Clinical Neurophysiology 146 (2023) 10–17

Contents lists available at ScienceDirect

Clinical Neurophysiology

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

A quantitative approach to evaluating interictal epileptiform discharges based on interpretable quantitative criteria

Fábio A. Nascimento [a,b,1], Jaden D. Barfuss [a,1], Alex Jaffe [c,1], M. Brandon Westover [b,2,*], Jin Jing [b,2]

[a] Department of Neurology, Washington University School of Medicine, St. Louis, MO, USA
[b] Department of Neurology, Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA
[c] Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA, USA

HIGHLIGHTS

- We created an automated algorithm that quantifies the six IFCN criteria for interictal epileptiform discharge (IED) identification.
- The model combines these criteria in a user-friendly way that accurately captures the likelihood that a waveform is epileptiform.
- This model may assist electroencephalographers decide whether a waveform is epileptiform, and trainees learn how to identify IEDs.

ARTICLE INFO

Article history:
Accepted 31 October 2022
Available online

Keywords:
EEG
Education
Interictal epileptiform discharges
Epileptiform discharges

ABSTRACT

Objective: To provide quantitative measures of the six International Federation of Clinical Neurophysiology (IFCN) criteria for interictal epileptiform discharge (IED) identification and estimate the likelihood of a candidate IED being epileptiform.
Methods: We designed an algorithm to identify five fiducial landmarks (onset, peak, trough, slow-wave peak, offset) of a candidate IED, and from these to quantify the six IFCN features of IEDs. Another model was trained with these features to quantify the probability that the waveform is epileptiform and incorporated into a user-friendly interface.
Results: The model's performance is excellent (area under the receiver operating characteristic curve (AUROC) = 0.88; calibration error 0.03) but lower than human experts (receiver operating characteristic (ROC) curve is below experts' operating points) or a deep neural-network model (SpikeNet; AUCROC = 0.97; calibration error 0.04). The six features were all significant ($p<0.001$), but not equally important when determining potential epileptiform nature of candidate IEDs: waveform asymmetry was the most (coefficient 0.64) and duration the least discriminative (coefficient 0.09).
Conclusions: Our approach quantifies the six IFCN criteria for IED identification and combines them in an easily interpretable, accessible fashion that accurately captures the likelihood that a candidate waveform is epileptiform.
Significance: This model may assist clinical electroencephalographers decide whether candidate waveforms are epileptiform and may assist trainees learn to identify IEDs.

1. Introduction

Providing high quality diagnostic testing and care for patients with suspected or established epilepsy requires accurate and reliable electroencephalography (EEG) interpretation. Correctly classifying sharp transients on EEG as epileptiform or non-epileptiform carries high weight in the clinical decision-making process in many scenarios (Fisher et al., 2014). Both over- and under-calling interictal epileptiform discharges (IEDs) can lead to epilepsy misdiagnosis and resultant harm to patients and healthcare systems (Benbadis,

Abbreviations: IFCN, International Federation of Clinical Neurophysiology; IED, interictal epileptiform discharge; AUROC, area under the receiver operating characteristic curve; ROC, receiver operating characteristic curve; EEG, electroencephalography; DTW, dynamic-time warping; GUI, graphic user interface; LOFO, leave one feature out; BG, background; LRM, logistic regression model.

* Corresponding author at: 50 Staniford St. Suite 401, Boston, MA 02114, USA.
E-mail address: mwestover@mgh.harvard.edu (M. Brandon Westover).
[1] Co-first authors.
[2] Co-senior authors.

https://doi.org/10.1016/j.clinph.2022.10.018
1388-2457/

2007). Notably, EEG misinterpretation is an issue not only when EEGs are read by neurologists without EEG training (Amin and Benbadis, 2019) – a common practice in the US (Adornato et al., 2011) – but also when studies are read by experts. The latter have been shown to have imperfect interrater reliability in identifying IEDs on routine EEG (Jing et al., 2020a; Nascimento et al., 2022).

In an effort to improve reliability in identifying IEDs, the International Federation of Clinical Neurophysiology (IFCN), in 2017, proposed a new operational definition of IEDs consisting of six criteria (Kane et al., 2017) (Fig. 1). By contrast with the older approach to IED identification consisting of a single step of recognition based on experience (Noachtar et al., 1999), the new IFCN criteria break down the evaluation of a candidate IED into a series of 6 binary decisions about the presence or absence of elementary features, and recommends that a waveform be considered an IED if 4 or more of the 6 features are present. Preliminary evidence suggests that this simple approach is useful for clinical practice and by electroencephalographers (Kural et al., 2020a).

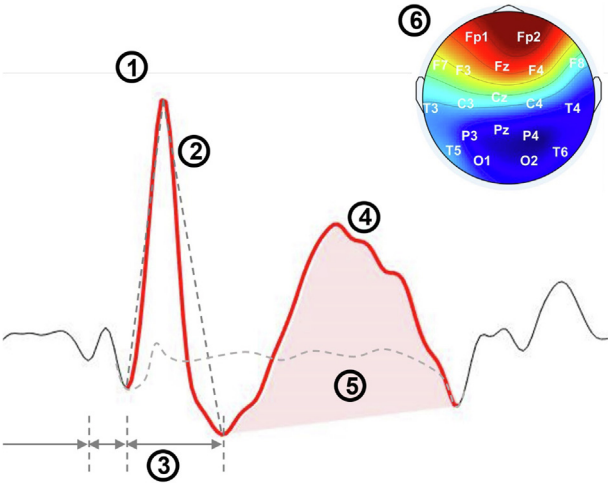

Fig. 1. The six International Federation of Clinical Neurophysiology (IFCN) criteria for identifying an interictal epileptiform discharges (IED). Epileptiform patterns have to fulfill at least four of these criteria: (1) the waveform is spiky; (2) the waveform is asymmetric; (3) the waveform duration is different than the background activity; (4) the waveform has an after-going slow wave; (5) the waveform disrupts the background activity; and (6) the waveform has a physiological field that suggests origination from the brain.

Nevertheless, the definitions given for the six IFCN features are qualitative, and thus remain reliant on expert judgement. To overcome this limitation, in this study we developed ways to quantify each of the six IFCN criteria and combine them into an interpretable algorithm that estimates the likelihood of candidate IEDs being epileptiform. We demonstrate that this model agrees well (calibration error 0.03, area under the receiver operating characteristic curve (AUROC) = 0.88, with 95 % confidence interval (CI) [0.86–0.89]) with expert consensus. Further, the model performs comparably with a better-than-human-experts black-box deep-neural network IED detector (*SpikeNet*, calibration error 0.04, AUROC = 0.97). We also provide an open-source MATLAB-based graphical user interface (Fig. 2) to allow reproducibility of our results, to assist electroencephalographers in identifying IEDs, and to serve as an educational resource to trainees.

2. Methods

2.1. Gold standard

In previous work (Jing et al., 2020b), we collected 13,262 candidate IEDs from a total of 1,051 EEG studies (51.4 % females0) (Table 1). All candidate IEDs were independently rated as IED or non-IED by eight experts (median experience in reading EEGs of 9.5 years, range 4 to 16 years) in standard visual analysis. This data generated a gold standard that consisted of two measures. The first is a binary measure whether a given sharp transient is an IED or a non-IED, based on expert consensus. By this definition a candidate

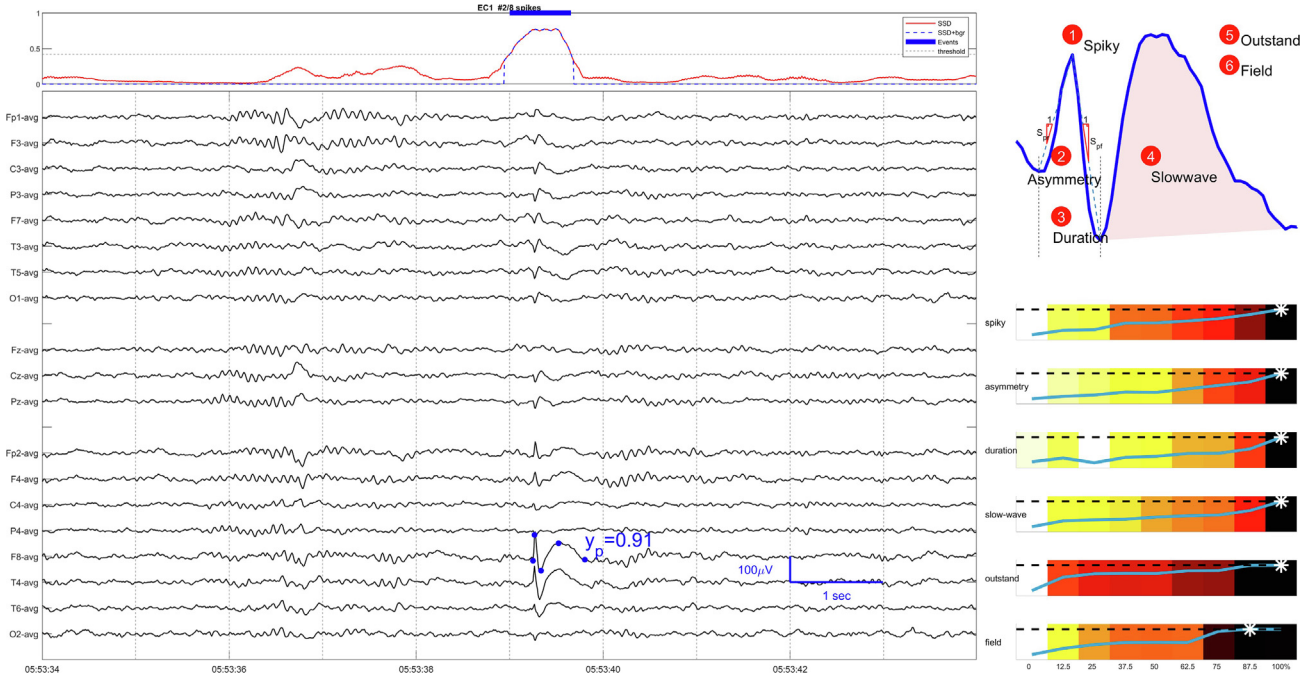

Fig. 2. MATLAB-based graphic user interface (GUI) for assessment of a candidate interictal epileptiform discharge (IED). It implements the six-feature logistic regression model. Operators can select the waveform on continuous electroencephalography (EEG) to get a model prediction score y_p representing the IED likelihood, as well as the individual measures for the six International Federation of Clinical Neurophysiology (IFCN) features in their own heat maps determined from our training data.

Table 1
Demographics of electroencephalography (EEG) samples.

	No.				
Age	**Total**	**%Female**	**rEEG**	**%EMU**	**%ICU**
0–1	30	53	100	0	0
1–5	84	49	95	4	1
5–13	267	49	98	1	1
13–18	117	48	97	3	0
18–30	121	60	94	4	2
30–50	107	50	93	4	3
50–65	133	44	92	2	6
65–75	98	63	91	0	9
75+	94	52	90	2	8
Total	**1051**	**51**	**95**	**2**	**3**

rEEG, routine electroencephalography; EMU, epilepsy monitoring unit; ICU, intensive care unit.

waveform is considered an IED if four or more experts scored it as being epileptiform. The second measure is the probability that a waveform is an IED, calculated as the percentage of experts who classified it as epileptiform (n/8).

2.2. Model design

2.2.1. Templates with fiducial points

Three of the authors (FN, MBW, JJ; experience in reading EEG of 1–10 years) manually marked five fiducial points (onset, peak, trough, slow-wave peak, and offset; Fig. 3) of waveforms in a subset of 2,511 candidate IEDs (evenly sampled from each group with vote 0 to 8). These annotations were used to design and train a model that automatically identifies fiducial points in previously unseen candidate IEDs. Given a one-second, 19-channel, average montage EEG epoch, the algorithm first identifies the five fiducial points in each of the 19 channels. The algorithm secondly utilizes a scoring metric to distinguish which of the 19 sets of five fiducial points is most likely to be the origin of the candidate IED in question. We describe below each of these steps in detail.

2.2.2. Identification of candidate fiducial points: Dynamic time warping

To identify the five fiducial points, dynamic-time warping (DTW) is used to match a series of template IEDs that have already had their fiducial points annotated by experts. Templates were selected from among waveforms that were scored as IEDs (based on receiving at least 4/8 expert votes). DTW attempts to match a template waveform to a candidate IED waveform by nonlinearly shrinking and stretching (“warping”) different parts of the time axis, subject to a limit (the *Sakoe-Chiba* parameter) on the maximum amount of warping permitted; we empirically set this limit to 5 % of the template length (Berndt and Clifford, 1994). As shown in the example in Fig. 4**A**, by aligning points in the candidate IED with fiducial points in the template, the warping process identifies potential fiducial points in the candidate IED. To decide which template best matches a given candidate IED, and in which EEG channel, we run DTW with each of the available templates and select the template $\hat{t}$ yielding the smallest *Euclidean* distance Δ_k, expressed by:

$$\Delta_k = \sum_i |s_i - t_k| \tag{1.1}$$

$$\hat{t} = \min_k [\Delta_k(t_k)], k = 1, 2, 3, \cdots K \tag{1.2}$$

where s represents the candidate IED waveform, t represents an IED template, i indexes the matched points on the candidate IED, j indexes matched points on the template IED, and k indexes the templates.

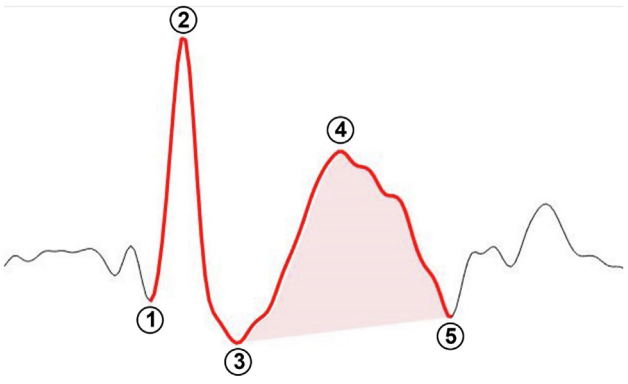

Fig. 3. The five fiducial points of a candidate interictal epileptiform discharge (IED): (1) onset; (2) peak; (3) trough; (4) slow-wave peak; and (5) offset.

2.2.3. Refinement of fiducial points: Secondary peak detection

At this stage, the candidate and template IEDs are aligned, but there remains the possibility that a fiducial point on the template is matched to more than one point on the event. For example, in Fig. 4**A**, the fifth fiducial point in the template is matched to seven adjacent points in the event. Considering that fiducial points 1, 3, and 5 are always troughs and 2 and 4 are always peaks, the minima of the matched event points to template fiducial points 1, 3, and 5 are chosen while the maxima of the matched event points to template fiducial points 2 and 4 are chosen.

Sometimes template matching fails to find five fiducial points in the event signal, such as when multiple fiducial points on the template are mapped to the same fiducial point on the event. Whenever the template matching does not return five distinct points in the event signal to match with the five distinct points in the template signal, the algorithm reports an error has occurred. For instance, in Fig. 4**A**, if the third and fourth fiducial points in the template both matched to the fourth fiducial point in the event, yielding only four distinct fiducial points in the event signal, the algorithm reports an error has occurred. In this case, a secondary peak detection method is used. This peak detection method includes two steps: finding the IED points (fiducial points 1–3); then finding the slow-wave points (fiducial points 4, 5). To locate potential IED points, troughs are found on either side of the peaks found in the candidate IED by looking for sign changes in the first derivative before and after the peak. Each set of points is scored by S_1 according to the following metric:

$$S_1 = \frac{a}{T} + b\left(P^+ + P^-\right) \tag{1.3}$$

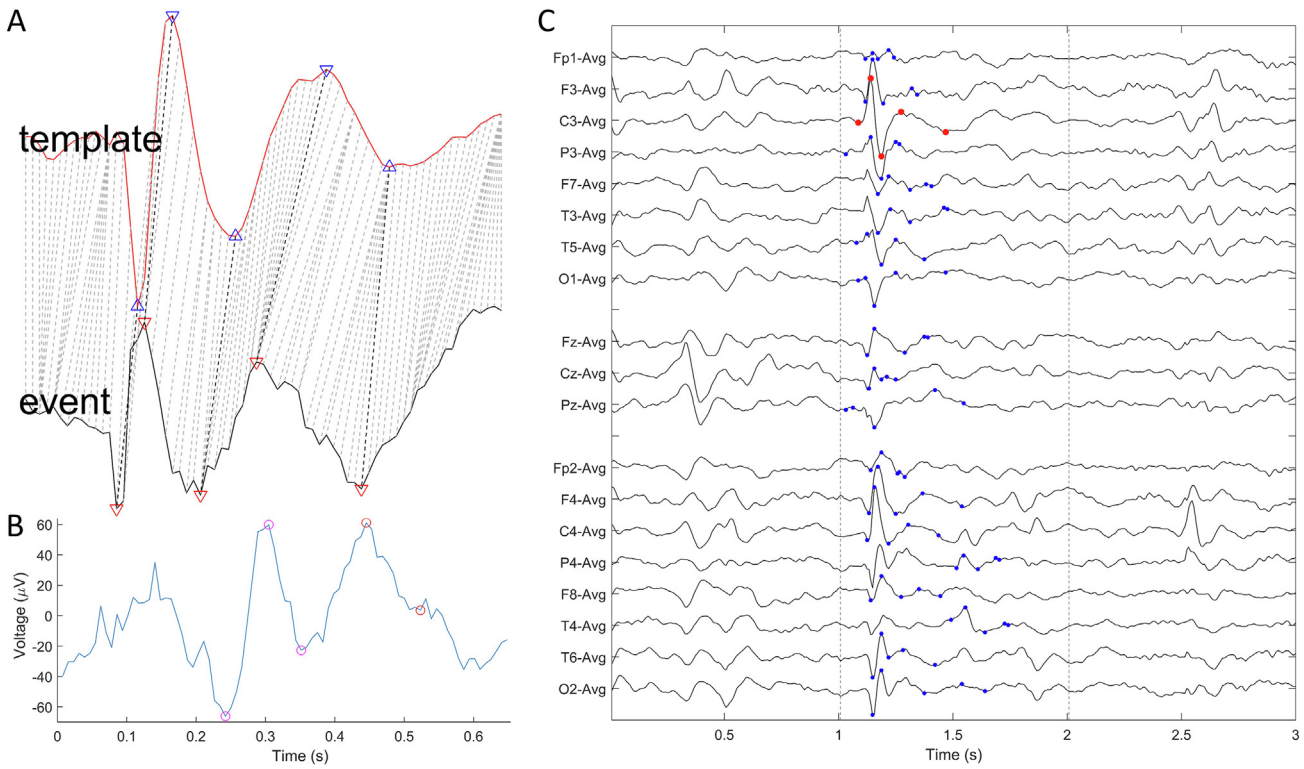

Fig. 4. Method for automated identification of the five fiducial points of a candidate interictal epileptiform discharge (IED). (A) Example of template matching using dynamic time warping (DTW) to find fiducial points. The template in red has its points mapped to the event in black with gray dashed lines. The template has labeled fiducial points marked by blue triangles which are connected via black dashed lines to the detected fiducial points of the event. (B) An example of the secondary peak detection of fiducial points in the event of a template matching error. The IED points are marked in magenta while the slow wave points are marked in red. (C) A candidate IED with overlaid detected fiducial points. The points with the highest amplitude are marked in red while others are marked in blue.

where T is the time between the two troughs on either side of the peak, P^+ and P^- are the voltage differences between the peak and the succeeding trough and between the peak and the preceding trough, respectively, and a and b are empirically determined constants designed to keep the two components of the equation the same order of magnitude. The set of IED points with the highest score is then chosen.

Next, the final two fiducial points corresponding to the slow wave are found. After passing the portion of the signal after the IED through a 5-point moving average filter, the derivative of the smoothed signal is obtained. The fourth fiducial point is chosen to be where the derivative of the signal transitions from positive to negative, as this corresponds to a peak. The fifth is chosen to occur when the derivative after the fourth point transitions from negative back to positive, as this corresponds to a trough. We observe an example of a signal with the fiducial points selected via this secondary method in Fig. 4**B**.

2.2.4. Identification of EEG channel best matching the cranial origin

The above methods for identifying candidate fiducial points are applied to each of the 19 channels, resulting in a list of fiducial points $\boldsymbol{p} = (p_1, p_2, p_3, p_4, p_5)$ for each channel. The fiducial points are scored by S_2 by:

$$S_2 = range(\boldsymbol{p}) \tag{1.4}$$

which is the difference between the maximum and the minimum voltage of the fiducial points $\boldsymbol{p}$. The channel whose points have the highest score is selected as indicating the channel of cranial origin of the candidate IED. Fig. 4**C** shows an example of a 19-channel averaged referenced IED candidate with the detected fiducial points overlaid. Those marked in red indicate the selected channel.

2.2.5. Quantification of the six IFCN features

Once the five fiducial points are identified, the six IFCN features are measured using the following methods (Fig. 5):

Spikiness:

Spikiness was quantified using the angle in radians at the peak of the IED (Fig. 5**A**). The angle was formed by using the X (seconds), Y (μV) coordinates at three points including the peak (point B) and 1/128 seconds before (point A) and after point C) the peak.

$$Spikiness = \arctan((m1 - m2)/(1 + m1 * m2)), \tag{1.5}$$

where $m1$ is the line length between point A and B, and $m2$ the line length between point B and C.

Asymmetry:

Asymmetry of the upward and downward slopes of the candidate IED was measured by calculating the difference between the slope of the rising half-wave and the slope of the falling half-wave (Fig. 5**B**):

$$Asymmetry = |s1 - s2|, \tag{1.6}$$

where $s1$ is the rising slope and $s2$ the falling slope, in microvolts per second. The rising slope $s1$ was taken from a line through the onset and peak of the IED. The falling slope $s2$ was taken from a line drawn through the peak and trough of the IED.

Difference in duration:

The difference in wave duration from typical surrounding background waves was measured by the ratio between the duration of the IED and the average duration of the background waveforms (Fig. 5**C**):

$$Duration = D_{ED} / D_{BG}, \tag{1.7}$$

where D_{BG} is the estimated average background event duration, and D_{ED} is the peak duration of IED, in milliseconds. This was accom-

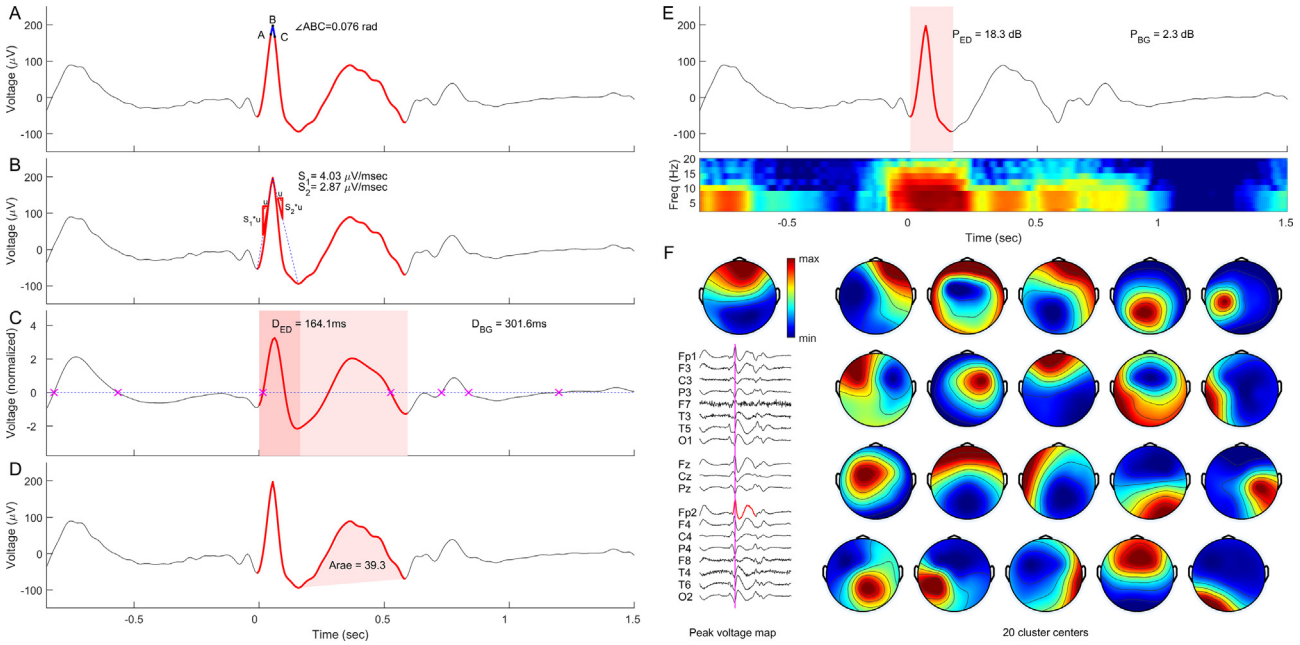

Fig. 5. The six quantitative measures of a candidate interictal epileptiform discharge (IED) based on the six International Federation of Clinical Neurophysiology (IFCN) criteria for IED identification. (A) *Spikiness*: the angle in radians between the peak and 1/128 sec before and after the peak. (B) *Asymmetry*: the absolute value of the slope of the rising half-wave minus the slope of the falling half-wave. (C) *Duration*: the duration of the IED divided by the average duration of the background events. (D) *Slow wave*: the area under the curve between the IED trough and the offset points. (E) *Disruption*: the power of the IED minus the average power of the background. (F) *Field*: the minimum *Euclidean* distance from the peak voltage vector in average montage to a pre-defined collection of IED exemplars that suggest origination from the brain.

plished by first smoothing the signal using boxcar averaging thereby enhancing the signal-to-noise ratio by replacing a group of data points with an average. The EEG data was then z-normalized to remove offset and center around 0 baseline. All zero crossings (i.e., points that signal intersects the 0 baseline) were identified in a local 3-second window centered at the IED event. The average time between each zero-crossing point D_{BG}, excluding the potential IED, was used to represent the duration of the background. The duration of the IED D_{ED} was taken by measuring the time between the IED onset and trough points.

Presence of slow wave:

The slow wave was measured by its area in microvolt*second (Fig. 5**D**):

$$Slow\,wave = \int_{x=p3}^{p5} IED(x), \tag{1.8}$$

where $IED(x)$ is the candidate IED event as function of time x, $p3$ the IED trough point, and $p5$ the IED offset point. The slow wave curve starts at the trough and ends at the offset point of the waveform. The area was represented by the space between the slow wave curve and the line that connected the trough and offset points.

Disruption of background:

The disruption of background activity was measured by taking the difference in signal power of the IED and that of the background (Fig. 5**E**):

$$\text{Disruption} = P_{BG} - P_{ED}, \tag{1.9}$$

where P_{BG} is the power of background, and P_{ED} the power of IED event. Power was calculated in dB scale using spectrogram by first summing up among all spectral frequencies, and then normalized by its duration of interest. The power of the background activity P_{BG} was measured as the mean power from a local 3-second window centered at, and excluding the IED. The power of the IED P_{ED} was measured by taking the mean power from IED onset to trough points.

Physiological field:

The feature that the physiological field of the IED originates from the brain was quantified by calculating the minimum *Euclidean* difference Δ_k between the actual field S_i and top K = 20 exemplar fields T_k (Fig. 5**F**):

$$\Delta_k = \sum_i |S_i - T_k|, \tag{1.10}$$

$$Field = \min_k [\Delta_k(T_k)],\ k = 1, 2, 3, \cdots K. \tag{1.11}$$

The physiological field was quantified using the peak voltage vector of IED with all its channels on average referential montage. These 20 exemplar fields were identified by clustering the generated fields of all the IEDs (among 297 IEDs) that experts unanimously agreed were IEDs (6/8 to 8/8 IEDs). The difference could then be calculated by comparing the peak voltage vector of the IED of interest to each of the 20 exemplars. The minimum *Euclidean* distance identified was then selected to quantify how close that the physiological field of the IED originates from the brain to an existing pattern.

2.2.6. Logistic regression model and statistical analysis

The six quantitative measures were then combined to develop logistic regression models using soft labels with 10-fold cross validation. Each time 9 folds of data were used to train the model, leaving the 10th fold as test data. We tested our models in 2,511 three-second 19-channel EEG epochs on average montage wherein candidate IEDs were annotated by experts in the middle second in the chosen channel. Performance metrics on test data including AUROC and calibration error relative to the gold standard (i.e., the average difference between the predicted probability and the "observed probability", as measured by the percentage of experts who agree that a given waveform is an IED) were reported. The measures were further evaluated by comparing the performance of our logistic regression model against the deep neural network model *SpikeNet* that has previously demonstrated better-than-

expert performance in discriminating IED from non-IEDs in a large and diverse set of labeled examples (Jing et al., 2020b).

2.2.7. User interface

A MATLAB-based graphic user interface (GUI) was created to display IED with context EEG (Fig. 2). One can annotate any event in the continuous EEG, and get instantly the model prediction plus the six IFCN measures shown in their own domain determined by our training data.

3. Results

3.1. Performance of fiducial point identification algorithm

The primary template matching method (DTW) successfully found five fiducial points in 63 % of the channels for all IEDs. The secondary peak detection method was used in the remaining 37 % of cases. The accuracy of our algorithm in terms of its channel selector is presented in an overall fashion and in relationship to the 1,181 IEDs whose points were determined to be epileptiform by over half of experts (Table 2). A channel is marked correct if it is exactly the channel chosen by experts or if it is within neighboring distance in the 10–20 EEG electrode placement scheme. Given these two scenarios of channel accuracy, fiducial point accuracy is determined for a channel that is marked correct. Fiducial points are marked correct if either the sum of absolute differences between the expert points and the selected points is less than 75 milliseconds or the sum of absolute differences between all expert fiducial points and all selected fiducial points is less than 200 milliseconds. Some fiducial points, corresponding to expert points deemed to be epileptiform by over half of experts, initially marked incorrect by our automated accuracy metric, were reviewed by an expert and retroactively deemed correct.

3.2. Performance of channel selection algorithm

Channel selection accuracy and fiducial point findings accuracy are higher when experts are more confident they have marked an IED. Furthermore, as the channel selection accuracy qualification becomes more lenient, fiducial point selection accuracy for those channels deemed correct decreases. This data is summarized in Table 2.

3.3. Performance of IED probability estimation

In comparison to expert consensus, our model had excellent performance (AUROC = 0.88, with 95 % CI [0.86–0.89]) and calibration (calibration error 0.03). The interpretable IFCN feature-based model performed marginally below the deep neural network *SpikeNet* (AUROC = 0.97, calibration error 0.04), the latter being a state-of-the-art deep neural network, and below experts (their operating points are above the interpretable model's ROC curve) (Fig. 6).

As shown in Table 3 multivariate analysis, all six IFCN features are significant in discrimination of IEDs from non-IEDs (p<0.001 for each feature). Waveform asymmetry was the most important feature (coefficient of 0.64), while duration relative to background waveforms was least important (coefficient of 0.09). The remaining features had coefficients ranging from 0.20 to 0.41 (Table 3).

Table 2
Fiducial point finder and channel selector accuracy.

Metric	N = 2511, 3-sec, average, all votes	N = 1181, 3-sec, average, votes > 4
Expert channels chosen	44.8 %	49.9 %
Expert /neighbor channels chosen	73.4 %	80.9 %
Fiducial points correct given expert channels chosen	68.2 %	72.0 %
Fiducial points correct given expert/ neighbor channels chosen	58.2 %	64.8 %
Adjusted fiducial points correct given expert channels chosen	–	91.9 %
Adjusted fiducial points correct given expert/neighbor channels chosen	–	77.3 %

4. Discussion

Our model analyzes candidate IEDs as to their probability of being epileptiform based on the operational criteria for IED identification proposed by the IFCN (Kane et al., 2017). This new method to evaluate IEDs, which is based upon quantification of each of the six IFCN criteria (Kane et al., 2017), has excellent performance (calibration error 0.03, AUROC = 0.88, with 95 % CI [0.86–0.89]) relative to expert consensus. Although its performance lies below the *SpikeNet* (Jing et al., 2020b) performance (calibration error 0.04, AUROC = 0.97), an artificial intelligence (AI)-based deep neural network algorithm, it surpasses the *SpikeNet* (Jing et al., 2020b) in interpretability making it a better companion tool for EEG readers in the evaluation of IEDs.

It is expected that there should be a tradeoff between performance and interpretability: models that perform better (i.e., that can make more correct predictions) are typically more complex and less interpretable (i.e., allow less human understanding), and vice versa. Our model compromises its performance – as compared to deep neural network algorithms – for increased understandability. The major advantage of this tradeoff is that the latter creates continuity in IED classification between interpreters and the interpretation software. Nonetheless, it is likely possible that our interpretable model could be further improved by optimizing one or more of the component algorithms (e.g., template matching) or by further fine-tuning parameters without compromising interpretability, and this is an ongoing focus of our research.

Human readers applying the six IFCN operational criteria to identify IEDs (Kane et al., 2017) were able to achieve diagnostic performance similar to conventional expert scoring (i.e., scoring without explicit reference to IFCN criteria) in a recent clinical validation study (Kural et al., 2020a). One of the main advantages associated with adoption of IFCN criteria relates to the decomposition of the complex implicit judgement involved in IED identification into a series of simpler tasks (Nascimento et al., 2022). Despite its success in standardizing the approach to identify IEDs, utilizing the IFCN criteria still requires subjective judgement. This subjectivity is reflected in moderate levels of interrater variability for some criteria - namely features three (waveform asymmetry) and five (background disruption) (Kural et al., 2020a). By quantifying all six IFCN criteria, our model addresses this issue thereby amplifying the objectivity linked with the use of these criteria.

Quantifying the IFCN criteria in our model also allows us to better understand the relative importance of each individual criterion in determining the likelihood that a sharp transient is epileptiform. The potential different importances of different combinations of IFCN criteria in accurately identifying IEDs have been investigated previously. In a recent study utilizing video-EEG data as the diagnostic gold standard (Kural et al., 2020b), the best performing set of criteria was the combination of spiky morphology, aftergoing slow-wave, and physiological field on voltage map. This combination produced high specificity (97 %) while maintaining reasonable sensitivity (89 %) as well as substantial interrater agreement (0.66). It is unclear why some criteria – in combination or isolation - result in increased diagnostic performance because they are more easily recognized by experts (i.e., improved interrater agreement) and/or due to their inherent level of association with IEDs.

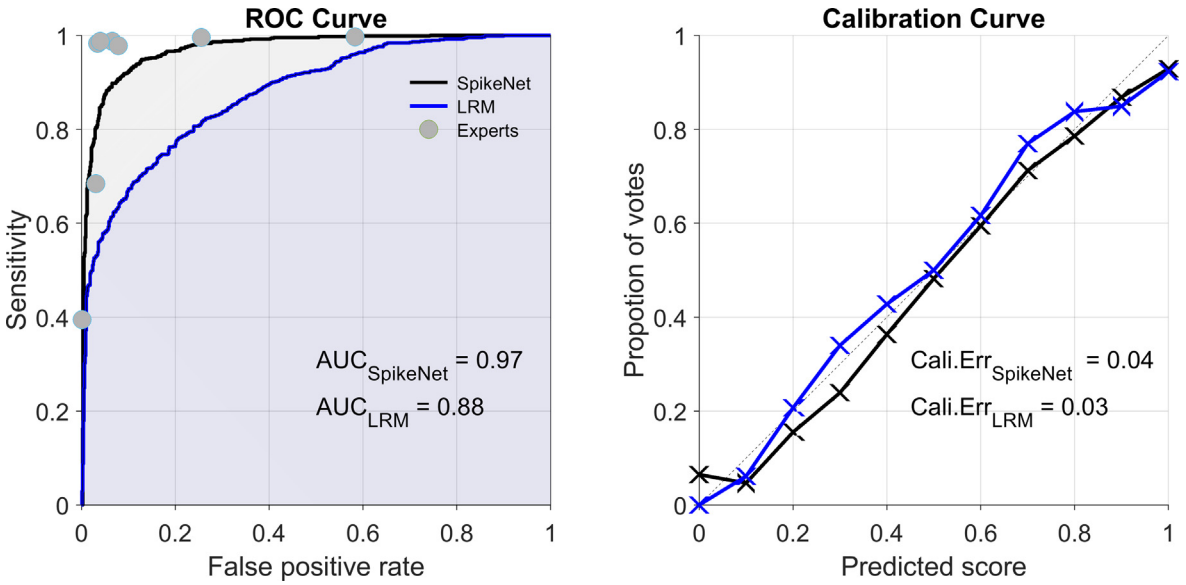

Fig. 6. Model's accuracy and calibration based on expert consensus and compared with *SpikeNet*. Our model preforms with a lower accuracy than *SpikeNet* with an area under the receiver operating characteristic curve (AUROC) at 0.88 but a better calibration error at 0.03. ROC, receiver operating characteristic; LRM, logistic regression model.

Table 3
Performance of the six International Federation of Clinical Neurophysiology (IFCN) quantitative features of a candidate interictal epileptiform discharge (IED).

Feature	Coefficient	OR	95 % CI	p-value	LOFO AUROC
Spikiness	0.26	1.30	[1.26,1.34]	<0.001	0.795
Asymmetry	0.64	1.90	[1.77,2.04]	<0.001	0.772
Duration	0.09	1.09	[1.06,1.13]	<0.001	0.805
Slow wave	0.20	1.23	[1.18,1.28]	<0.001	0.799
Disruption of BG	0.41	1.51	[1.46,1.56]	<0.001	0.795
Physiological Field	0.27	1.31	[1.27,1.35]	<0.001	0.784

OR: odds ratio; CI: confidence interval; LOFO: leave one feature out; AUROC: area under the receiver operating characteristic curve; BG, background.

From a practical standpoint, we envision our model being used in clinical practice in two ways. It may be utilized in combination with conventional visual EEG analysis. Experts would select a candidate IED, run our model, and use the algorithm-generated probability (overall and criterion-specific) that this waveform is epileptiform in their decision-making. Alternatively, it may be used after the EEG data is processed by AI-based algorithms, such as *SpikeNet* (Jing et al., 2020b). Experts would visually inspect automated detections of candidate IED clusters, run our model, and ultimately decide whether these waveforms are epileptiform. This approach where experts work in conjunction with AI-based algorithms (termed "hybrid") has been shown to be accurate and suitable for clinical implementation (Kural et al., 2022a).

From an educational standpoint, our model holds potential to be instrumental in teaching trainees how to identify IEDs specifically through implementation of the six IFCN criteria. This task is of paramount importance given the deleterious effects of EEG misinterpretation and epilepsy diagnosis (Benbadis, 2007) and the well-known profound lack of consistency in trainee EEG teaching in the US (Nascimento and Gavvala, 2021). Because of its user-friendly interface and easy interpretability, trainees may utilize our model with minimal prior training. Further, it provides trainees with the probability assigned with each IFCN criterion for a particular candidate IED. We believe this instant feedback has major educational benefits especially when trainees are initially exposed to the six IFCN criteria. Notably, teaching trainees how to apply the IFCN criteria on candidate IEDs has been shown to improve diagnostic performance as well as interrater agreement (Kural et al., 2022b).

Our work has important limitations. First, the dataset used to train our model is relatively small in scale. Second, our model was not explicitly evaluated for its ability to identify difficult "corner-cases", such as differentiating benign waveforms that closely resemble IEDs (e.g., vertex sharp waves in children). Third, our model is designed to evaluate candidate IEDs, meaning waveforms selected by an expert (or by another algorithm) as potentially being IEDs. The algorithm is not designed to be fast enough to scan continuous EEG data, nor has it been tested in that context. Fourth, we highlight that there may be other methods to assess the IED criterion related to disruption of the background, some of which may be different from the signal power-based approach used in our model. Lastly, the method by which our model evaluates whether a candidate IED has a physiological field is different from the method performed by electroencephalographers, which is based on visual assessment of the waveform's voltage topography and determination of whether it is compatible with cerebral generation.

5. Conclusion

Our model accurately quantifies the six IFCN criteria in an easily accessible and readily interpretable manner. We believe its use can have major benefits in clinical practice and education. In clinical practice, the model may help electroencephalographers rate candidate IEDs and potentially improve interrater agreement. In education, our model has the potential to help trainees learn and apply the six IFCN criteria. The improvement in accuracy and reliability

of IED identification will ultimately positively impact the diagnosis and care of patients with seizures and epilepsy.

Disclosure

F. Nascimento, J. Barfuss, A. Jaffe, and J. Jing report no disclosures relevant to the manuscript. M. Westover is a co-founder of Beacon Biosignals, which played no role in this work.

Author contributions

F. Nascimento, J. Barfuss, and A. Jaffe collected and analyzed the data, and drafted the manuscript. M. Westover and J. Jing conceptualized and designed the study, collected and analyzed the data, revised the manuscript, supervised the study, and provided final approval of the version being submitted. All authors consent to take responsibility for the content of the work. All authors confirm that this manuscript does not overlap with previous publications and the manuscript, data, figures, and tables have not been published previously.

Funding source

MBW was supported by the Glenn Foundation for Medical Research and American Federation for Aging Research (Breakthroughs in Gerontology Grant); American Academy of Sleep Medicine (AASM Foundation Strategic Research Award); NIH (R01NS102190, R01NS102574, R01NS107291, RF1AG064312, R01AG062989, R01AG073410); and NSF (SCH-2014431).

References

Amin U, Benbadis SR. The role of EEG in the erroneous diagnosis of epilepsy. J Clin Neurophysiol 2019;36(4):294–7.

Adornato BT, Drogan O, Thoresen P, Coleman M, Henderson VW, Henry KA, Liu L, Mortimer JA, Schneck MJ, Borenstein AR. The practice of neurology, 2000–2010: report of the AAN member research subcommittee. Neurology 2011;77(21):1921–8.

Benbadis SR. Errors in EEGs and the misdiagnosis of epilepsy: importance, causes, consequences, and proposed remedies. Epilepsy Behav 2007;11(3):257–62.

Berndt DJ, Clifford J. Using dynamic time warping to find patterns in time series. In: KDD Workshop, Seattle, WA, 1994, vol. 10. p 359–70.

Fisher RS, Acevedo C, Arzimanoglou A, Bogacz A, Cross JH, Elger CE, Engel Jr J, Forsgren L, French JA, Glynn M, Hesdorffer DC, Lee BI, Mathern GW, Moshe SL, Perucca E, Scheffer IE, Tomson T, Watanabe M, Wiebe S. ILAE official report: a practical clinical definition of epilepsy. Epilepsia 2014;55(4):475–82.

Jing J, Herlopian A, Karakis I, Ng M, Halford JJ, Lam A, Maus D, Chan F, Dolatshahi M, Muniz CF, Chu C, Sacca V, Pathmanathan J, Ge W, Sun H, Dauwels J, Cole AJ, Hoch DB, Cash SS, Westover MB. Interrater reliability of experts in identifying interictal epileptiform discharges in electroencephalograms. JAMA Neurol 2020a;77(1):49–57.

Jing J, Sun H, Kim JA, Herlopian A, Karakis I, Ng M, Halford JJ, Maus D, Chan F, Dolatshahi M, Muniz C, Chu C, Sacca V, Pathmanathan J, Ge W, Dauwels J, Lam A, Cole AJ, Cash SS, Westover MB. Development of expert-level automated detection of epileptiform discharges during electroencephalogram interpretation. JAMA Neurol 2020b;77(1):103–8.

Kane N, Acharya J, Beniczky S, Caboclo L, Finnigan S, Kaplan PW, Shibasaki H, Pressler R, van Putten MJAM. A revised glossary of terms most commonly used by clinical electroencephalographers and updated proposal for the report format of the EEG findings. Revision 2017. Clin Neurophysiol Pract 2017;2:170–85.

Kural MA, Duez L, Hansen VS, Larsson PG, Rampp S, Schulz R, Tankisi H, Wennberg R, Bibby BM, Scherg M, Beniczky S. Criteria for defining interictal epileptiform discharges in EEG: A clinical validation study. Neurology 2020a;94(20):e2139–47.

Kural MA, Tankisi H, Duez L, Hansen VS, Udupi A, Wennberg R, Rampp S, Larsson PG, Schulz R, Beniczky S. Optimized set of criteria for defining interictal epileptiform EEG discharges. Clin Neurophysiol 2020b;131(9):2250–4.

Kural MA, Jing J, Furbass F, Perko H, Qerama E, Johnsen B, Fuchs S, Westover MB, Beniczky S. Accurate identification of EEG recordings with interictal epileptiform discharges using a hybrid approach: Artificial intelligence supervised by human experts. Epilepsia 2022a;63(5):1064–73.

Kural MA, Aydemir ST, Levent HC, Olmez B, Ozer IS, Vlachou M, Witt AH, Yilmaz AY, Beniczky S. The operational definition of epileptiform discharges significantly improves diagnostic accuracy and inter-rater agreement of trainees in EEG reading. Epileptic Disord 2022b;24(2):353–8.

Nascimento FA, Gavvala JR. Education research: Neurology resident EEG education: a survey of US neurology residency program directors. Neurology 2021;96(17):821–4.

Nascimento FA, Jing J, Beniczky S, Benbadis SR, Gavvala JR, Yacubian EMT, Wiebe S, Rampp S, van Putten MJAM, Tripathi M, Cook MJ, Kaplan PW, Tatum WO, Trinka E, Cole AJ, Westover MB. One EEG, one read – a manifesto towards reducing interrater variability among experts. Clin Neurophysiol 2022;133:68–70.

Noachtar S, Binnie C, Ebersole J, Mauguiere F, Sakamoto A, Westmoreland B. A glossary of terms most commonly used by clinical electroencephalographers and proposal for the report form for the EEG findings. The International Federation of Clinical Neurophysiology. Electroencephalogr Clin Neurophysiol Suppl 1999;52:21–41.