first, we applied a fourth-order Butterworth notch filter at 50/60 Hz to eliminate power line interference and a fourth-order 1 Hz high-pass filter to remove DC offset and baseline fluctuations. Next, we downsampled all the EEGs to 128 Hz and then applied the Common Average Reference montage to enhance the signal quality. Further, noise statistics-based artifact rejection technique is applied to remove high amplitude noise.[6] Next, we extract the IED and background segments (non-IEDs) from the preprocessed (prep) EEG data to train the ConvNet-based IED detector.

According to the International League Against Epilepsy (ILAE), IEDs range between duration of 0.02–0.2 s.[2] Therefore, the IEDs were extracted as 0.5 s ($0.5 \times 128 = 64$ samples) nonoverlapping segments from epileptic EEGs. The background

Table 2. Description of the six scalp EEG datasets.

Dataset/ country	Epileptic/nonepileptic						
	Number of subjects	Number of EEGs	Gender (age in years)			Fs (Hz)	Dataset type/ duration (min)
			Male	Female	Unknown		
MGH US	84	93	43 (35.2±27.2)	41 (37.1±28.2)	—	128, 200, 256	Private
	461	461	—	—	461 (31.2±9.8)		28.8±7.1
NUH SG	65	65	32 (50.3±20.2)	33 (56.4±19.7)	—	250	Private
	99	99	60 (48.8±17.9)	39 (50.9±20.5)	—		18.8±9.0
NNI SG	119	119	55 (44.5±19.0)	64 (47.0±21.2)	—	200	Private
	118	118	60 (44.0±16.8)	58 (51.2±18.4)	—		26.7±1.8
TUH US	42	260	12 (61.8±12.7)	30 (59.1±16.0)	—	200, 256, 500	Public
	30	44	13 (52.3±14.2)	17 (53.2±19.8)	—		14.5±7.3
Fortis IN	36	36	25 (37.0±14.7)	10 (38.4±17.4)	1 (25)	500	Private
	342	342	185 (48.5±18.1)	147 (47.2±17.3)	10 (41.0±18.5)		20.7±5.5
LTMGH IN	44	44	26 (51.2±24.3)	18 (46.3±24.7)	—	256	Private
	626	626	365 (41.0±16.7)	261 (41.4±19.0)	—		13.9±1.5
Total	390	617	—	—	—	—	—
	1676	1691					801.16 h

Notes: All the EEGs contain 19 channel recordings. F_s: sampling frequency. Age/duration are presented as mean±standard deviation.

segments were extracted from normal EEGs as 0.5 s segments with 75% overlapping window. We extracted the background segments only from EEGs without IEDs. The ConvNet-based IED detectors are then trained with these 0.5 s single-channel waveforms to classify them as IED or non-IED segments. Information about the various folds is shown in Table 3. In addition to the prep EEG signals as input features to the ConvNet-based IED detector, we also explore other signal representations: frequency sub-bands (δ, θ, α, β), CWT (db4–db10, scales D1–D5 and A1–A5), and DWT (Morlet and Mexican Hat, scales 1–10).

2.2.1. *Frequency sub-bands*

EEG signals can be filtered according to different frequency bands, providing valuable information for detecting IEDs. In this study, we applied a Kaiser window[32] based bandpass filter on the preprocessed EEG recordings, filtering the signals according to the five frequency bands: delta (δ: 1–4 Hz), theta (θ: 4–8 Hz), alpha (α: 8–13 Hz), and beta (β: 13–30 Hz). The bandpass filter is initially designed by the Matlab R2019b function, $bandpass(X, Fpass, Fs)$, where X is the input signal, $Fpass$ is the pass-band frequency range, and Fs is the sample rate of the input signal. After applying the bandpass filter according to different frequency bands on the entire EEG signal, we extracted 0.5 s segments from each of the filtered signals. The use of 1 Hz fourth-order high-pass filter during data preprocessing attenuates the delta band slightly; however, the filter is essential in removing baseline wander.

2.2.2. *Discrete wavelet transforms*

The DWT decomposes the input signal into various approximate and detailed coefficients to investigate the data at different frequencies and time interval windows. DWT utilizes wavelet and scaling functions related to highpass filters (HPF) and lowpass filters (LPF). The signal is deconstructed into different frequencies by consecutive HPF and LPF of the input signal. The input signal $x[i]$ is fed into a half-band HPF $h[i]$ and LPF $l[i]$. According to the Nyquist law, half of the samples can be discarded after filtering. Therefore, the signal can be sub-sampled by a factor of 2, just by dropping every alternative sample. This represents one decomposition step.

Table 3. MGH 5-fold cross-validation data distribution.

Fold	Epileptic EEGs	Nonepileptic EEGs	Annotated IEDs	IED events
1	19	92	4077	2920
2	19	92	3571	2757
3	18	92	3207	2831
4	19	92	4021	2781
5	18	93	3288	2881
Total	93	461	18,164	14,170

It is crucial to choose a mother wavelet that resembles IEDs. In this study, we applied the Daubechies family from db4 to db10 as they have a strong correlation to IEDs. For each candidate IED and non-IED of 500 ms (64 samples), DWT was performed to extract the feature vectors. Following the literature,[15] we extract coefficients from levels 1 to 5, corresponding to frequencies from 0 Hz to 128 Hz. Level 1 decomposition corresponds to 0–64 Hz and 64–128 Hz for approximate and detailed coefficients, respectively. Level 2 decomposition corresponds to the frequency range of 0–32 Hz and 32–64 Hz for approximate and detailed coefficients, respectively, and so on. This procedure resulted in coefficient vectors of different lengths at each level. To obtain coefficient vectors of the same length, we resample each vector into a 64-dimensional vector by 1D linear interpolation. In Fig. 1, we depict the detailed and approximated coefficients of level-5 wavelet.

2.2.3. *Continuous wavelet transforms*

The CWT measures the similarity (inner-product) between a signal and an analyzing function. In this study, we examined the Mexican hat and Morlet wavelets, which reflect the morphological characteristics of the original IEDs.[18,20] From each IED and non-IED segment of 500 ms (64 samples), we extracted CWT coefficients at ten scales (from 1 to 10). This procedure resulted in coefficients of different lengths at each level. Next, we resample the coefficients at each level to obtain a vector of length 64. The wavelet scales of a randomly selected IED and non-IED segment are shown in Fig. 1. The wavelet scales and corresponding frequency at each scale are listed in Table 4. The scales correspond to different frequencies for the Mexican hat and Morlet wavelet due to the fundamental difference in mother

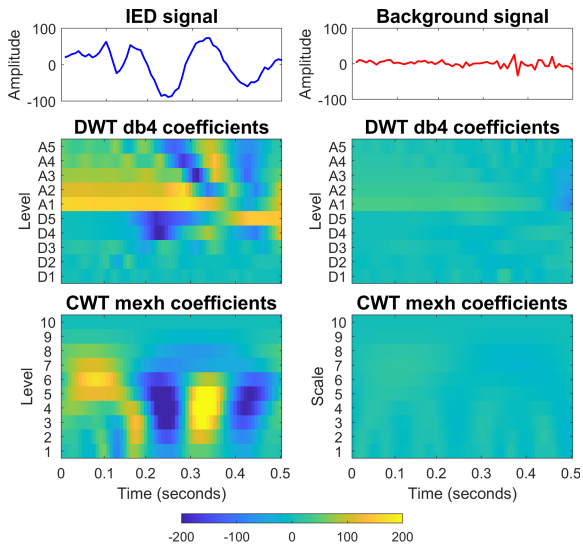


Fig. 1. db4 and Mexicant hat wavelet coefficients (magnitude) for IED and background segments.

Table 4. CWT scales and the corresponding frequencies.

No.	Scale	Morl frequency	Mexh frequency
1	0.016	61.115	16.105
2	0.022	43.591	11.487
3	0.031	31.092	8.193
4	0.043	22.177	5.844
5	0.06	15.818	4.168
6	0.085	11.282	2.973
7	0.119	8.047	2.121
8	0.166	5.740	1.513
9	0.233	4.094	1.079
10	0.327	2.92	0.769

Notes: morl: Morlet Wavelet; mexh: Mexican Hat Wavelet.

wavelets. Therefore, we have also investigated the scales (from 4 to 10) that contain frequency components below gamma band and high clinical relevance.[33] We also select Morlet scales from 5 to 10 and Mexican hat scales from 1 to 6 as they translate to similar frequency range. The workflow of the proposed approach is shown in Fig. 2.

2.3. *ConvNet-based IED detector*

We trained the IED detectors to estimate the probability of 0.5 s waveforms being an IED, whereas the EEG classifiers are built on features derived from IED detectors. We take the maximum of the output from 19 channels to merge the 19 IED detector predictions (for each channel) into one final output for each 0.5 s time segment. We extract the IED rate per minute for each EEG at various prediction thresholds. The EEG classifiers are built on these IED rates.

Along the lines of the ConvNet architecture suggested by Thomas *et al.*,[6] we develop the 1D ConvNet IED detector. To create the feature maps for 1D ConvNet and 2D ConvNet, we initially utilize 1D and 2D convolutionary filters. Then, we adopt the

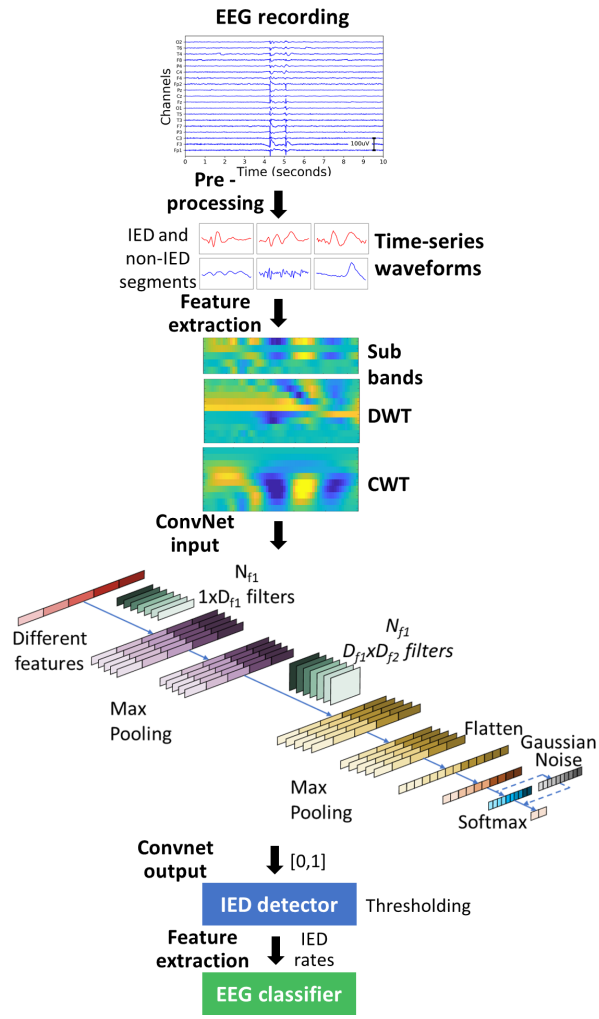


Fig. 2. Overall workflow of the ConvNet IED detector and EEG classifier. The proposed 1D ConvNet architecture has two convolutional, pooling and fully connected layers with multiple input features. During training, Gaussian noise is added to the last fully connected layer, before the softmax function.

activation function of rectified linear units (ReLU) to solve the issue of vanishing gradients and max-pooling for reduction of dimensionality. Next, the pooled feature maps are flattened into a single long vector and passed into a dense layer or fully connected layer (FC). The softmax is applied as the last activation function to normalize the FC outputs to 0 and 1, with predicted probability output '1' pointing to an IED. A representation of the ConvNet IED detector is depicted in Fig. 2.

We balanced the ConvNet training by maintaining an identical number of IED and background segments in each batch. We implemented dropout (with dropout rate set to 0.5) on each of the dense layers during the ConvNet training to eliminate the high variance issue. We selected the training samples randomly in batches. We specifically include non-IED waveforms in the training set that are wrongly labeled as IED in the first round of classification, as proposed by Thomas *et al.*[6,9] The optimal ConvNet hyperparameters are selected by performing Cross-Validation on the training data. The network is trained for each fold until the iteration at which the validation loss becomes minimal, which is the early stopping criterion. Table 5 summarizes the different parameter values and ranges evaluated for optimizing the ConvNet. The ConvNet was executed in Tensorflow 1.5.1.[34]

We investigate different types of ConvNet, followed by the data augmentation with noise injection.

- One-dimensional (1D) ConvNet: We evaluate the 1D ConvNet with different combinations of features: prep EEG, sub-bands, CWT, and DWT. We stack the different 64 sample feature coefficients extracted beside one another horizontally, leading to a long 1D vector. The input to the 1D ConvNet with multiple features is of dimension $\mathcal{R}^{1\times(64\times N_f)\times 1}$ where N_f represents the number of different features fed into the ConvNet.
- Two-dimensional (2D) ConvNet: Similarly as 1D ConvNet, the 2D ConvNet is built with different feature combinations. However, the significant differentiator is that instead of appending additional features horizontally, they are stacked vertically, making the input a 2D matrix. This configuration enables the ConvNet filters to leverage both temporal and spectral features simultaneously. The input to the 2D ConvNet has a dimension of $\mathcal{R}^{N_f\times 64\times 1}$.
- Multi-channel (MC) ConvNet: For MC ConvNet, instead of stacking different 64 sample features horizontally or vertically, we append them as additional channels mimicking RGB image where each color is represented as a separate channel. This generates multiple feature maps during convolution, which are then added up into one final feature map. The input to this MC ConvNet will have dimension of $\mathcal{R}^{1\times 64\times N_f}$.

Table 5. Hyperparameters evaluated for optimizing the ConvNet IED detector.

Parameters	Range/values
Number of convolution and pooling layers	1, 2, 3, 4
Number of fully connected layers	1, 2, 3
Number of convolution filters	32, 64, 128, 200
Dimension of convolution filters $D_f \in \{3, 5, 7\}$	$1 \times D_f \times 1$ (1D) $D_f \times D_f \times 1$ (2D) $1 \times D_f \times D_f$ (MC)
Number of hidden layer neurons	512, 128, 64
Activation function	ReLU
Dropout rate	0.5
Maximum number of iterations	20,000
Optimizer	Adam
Learning rate	10^{-4}
Loss function	Cross-entropy
Weights initialization	Xavier[35]
Layers trained with Gaussian noise	Input layer, FC1, FC2, before softmax
Gaussian noise, σ	0.001, 0.003, 0.005, 0.01, 0.02, 0.03, 0.04, 0.05, 0.10

- ConvNet with noise injection: The IED detector should be capable of handling input signals with perturbations. Neural networks are proven not to be very robust to noise. One way to improve the robustness of the neural network is to train them with random noise applied to their input or hidden units or at the output level.[22] Noise injection can be seen as augmenting the data at multiple levels of abstraction. Provided that the magnitude of the noise is carefully tuned, this approach can be highly effective. The most prevalent kind of noise adopted during the training of machine learning algorithms is injecting Gaussian noise to input variables, with zero mean and a given variance.[23] The amount of noise injected (standard deviation) is a configurable hyperparameter. Very low noise has no effect, whereas very high perturbation makes the mapping function too challenging for the network to learn. Noise is only added during training. Although injecting noise to the input signal is the most well-known and broadly contemplated approach, we added noise at different layers (input, convolutional, fully connected, and output layer). The proposed data augmentation approach with noise at the output layer is shown in Fig. 2.

2.4. *Performance assessment*

2.4.1. *IED detection*: 5-*fold CV on MGH dataset*

The IED detectors are trained with different input features (prep signal, frequency sub-bands, CWT coefficients, and DWT coefficients) and various ConvNet architectures. Then, we compare the performance of the different IED detectors by carrying out 5-fold CV on the annotated EEGs from MGH, where four folds are allocated for training and one for testing. We split the MGH dataset into five folds (see Table 3), more or less matching distribution of labeled IEDs, gender, and age. We split the folds in the same way as reported by Thomas *et al.*[6] We assign all the EEG recordings from the same subject into the same fold to ensure that EEG recordings of the same subject be either in testing or training at any given time. We evaluate the performance of the IED detectors based on four metrics: AUC, AUPRC, PPV, and FDR/min. Since the two classes are highly disproportional (background to IED segments ratio is around 1000:1), area-related metrics such as area under the precision-recall curve (AUPRC) and AUC could be deceiving. Considering this, we investigate the FDR/min for preset sensitivity values. The results reported are averages computed from 5-fold CV and, therefore, tend to be robust against stochasticity.

2.4.2. *EEG classification*: *LOIO and LOSO CV*

Next, we evaluate the efficiency of EEG classifier based on features extracted from IED detector. This is a two-stage process. In the first stage, we train, optimize, and evaluate the IED detector (ConvNet-based IED detectors with various input features and ConvNet architectures). In the second stage, we extract feature from EEGs based on the ConvNet IED detector, train the EEG classifier, find the best ConvNet threshold based on BAC on the training set, and analyze the performance of the EEG classifier.

In order to evaluate the datasets from different centers, we train and validate (to determine the stopping criteria and best hyperparameters) the ConvNet-based IED detectors on all five folds of MGH dataset. The model with minimal validation error is selected for further evaluation. We conduct the multi-center EEG classification in two ways, as illustrated below.

We conduct the LOIO CV by omitting the dataset from a target institution completely. We train the EEG classifier on the remaining institutions and test on the EEG dataset from the left-out center (see Fig. 3). To avoid class imbalance in any dataset during training, we randomly choose the same number of EEG recordings from each institution and merge them together. The maximum EEGs we select from each dataset is configured as the minimum of normal/epileptic EEG recordings available across all

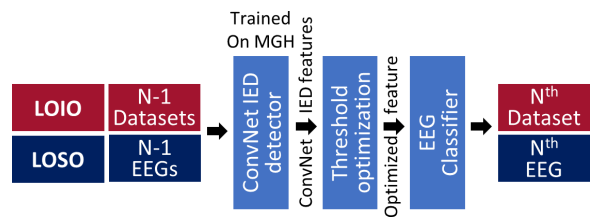


Fig. 3. LOIO and LOSO CV process on multiple datasets.

institutions to avoid institutional bias. During LOIO CV, we examine the EEGs from LTMGH only during testing, since EEGs from LTMGH are captured from nonstandard medical devices. We present the most widely used performance evaluation metrics, including AUC, BAC, and specificity.[10]

While performing the LOSO CV on each dataset separately, the classifier is trained by leaving an EEG and testing on the left out EEG (see Fig. 3). Finally, we consolidate and present the aggregated results. We report BAC and specificity at 80% sensitivity. When performing LOIO and LOSO CV assessments, the IED detectors are built on the entire MGH dataset, and they remain fixed across all the institutions. We utilize the training fold on EEG-level to develop the threshold-based-IED module and optimize the ConvNet threshold at which IED rates are extracted. For both LOIO and LOSO CV, we run 100 experiments in order to address the stochasticity introduced by random selection of training samples.

3. Results

3.1. 5-*Fold CV on MGH dataset*

We applied 5-fold CV on EEGs from MGH to assess the performance of 1D ConvNet IED detection system with different configurations of input features. The evaluation results are listed in Table 6. The input configurations are presented in Table 6, columns 1 to 4. Input features 'prep EEG' represents IED detector with only time-series EEG segments, 'prep EEG, δ, θ, α, β' represents IED detector with time-series EEG segment along with frequency subbands, configuration 'prep EEG, db4 (A1–A5, D1–D5) represents the IED detector with a combination of time series, and db10 approximate and detailed coefficients from level 1 to 5, and so on.

The IED detector performance is measured on the basis of mean FDR/min and precision at 90% sensitivity. The 1D ConvNet IED detector based on various frequency bands (prep EEG, δ, θ, α, β) and DWT-based coefficients (prep EEG, db10 (A1–A5, D1–D5)) provided the best overall results. The 1D ConvNet with prep EEG as features achieved a mean FDR/min of 0.36 and a precision of 0.748 and will act as a baseline. The 1D ConvNet with multiple features based on prep EEG and frequency bands (δ, θ, α, β) achieved a mean FDR/min of 0.230 and a precision of 0.790. With prep EEG, and db10 (A1–A5, D1–D5), the system achieved a mean FDR/min of 0.262 and a precision of 0.799. This illustrates that carefully selected multiple features such as frequency bands and wavelets, coupled with time-series signal, are useful to avoid false positives in IED detection. The sensitivity versus FDR/min plot for the best features from each family (prep EEG, sub-bands, DWTs, and CWTs) is displayed in Fig. 4. With the

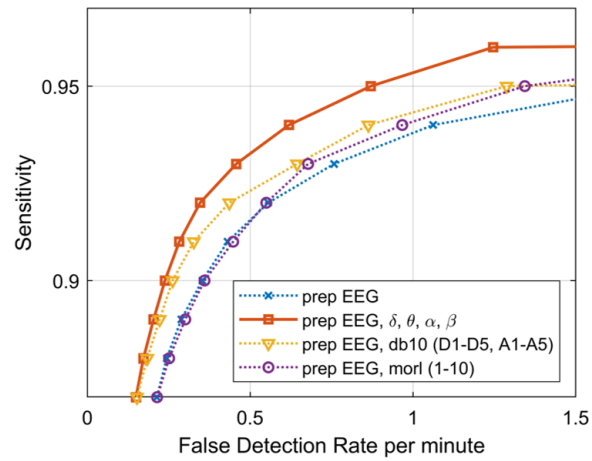


Fig. 4. Sensitivity versus FDR/min for the best features from each family (prep EEG, sub-bands, DWTs, and CWTs).

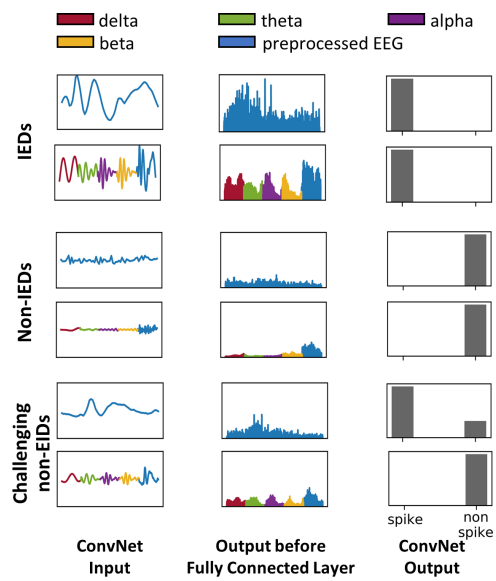


Fig. 5. ConvNet activations before the fully connected layer and predicted outputs for IED, non-IED and challenging non-IED inputs to 1D ConvNet with preprocessed EEG (first rows) and 1D ConvNet with preprocessed EEG, δ, θ, α, β (second rows).

combination of best features from each family (prep EEG, sub-bands, DWT, and CWT), we achieved an FDR/min of 0.297.

We selected the best performing feature combination (prep EEG, δ, θ, α, β) and evaluated them with different ConvNet types, namely 1D ConvNet, 2D DNN, and MC ConvNet. 2D ConvNet and MC ConvNet achieved a mean FDR/min of 0.270 (precision of 0.770) and 0.320 (precision of 0.760), respectively. Although all 3 ConvNet types performed better than the baseline, the 1D ConvNet IED detector trained on multiple frequency sub-bands surpassed the 2D ConvNet and MC ConvNet. We compared the contribution of neurons before the fully connected layer of the best performing ConvNet IED detector system (see Fig. 5). We observed that the preprocessed signal contributed the most, followed by other sub-band components.

We trained the 1D ConvNet with prep EEG with Gaussian noise at different layers (input, FC1, FC2, and output layer) and different standard deviation values. Their results are presented in Table 7. The

Table 6. IED detection results for different ConvNet input feature configurations.

Input dim $\mathcal{R}^{x\times y\times z}$	1D ConvNet input features	Performance measures			
		AUC	AUPRC	Precision	FDR/min
1×64×1	prep EEG	0.987	0.876	0.748	0.353
1×320×1	prep EEG, δ, θ, α, β	0.988	0.902	0.790	**0.230**
1×320×1	db4 (D1–D5)	0.986	0.888	0.747	0.359
1×320×1	db5 (D1–D5)	0.988	0.887	0.733	0.38
1×320×1	db6 (D1–D5)	0.988	0.871	0.723	0.395
1×320×1	db7 (D1–D5)	0.987	0.885	0.729	0.399
1×320×1	db8 (D1–D5)	0.986	0.895	0.744	0.367
1×320×1	db9 (D1–D5)	0.987	0.877	0.712	0.419
1×320×1	db10 (D1–D5)	0.987	0.89	0.733	0.389
1×384×1	prep EEG, db4 (D1–D5)	0.984	0.895	0.76	0.336
1×384×1	prep EEG, db5 (D1–D5)	0.988	0.895	0.766	0.32
1×384×1	prep EEG, db6 (D1–D5)	0.986	0.901	0.781	0.292
1×384×1	prep EEG, db7 (D1–D5)	0.987	0.89	0.75	0.352
1×384×1	prep EEG, db8 (D1–D5)	0.987	0.897	0.768	0.321
1×384×1	prep EEG, db9 (D1–D5)	0.987	0.891	0.767	0.317
1×384×1	prep EEG, db10 (D1–D5)	0.986	0.897	0.781	0.295
1×640×1	db4 (D1–D5, A1–A5)	0.988	0.893	0.755	0.342
1×640×1	db5 (D1–D5, A1–A5)	0.989	0.889	0.749	0.35
1×640×1	db6 (D1–D5, A1–A5)	0.988	0.897	0.757	0.338
1×640×1	db7 (D1–D5, A1–A5)	0.988	0.874	0.742	0.364
1×640×1	db8 (D1–D5, A1–A5)	0.985	0.9	0.77	0.315
1×640×1	db9 (D1–D5, A1–A5)	0.988	0.897	0.783	0.29
1×640×1	db10 (D1–D5, A1-A5)	0.986	0.897	0.764	0.319
1×704×1	prep EEG, db4 (D1–D5, A1–A5)	0.988	0.896	0.764	0.319
1×704×1	prep EEG, db5 (D1–D5, A1–A5)	0.986	0.893	0.747	0.358
1×704×1	prep EEG, db6 (D1–D5, A1–A5)	0.988	0.907	0.768	0.32
1×704×1	prep EEG, db7 (D1–D5, A1–A5)	0.987	0.885	0.754	0.341
1×704×1	prep EEG, db8 (D1–D5, A1–A5)	0.988	0.878	0.746	0.356
1×704×1	prep EEG, db9 (D1–D5, A1–A5)	0.987	0.9	0.792	0.272
1×704×1	prep EEG, db10 (D1–D5, A1–A5)	0.984	0.907	0.799	0.262
1×384×1	mexh (1–6)	0.986	0.869	0.665	0.558
1×448×1	prep EEG, mexh (1–6)	0.985	0.893	0.722	0.428
1×704×1	prep EEG, mexh (1–10)	0.987	0.88	0.712	0.432
1×448×1	prep EEG, morl (5–10)	0.989	0.885	0.72	0.394
1×512×1	prep EEG, morl (4–10)	0.989	0.891	0.74	0.359
1×704×1	prep EEG, morl (1–10)	0.988	0.865	0.65	0.595

Notes: Precision and FDR/min are reported for 90% sensitivity.

Table 7. IED detection results for 1D ConvNet with prep EEG trained with Gaussian noise at different layers and different standard deviations.

ConvNet layers trained with Gaussian Noise	FDR/min for sensitvity = 90%								
	For Gaussian noise, σ =								
	0.001	0.003	0.005	0.01	0.02	0.03	0.04	0.05	0.1
Input layer	0.367	0.373	0.375	0.389	**0.341**	0.374	0.367	0.363	0.344
FC1	0.408	0.347	**0.340**	0.360	0.370	0.369	0.393	0.359	0.425
FC2	0.425	0.367	0.377	0.364	0.390	0.344	**0.336**	0.423	0.398
Before softmax	0.357	0.390	0.368	**0.325**	0.335	0.392	0.426	0.332	0.355
FC2, before softmax	**0.336**	0.360	0.370	0.367	0.343	0.351	0.366	0.361	0.402
FC1, FC2, before softmax	0.365	0.331	0.394	0.353	0.376	**0.329**	0.368	0.367	0.330
Input layer, FC1, FC2, before softmax	0.410	0.355	0.368	0.389	**0.341**	0.374	0.367	0.363	0.344

1D ConvNet with prep EEG trained with Gaussian noise at the output layer (before softmax) with σ = 0.01 attained a mean FDR/min of 0.32, performing better than 1D ConvNet without any data augmentation. This illustrates that conscientiously chosen noise parameters help to minimize FDR/min in IED detection. The additional parameters besides the range/values mentioned in Table 5 did not improve the EEG level classification, and as the computational complexity of tuning, all those parameters become unwieldy, we ignored those parameters in further analysis.

3.2. *LOIO and LOSO CV*

We selected the best ConvNet IED detector type and input feature combination based on IED detection results (FDR/min). The particular ConvNet is

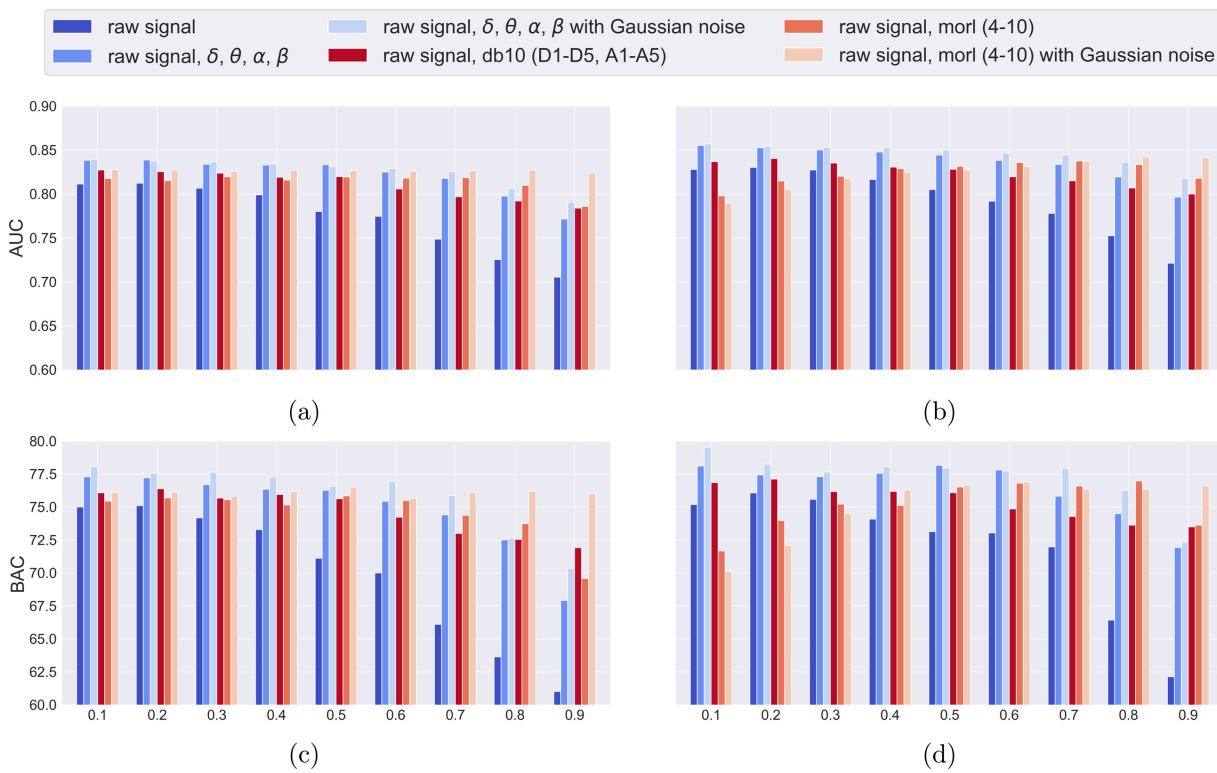


Fig. 6. LOIO and LOSO CV (AUC and BAC) for the five datasets for various ConvNet thresholds and input features: (a) LOIO AUC, (b) LOSO AUC, (c) LOIO BAC, and (d) LOSO BAC.

Table 8. LOIO and LOSO CV results for multiple institutions (AUC and BAC) with features extracted from 1D ConvNet-based IED detector.

ConvNet input features	Data augmentation Gaussian noise	LOIO CV		LOSO CV	
		AUC	BAC (specificity)	AUC	BAC (specificity)
prep EEG	—	0.811	75.0%±5.0 (70.0)	0.827	75.2%±5.4 (70.4)
prep EEG	$\sigma = 0.01$	0.817	75.3%±4.5 (70.6)	0.832	75.9%±5.0 (71.8)
prep EEG, δ, θ, α, β	—	0.838	77.3%±5.8 (74.6)	0.855	78.1%±5.4 (76.2)
prep EEG, $\boldsymbol{\delta}$, $\boldsymbol{\theta}$, $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$	**$\boldsymbol{\sigma = 0.01}$**	**0.839**	**78.1%±6.3 (76.2)**	**0.856**	**79.5%±6.3 (79.0)**
prep EEG, db10 (D1–D5, A1–A5)	—	0.827	76.1% ±3.6(72.2)	0.836	76.9%±4.1 (73.8)
prep EEG, db10 (D1–D5, A1–A5)	$\sigma = 0.01$	0.828	76.1% ±3.8 (72.0)	0.844	77.1%±5.4 (74.2)
prep EEG, morl (4–10)	—	0.818	75.4%±6.7 (70.9)	0.800	71.7%±6.6 (63.4)
prep EEG, morl (4–10)	$\sigma = 0.01$	0.827	76.1%±6.2 (72.2)	0.800	70.1%±6.4 (60.2)

Notes: BAC and Specificty are reported for 80% sensitivity.

chosen based on the lowest validation loss during ConvNet training. The mean AUC, BAC, and specificity results for the different datasets and eight input configurations applied for a fixed 1D ConvNet for 100 random instances of EEG classification are summarized in Table 8. The consolidated evaluation results on five institutions show that the ConvNet-based IED detector trained on various features (sub-bands or DWT) performed better than the ConvNets that operate directly on the prep EEG signals. The mean LOIO CV and LOSO CV results obtained for five different institutions at various ConvNet thresholds are shown in Fig. 6.

The 1D ConvNet trained on prep EEG achieved a mean LOIO CV BAC of 75.0% (AUC of 0.811), and LOSO CV BAC of 75.2% (AUC of 0.827), which will be our baseline. For the best system configuration, the suggested workflow has attained a mean LOIO CV BAC of 78.1% (AUC of 0.839) and LOSO CV BAC of 79.5% (AUC of 0.856). The combination of the preprocessed signal with frequency bands together with Gaussian noise at the output layer performed the best. From Fig. 6 and Table 8, we see that the suggested pipeline generalizes effectively across multiple datasets.

4. Discussion

4.1. *IED detection: Comparison of system configurations*

The ConvNet with prep EEG, δ, θ, α, β and prep EEG, db10 (A1–A5, D1–D5) achieved better performance than other combinations of 1D ConvNet systems for detecting IEDs. The ConvNet with prep EEG, δ, θ, α, β delivered the best FDR/min of 0.230 (AUPRC of 0.902) followed by ConvNet with prep EEG, db10 (A1–A5, D1–D5) with FDR/min of 0.26 (AUPRC of 0.9) at 90% sensitivity. The ConvNet with prep EEG achieved an FDR/min of 0.353 (AUPRC of 0.876), lower performance than ConvNet trained on prep EEG with sub-band or db10 wavelets. ConvNet, based on both continuous wavelets (Mexican and Morlet) with scales from 1 to 10, leads to the lowest performance. By training the ConvNet only with prep EEG and Morlet scales 4 to 10 that contain frequency components of clinical interest,[33] we achieved an FDR/min of 0.359, similar to ConvNet trained with prep EEG.

As there is substantial variation in the IED morphology across patients and within the EEGs of the same patient, the IEDs might not be correlating well with the wavelet families we have evaluated. Moreover, wavelets might be more sensitive to certain types of artifacts. As the 1D ConvNet (input dimension: $\mathcal{R}^{1\times320\times1}$) with prep EEG combined with sub-bands (δ, θ, α, β) provided the best IED detection results when compared to DWT- and CWT-based features, we investigated the prep EEG, δ, θ, α, β features with different ConvNet input combination, namely 2D ConvNet (input dimension: $\mathcal{R}^{5\times64\times1}$) and MC ConvNet(input dimension: $\mathcal{R}^{1\times64\times5}$). Both 2D ConvNet and MC ConvNet performed worse than 1D ConvNet with multiple features. As the feature representations are not extracted in a fixed linear range, the 2D or MC convolutional filters might be ineffective in capturing the additional information. However, irrespective of the ConvNet type, ConvNet (1D ConvNet, 2D ConvNet, or MC ConvNet) with

multiple sub-bands consistently outperformed 1D ConvNet with prep EEG in terms of AUPRC, precision, and FDR/min. This shows that utilizing various EEG frequency bands improves the IED detection by reducing FPs, similar to what has been reported in an earlier study.[7] The 1D ConvNet with prep EEG and Gaussian noise of $\sigma = 0.01$ at the output layer during training performed better than other combinations of noise and achieved an FDR/min of 0.325 at 90% sensitivity and AUPRC of 0.888. When the noise is added at FC1, FC2 and output layer with $\sigma = 0.03$ and 0.1, we achieved an FDR/min of 0.329 and 0.330, respectively. This shows that noise injection could potentially improve the overall system performance.

We can observe the contribution of neurons before the fully connected ConvNet layer (after max-pooling from the second convolutional layer) in Fig. 5. For the IED segments, the neurons before the fully connected layer are highly active in the prep EEG band, accompanied by other sub-bands. Several neurons have high activities, thereby leading effectively to the detection of the IEDs. However, for non-IEDs, most of the neurons before the fully connected layer are dormant, thereby predicting the segment as non-IED. On the other hand, a non-IED or an IED-like artifact could quickly produce high neuronal activity in the prep EEG part before the fully connected layer, hence causing the prep EEG signal-based IED detector to misclassify them as potential IED segments. As our proposed system incorporates information from multiple features, neurons in the sub-bands exhibit minimal activity during convolution, thereby giving additional valuable information to the neurons in the fully connected layers to make proper voting on the final prediction. This supplementary information, which is not so evident in the full-spectrum EEG data, aids in the successful detection of challenging background segments and artifacts.

4.2. *EEG classification: Comparison of system configurations*

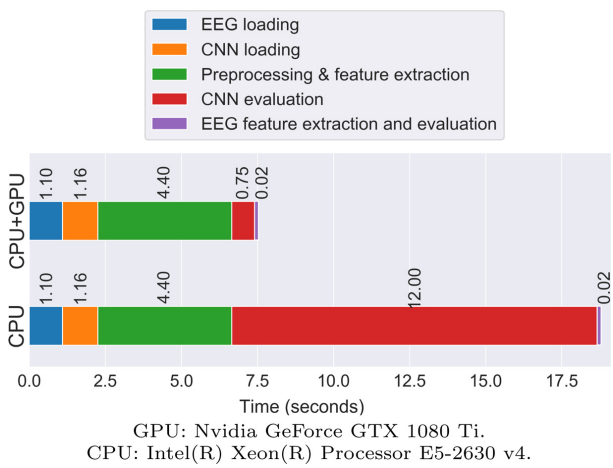
We have implemented an automated diagnostic system for epilepsy and have tested it on routine clinical EEGs recorded from multiple institutions. We analyzed the system in two ways: LOIO CV and LOSO CV. The proposed approach generalizes well across datasets from different centers. Hence, they are expected to perform well on datasets from centers that did not participate in this study. The proposed 1D ConvNet-based multi-feature system is evaluated on the dataset from five distinct centers and has attained a mean LOIO BAC of 78.1% (AUC of 0.839) and LOSO BAC of 79.5% (AUC of 0.856). In terms of absolute numbers, we have correctly classified 73 and 93 additional EEGs in LOIO and LOSO, respectively, compared to baseline. The system proposed in this study yields high classification accuracy across a wide range of ConvNet thresholds applied to derive ConvNet IED rates for EEG level classification. This enables the clinicians to set different operating points (thresholds) to extract ConvNet IED rates by making a trade-off between high sensitivity and FDR/min on the IED level and high BAC on the EEG level.

4.3. *Comparison with the state-of-the-art literature*

The inter-rater agreement (IRA) reported for classifying epileptic EEGs are 80.9%,[36] 77.0%,[37] 74.0%,[38] and 88.6%[39] in the current literature (see Table 6 from Thomas *et al.*[10]). IRA reported by Piccinelli *et al.* is performed only on pediatric patients with idiopathic epilepsy. The proposed approach obtained an LOIO CV BAC of 78.1% and LOSO CV BAC of 79.5%, which is in good accordance with the IRA of human experts.[10] The efficiency of EEG classification across various centers (LOSO CV BAC) achieved in our cross-institutional analysis is superior to those described in the previous works, and the proposed system is reaching human-level performance. The majority of the EEG classification studies in the literature have one or more of the following limitations: performed on the homogenous datasets, performed by experts with similar training or from the same center, analyzed on idiopathic generalized epilepsy of childhood, and EEG with seizures that is easier to detect than EEG with IEDs or between seizures. In our recent multi-center study,[10] we achieved a LOIO CV BAC of 75.5% (specificity of 71.0%) and LOSO CV of 74.8% (specificity of 69.6%) with ConvNet feature-based EEG classification system. This study has achieved up to 4.3% improvement in BAC (8.6% improvement in specificity). The majority of the studies focus on one specific type of feature and ignore the potential of data augmentation. Deep ConvNets rely heavily on big data to

prevent overfitting, but big data is scarce in the field of EEG analysis. Data augmentation techniques[40] such as neural style transfer learning,[41] feature space augmentation, noise boosting, generative adversarial network (GAN)[42] have not been extensively studied for the problem of EEG classification. This would be an interesting direction to pursue in the future.

It takes about 10 min to train the CNN IED detector on a GPU. The computational time of the proposed system for evaluating an EEG is shown in Fig. 7. The proposed system approximately takes 4.5 ± 0.07 s for preprocessing and 0.75 ± 0.01 s on a CPU+GPU system for the ConvNet IED detector to evaluate a 19-channel 30-min scalp EEG collected at a sampling rate of 128 Hz. On the other hand, it takes around 12.0 ± 0.11 s for evaluation on a CPU only system. It takes about 10 mins for an expert to review a similar routine scalp EEG recording. On the other hand, our IED detection and EEG classification system can evaluate the EEG in few seconds, providing invaluable information to the neurologists, thus proving to be an effective tool for epilepsy diagnosis. The assessment was carried out in Python v3.5, on a Nvidia GeForce GTX 1080 Ti GPU and Intel(R) Xeon(R) Processor E5-2630 v4 CPU. As the IED detectors are developed to predict the waveforms at channel level, the recommended framework is generalizable to clinical scalp EEG recorded with arbitrary number of channels. This has a potential significance in the realm of wearable devices used for continuous health monitoring, especially to detect and diagnose epilepsy. Defective or EEG channels with low signal-to-noise ratio can also be discarded when detecting IEDs.



GPU: Nvidia GeForce GTX 1080 Ti.
CPU: Intel(R) Xeon(R) Processor E5-2630 v4.

Fig. 7. Computational time for evaluating a 19-channel 30-min EEG recorded at 128 Hz.

4.4. *Limitations of the study*

The study has a few limitations. First, the study does not leverage spatial information when combining information from multiple channels. Second, the IEDs applied to build the IED detector are labeled by clinicians from the same center. Third, data augmentation techniques such as GAN and noise boosting are not extensively studied. Finally, this study is mainly focused on identifying epilepsy based only on interictal IED events.

5. Conclusion

We have developed a generalized epileptic EEG diagnostic tool that efficiently detects IED patterns. We have cross-validated the proposed system on six institutions and attained a mean LOIO CV BAC of 78.1% (AUC of 0.839) and LOIO CV BAC of 79.5% (AUC of 0.856). Hence, the recommended EEG classification pipeline could serve as an invaluable tool to assist clinicians in the rapid analysis of EEGs with epilepsy. This system could also be employed to monitor the effects of antiepileptic drug administration on the characteristics of interictal epileptiform events.

We could interpolate the EEG signals as a 2D image grid in the ConvNet, independent of the number of channels. We can also improve the IED detector performance and reach the human-expert IRA by training it with more labeled quality data collected from multiple centers and several neurologists. Once the automated system reaches the IRA for epileptic EEG interpretation, the system can be considered clinically more viable. Noise injection and other forms of data augmentation and regularization is an important field that it merits its own separate study. This study can be extended to detect ictal events and other abnormalities in both routine EEG and ambulatory and critical care EEG. Although the primary epilepsy diagnostics for routine clinical scalp EEG is typically based on IED, other potential epilepsy linked features of EEG such as seizures, HFOs, slowing, and PLEDs can be extensively investigated in the future to identify epilepsy.

Acknowledgments

The data collection at M. G. H. was carried out under the guidance of Dr. Cash, Dr. Jing, and Dr. Westover. The NNI and NUH dataset collection were 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 A . Acronyms

Table A.1. Acronyms.

EEG	Electroencephalogram
IED	Interictal epileptiform discharge
ConvNet	Convolutional neural network
CV	Cross-validation
AUC	Area under the curve
AUPRC	Area under Precision-Recall curve
ROC	Receiver operating characteristics
FP	False Positive
BAC	Balanced accuracy
LOSO	Leave-one-subject-out
LOIO	Leave-one-institution-out
1D/2D	One-/two-dimensional
MC	Multi-channel
MGH	Massachusetts General Hospital
NUH	National University Hospital
NNI	National Neuroscience Institute
TUH	Temple University Hospital
LTMGH	Lokmanya Tilak Municipal General Hospital
db	Daubachies
mexh	Mexican Hat
morl	Morlet
prep	Preprocessed

References

1. R. D. Thijs, R. Surges, T. J. O'Brien and J. W. Sander, Epilepsy in adults, *The Lancet* (2019).
2. J. Pillai and M. R. Sperling, Interictal EEG and the diagnosis of epilepsy, *Epilepsia* **47** (2006) 14–22.
3. B. Frauscher and J. Gotman, Sleep, oscillations, interictal discharges, and seizures in human focal epilepsy, *Neurobiol. Dis.* **127** (2019) 545–553.
4. J. J. Halford, Computerized epileptiform transient detection in the scalp electroencephalogram: Obstacles to progress and the example of computerized ecg interpretation, *Clin. Neurophysiol.* **120**(11) (2009) 1909–1915.
5. N. B. Fountain and J. M. Freeman, EEG is an essential clinical tool: Pro and con, *Epilepsia* **47** (2006) 23–25.
6. J. Thomas, J. Jin, P. Thangavel, E. Bagheri, R. Yuvaraj, J. Dauwels, R. Rathakrishnan, J. J. Halford, S. S. Cash and B. Westover, Automated detection of interictal epileptiform discharges from scalp electroencephalograms by convolutional neural networks, *Int. J. Neural Syst.* **30**(11) (2020) 2050030.
7. T. Prasanth, J. Thomas, R. Yuvaraj, J. Jing, S. S. Cash, R. Chaudhari, T. Y. Leng, R. Rathakrishnan, S. Rohit, V. Saini *et al.*, Deep learning for interictal epileptiform spike detection from scalp EEG frequency sub bands, in *2020 42nd Annual Int. Conf. IEEE Engineering in Medicine & Biology Society* (*EMBC*) (IEEE, 2020), pp. 3703–3706.
8. J. Jing, H. Sun, J. A. Kim, A. Herlopian, I. Karakis, M. Ng, J. J. Halford, D. Maus, F. Chan, M. Dolatshahi *et al.*, Development of expert-level automated detection of epileptiform discharges during electroencephalogram interpretation, *JAMA Neurol.* **77**(1) (2019) 103–108.
9. J. Thomas, L. Comoretto, J. Jin, J. Dauwels, S. S. Cash and M. B. Westover, EEG classification via convolutional neural network-based interictal epileptiform event detection, in *2018 40th Annual Int. Conf. IEEE Engineering in Medicine and Biology Society* (*EMBC*) (IEEE, 2018), pp. 3148–3151.
10. J. Thomas, P. Thangavel, W. Y. Peh, J. Jing, R. Yuvaraj, S. S. Cash, R. Chaudhari, S. Karia, R. Rathakrishnan, V. Saini *et al.*, Automated adult epilepsy diagnostic tool based on interictal scalp electroencephalogram characteristics: A six-center study, *Int. J. Neural Syst.* **31**(5) (2021) 2050074.
11. I. Ullah, M. Hussain, H. Aboalsamh *et al.*, An automated system for epilepsy detection using EEG brain signals based on deep learning approach, *Expert Syst. Appl.* **107** (2018) 61–71.
12. Y. Roy, H. Banville, I. Albuquerque, A. Gramfort, T. H. Falk and J. Faubert, Deep learning-based electroencephalography analysis: A systematic review, *J. Neural Eng.* **16**(5) (2019) 051001.
13. T. X. Le, T. T. Le, V. V. Dinh, Q. L. Tran, L. T. Nguyen and D. T. Nguyen, Deep learning for epileptic spike detection, *VNU J. Sci. Comput. Sci. Commun. Eng.* **33**(2) (2017) 1–13.
14. K. Fukumori, H. T. T. Nguyen, N. Yoshida and T. Tanaka, Fully data-driven convolutional filters with deep learning models for epileptic spike detection, in *ICASSP 2019-2019 IEEE Int. Conf. Acoustics, Speech and Signal Processing* (*ICASSP*) (IEEE, 2019), pp. 2772–2776.
15. K. Indiradevi, E. Elias, P. Sathidevi, S. D. Nayak and K. Radhakrishnan, A multi-level wavelet approach for automatic detection of epileptic spikes in the electroencephalogram, *Comput. Biol. Med.* **38**(7) (2008) 805–816.

16. D. Gajic, Z. Djurovic, J. Gligorijevic, S. Di Gennaro and I. Savic-Gajic, Detection of epileptiform activity in EEG signals based on time-frequency and non-linear analysis, *Front. Comput. Neurosci.* **9** (2015) 38.
17. J. Le Douget, A. Fouad, M. M. Filali, J. Pyrzowski and M. Le Van Quyen, Surface and intracranial EEG spike detection based on discrete wavelet decomposition and random forest classification, in *Engineering in Medicine and Biology Society* (*EMBC*), *2017 39th Annual Int. Conf. IEEE* (IEEE, 2017), pp. 475–478.
18. Ö. Türk and M. S. Özerdem, Epilepsy detection by using scalogram based convolutional neural network from EEG signals, *Brain Sci.* **9**(5) (2019) 115.
19. N. T. Anh-Dao, N. Linh-Trung, L. Van Nguyen, T. Tran-Duc, B. Boashash *et al.*, A multistage system for automatic detection of epileptic spikes, *REV J. Electron. Commun.* **8**(1–2) (2018).
20. M. Latka, Z. Was, A. Kozik and B. J. West, Wavelet analysis of epileptic spikes, *Phys. Rev. E* **67**(5) (2003) 052902.
21. Y. Bengio, F. Bastien, A. Bergeron, N. Boulanger-Lewandowski, T. Breuel, Y. Chherawala, M. Cisse, M. Côté, D. Erhan, J. Eustache *et al.*, Deep learners benefit more from out-of-distribution examples, in *Proc. Fourteenth Int. Conf. Artificial Intelligence and Statistics* (2011), pp. 164–172.
22. I. Goodfellow, Y. Bengio, A. Courville and Y. Bengio, *Deep Learning* (MIT Press, Cambridge, 2016).
23. M. E. Akbiyik, Data augmentation in training CNNs: Injecting noise to images (2020).
24. S. Roy, I. Kiral-Kornek and S. Harrer, Chrononet: A deep recurrent neural network for abnormal EEG identification, in *Conf. Artificial Intelligence in Medicine in Europe* (Springer, 2019), pp. 47–56.
25. W. Y. Peh, J. Thomas, E. Bagheri, R. Chaudhari, S. Karia, R. Rathakrishnan, V. Saini, N. Shah, R. Srivastava, Y.-L. Tan *et al.*, Multi-center validation study of automated classification of pathological slowing in adult scalp electroencephalograms via frequency features, *Int. J. Neural Syst.* **31**(6) (2021) 2150016.
26. I. Obeid and J. Picone, The temple university hospital EEG data corpus, *Front. Neurosci.* **10** (2016) 196.
27. A. Antoniades, L. Spyrou, D. Martin-Lopez, A. Valentin, G. Alarcon, S. Sanei and C. C. Took, Detection of interictal discharges with convolutional neural networks using discrete ordered multichannel intracranial EEG, *IEEE Trans. Neural Syst. Rehab. Eng.* **25**(12) (2017) 2285–2294.
28. A. Antoniades, L. Spyrou, D. Martin-Lopez, A. Valentin, G. Alarcon, S. Sanei and C. C. Took, Deep neural architectures for mapping scalp to intracranial EEG, *Int. J. Neural Syst.* **28**(8) (2018) 1850009.
29. L. Spyrou, S. Kouchaki and S. Sanei, Multiview classification and dimensionality reduction of scalp and intracranial EEG data through tensor factorisation, *J. Signal Process. Syst.* **90**(2) (2018) 273–284.
30. L.-C. Lin, C.-S. Ouyang, R.-C. Wu, R.-C. Yang and C.-T. Chiang, Alternative diagnosis of epilepsy in children without epileptiform discharges using deep convolutional neural networks, *Int. J. Neural Syst.* **30**(5) (2020) 1850060.
31. M. C. Tjepkema-Cloostermans, R. C. de Carvalho and M. J. van Putten, Deep learning for detection of focal epileptiform discharges from scalp EEG recordings, *Clin. Neurophysiol.* **129**(10) (2018) 2191–2196.
32. A. V. Oppenheim, *Discrete-Time Signal Processing* (Pearson Education India, 1999).
33. H. Adeli, Z. Zhou and N. Dadmehr, Analysis of EEG records in an epileptic patient using wavelet transform, *J. Neurosci. Meth.* **123**(1) (2003) 69–87.
34. M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin *et al.*, Tensorflow: Large-scale machine learning on heterogeneous distributed systems, preprint (2016), arXiv:1603.04467.
35. X. Glorot and Y. Bengio, Understanding the difficulty of training deep feedforward neural networks, in *Proc. Thirteenth Int. Conf. Artificial Intelligence and Statistics* (2010), pp. 249–256.
36. J. Jing, A. Herlopian, I. Karakis, M. Ng, J. J. Halford, A. Lam, D. Maus, F. Chan, M. Dolatshahi, C. F. Muniz *et al.*, Interrater reliability of experts in identifying interictal epileptiform discharges in electroencephalograms, *JAMA Neurol.* **77**(1) (2020) 49–57.
37. A. C. Grant, S. G. Abdel-Baki, J. Weedon, V. Arnedo, G. Chari, E. Koziorynska, C. Lushbough, D. Maus, T. McSween, K. A. Mortati *et al.*, EEG interpretation reliability and interpreter confidence: A large single-center study, *Epilepsy Behav.* **32** (2014) 102–107.
38. J. Z. Ding, R. Mallick, J. Carpentier, K. McBain, N. Gaspard, M. B. Westover and T. A. Fantaneanu, Resident training and interrater agreements using the ACNS critical care EEG terminology, *Seizure* **66** (2019) 76–80.
39. P. Piccinelli, M. Viri, C. Zucca, R. Borgatti, A. Romeo, L. Giordano, U. Balottin and E. Beghi, Inter-rater reliability of the EEG reading in patients with childhood idiopathic epilepsy, *Epilepsy Res.* **66**(1–3) (2005) 195–198.
40. C. Shorten and T. M. Khoshgoftaar, A survey on image data augmentation for deep learning, *J. Big Data* **6**(1) (2019) 60.
41. Y. Jing, Y. Yang, Z. Feng, J. Ye, Y. Yu and M. Song, Neural style transfer: A review, *IEEE Trans. Visual. Comput. Graph.* **26**(11) (2019) 3365–3385.
42. I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville and Y. Bengio, Generative adversarial nets, in *Advances in Neural Information Processing Systems* (2014), pp. 2672–2680.