

Development of Expert-Level Automated Detection of Epileptiform Discharges During Electroencephalogram Interpretation

Jin Jing, PhD; Haoqi Sun, PhD; Jennifer A. Kim, MD, PhD; Aline Herlopian, MD; Ioannis Karakis, MD, PhD; Marcus Ng, MD; Jonathan J. Halford, MD; Douglas Maus, MD, PhD; Fonda Chan, MD; Marjan Dolatshahi, MD; Carlos Muniz, MD; Catherine Chu, MD; Valeria Sacca, PhD; Jay Pathmanathan, MD; Wendong Ge, PhD; Justin Dauwels, PhD; Alice Lam, MD, PhD; Andrew J. Cole, MD; Sydney S. Cash, MD, PhD; M. Brandon Westover, MD, PhD

IMPORTANCE Interictal epileptiform discharges (IEDs) in electroencephalograms (EEGs) are a biomarker of epilepsy, seizure risk, and clinical decline. However, there is a scarcity of experts qualified to interpret EEG results. Prior attempts to automate IED detection have been limited by small samples and have not demonstrated expert-level performance. There is a need for a validated automated method to detect IEDs with expert-level reliability.

OBJECTIVE To develop and validate a computer algorithm with the ability to identify IEDs as reliably as experts and classify an EEG recording as containing IEDs vs no IEDs.

DESIGN, SETTING, AND PARTICIPANTS A total of 9571 scalp EEG records with and without IEDs were used to train a deep neural network (SpikeNet) to perform IED detection. Independent training and testing data sets were generated from 13 262 IED candidates, independently annotated by 8 fellowship-trained clinical neurophysiologists, and 8520 EEG records containing no IEDs based on clinical EEG reports. Using the estimated spike probability, a classifier designating the whole EEG recording as positive or negative was also built.

MAIN OUTCOMES AND MEASURES SpikeNet accuracy, sensitivity, and specificity compared with fellowship-trained neurophysiology experts for identifying IEDs and classifying EEGs as positive or negative or negative for IEDs. Statistical performance was assessed via calibration error and area under the receiver operating characteristic curve (AUC). All performance statistics were estimated using 10-fold cross-validation.

RESULTS SpikeNet surpassed both expert interpretation and an industry standard commercial IED detector, based on calibration error (SpikeNet, 0.041; 95% CI, 0.033-0.049; vs industry standard, 0.066; 95% CI, 0.060-0.078; vs experts, mean, 0.183; range, 0.081-0.364) and binary classification performance based on AUC (SpikeNet, 0.980; 95% CI, 0.977-0.984; vs industry standard, 0.882; 95% CI, 0.872-0.893). Whole EEG classification had a mean calibration error of 0.126 (range, 0.109-0.1444) vs experts (mean, 0.197; range, 0.099-0.372) and AUC of 0.847 (95% CI, 0.830-0.865).

CONCLUSIONS AND RELEVANCE In this study, SpikeNet automatically detected IEDs and classified whole EEGs as IED-positive or IED-negative. This may be the first time an algorithm has been shown to exceed expert performance for IED detection in a representative sample of EEGs and may thus be a valuable tool for expedited review of EEGs.

JAMA Neurol. 2020;77(1):103-108. doi:10.1001/jamaneurol.2019.3485
Published online October 21, 2019.

← Related article [page 49](#)

+ Supplemental content

Author Affiliations: Author affiliations are listed at the end of this article.

Corresponding Author: M. Brandon Westover, MD, PhD, Department of Neurology, Massachusetts General Hospital, 55 Fruit St, Boston, MA 02114 (mwestover@mgh.harvard.edu).

Interictal epileptiform discharges (IEDs) in an electroencephalogram (EEG) are a hallmark of epilepsy.¹⁻³ Identification of IEDs by fellowship-trained clinical neurophysiologists is the standard. However, subspecialists are scarce—most EEGs performed in the United States are read by nonspecialists⁴ and, in much of the world, EEG services are unavailable. Interictal epileptiform discharges are also biomarkers for delayed cerebral ischemia in subarachnoid hemorrhage⁵⁻⁷ and risk for posttraumatic epilepsy,⁶ and may contribute to cognitive deterioration in Alzheimer disease.⁸ It is usually impractical to manually annotate IEDs thoroughly, particularly in EEGs recorded over hours or days. There is a need for methods that detect IEDs automatically without compromising accuracy.

Prior automation efforts have been limited by lack of large, well-annotated data sets with which to train and evaluate IED detection algorithms.⁹⁻¹² We therefore created a set of 13 262 candidate IEDs from 1051 EEGs, spanning a range of ages and settings, and recruited 8 physician subspecialists to independently annotate each candidate IED. Using these data and 8520 additional control EEGs reported in the medical record to contain no IEDs, we trained a computer program to identify individual IEDs and classify an entire EEG as containing IEDs vs none, with accuracy exceeding that of typical fellowship-trained clinical neurophysiologists. Thus, a total of 9571 scalp EEG records with and without IEDs were used to develop a deep neural network.

Key Points

Question Can a computer program be trained to identify interictal epileptiform discharges and classify an electroencephalogram as containing interictal epileptiform discharges with accuracy equivalent or superior to that of physicians with subspecialty training in clinical neurophysiology?

Findings In this diagnostic study of interictal epileptiform discharges, a deep neural network was trained using 9571 scalp electroencephalogram recordings. The algorithm appeared to perform at or above the accuracy, sensitivity, and specificity of fellowship-trained clinical experts.

Meaning This computer program appeared to be able to classify electroencephalograms and detect individual interictal epileptiform discharges more accurately than human experts and may help with diagnostic testing for epilepsy and warn of clinical decline in critically ill patients, particularly in settings without available electroencephalogram expertise.

Methods

Data

We used 1051 EEGs with and without IEDs as reported by experts (median duration, 52.5 minutes; range, 0.54-833 minutes) performed at Massachusetts General Hospital during clinical care (Table). The study was approved by the

Table. Characteristics of Patients and EEGs Used in Training and Testing of SpikeNet

Patient Age Range, y	No. (%)				
	Total	Female	EEG ^a	EMU	ICU
1051 EEGs Annotated by 8 Experts^b					
0 to <1	30 (3)	16 (53)	30 (100)	0	0
1 to <5	84 (8)	41 (49)	80 (95)	3 (4)	1 (1)
5 to <13	267 (25)	131 (49)	262 (98)	3 (1)	2 (1)
13 to <18	117 (11)	56 (48)	114 (97)	3 (3)	0
18 to <30	121 (12)	73 (60)	114 (94)	5 (4)	2 (2)
30 to <50	107 (10)	54 (50)	100 (93)	4 (4)	3 (3)
50 to <65	133 (13)	58 (44)	122 (92)	3 (2)	8 (6)
65 to <75	98 (9)	62 (63)	89 (91)	0	9 (9)
≥75	94 (9)	49 (52)	85 (90)	2 (2)	7 (7)
Totals	1051 (100)	540 (51)	996 (95)	23 (2)	32 (3)
8520 Control EEGs^c					
0 to <1	183 (2)	73 (40)	183 (100)	0	0
1 to <5	482 (6)	205 (43)	482 (100)	0	0
5 to <13	617 (7)	248 (40)	617 (100)	0	0
13 to <18	428 (5)	192 (45)	428 (100)	0	0
18 to <30	1121 (13)	539 (48)	1121 (100)	0	0
30 to <50	1569 (18)	823 (52)	1569 (100)	0	0
50 to <65	1761 (21)	864 (49)	1761 (100)	0	0
65 to <75	1185 (14)	560 (47)	1185 (100)	0	0
≥75	1174 (14)	579 (49)	1174 (100)	0	0
Totals	8520 (100)	4083 (48)	8520 (100)	0	0

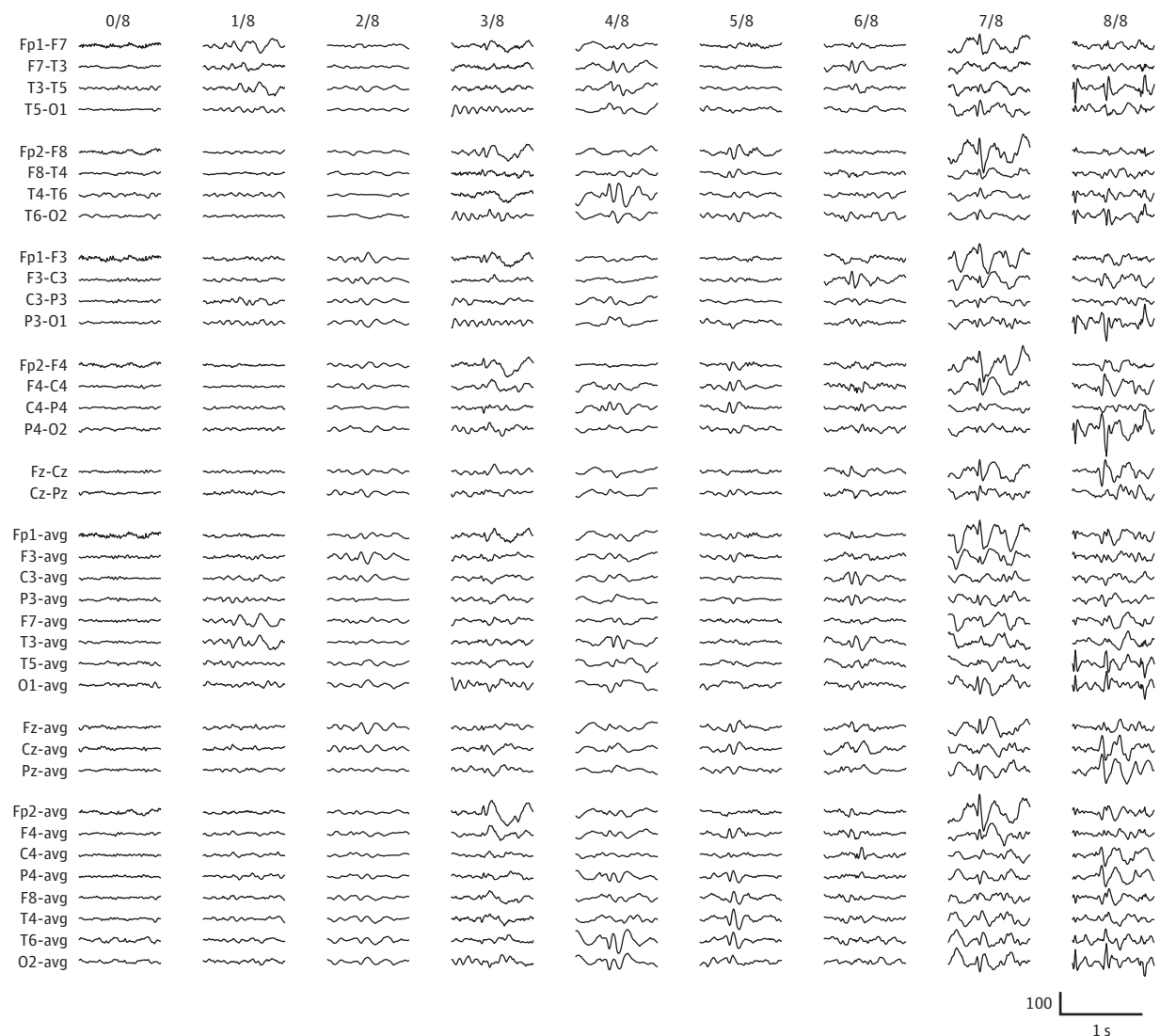
Abbreviations:
EEG, electroencephalogram;
EMU, epilepsy monitoring unit;
ICU, intensive care unit.

^a Routine EEG (outpatient and inpatient service).

^b Used in training and testing with patient-wise 10-fold cross-validation.

^c Used only to augment training data.

Figure 1. Example Electroencephalographic Segments of Interictal Epileptiform Discharges Identified at Each Expert Agreement Level



Standard banana, bipolar montage 1-second displays with expert agreement ratios: number of experts identifying that segment as containing a spike divided by total number of experts who performed annotations displayed across the top (eg, 0/8 indicates none of the 8 experts).

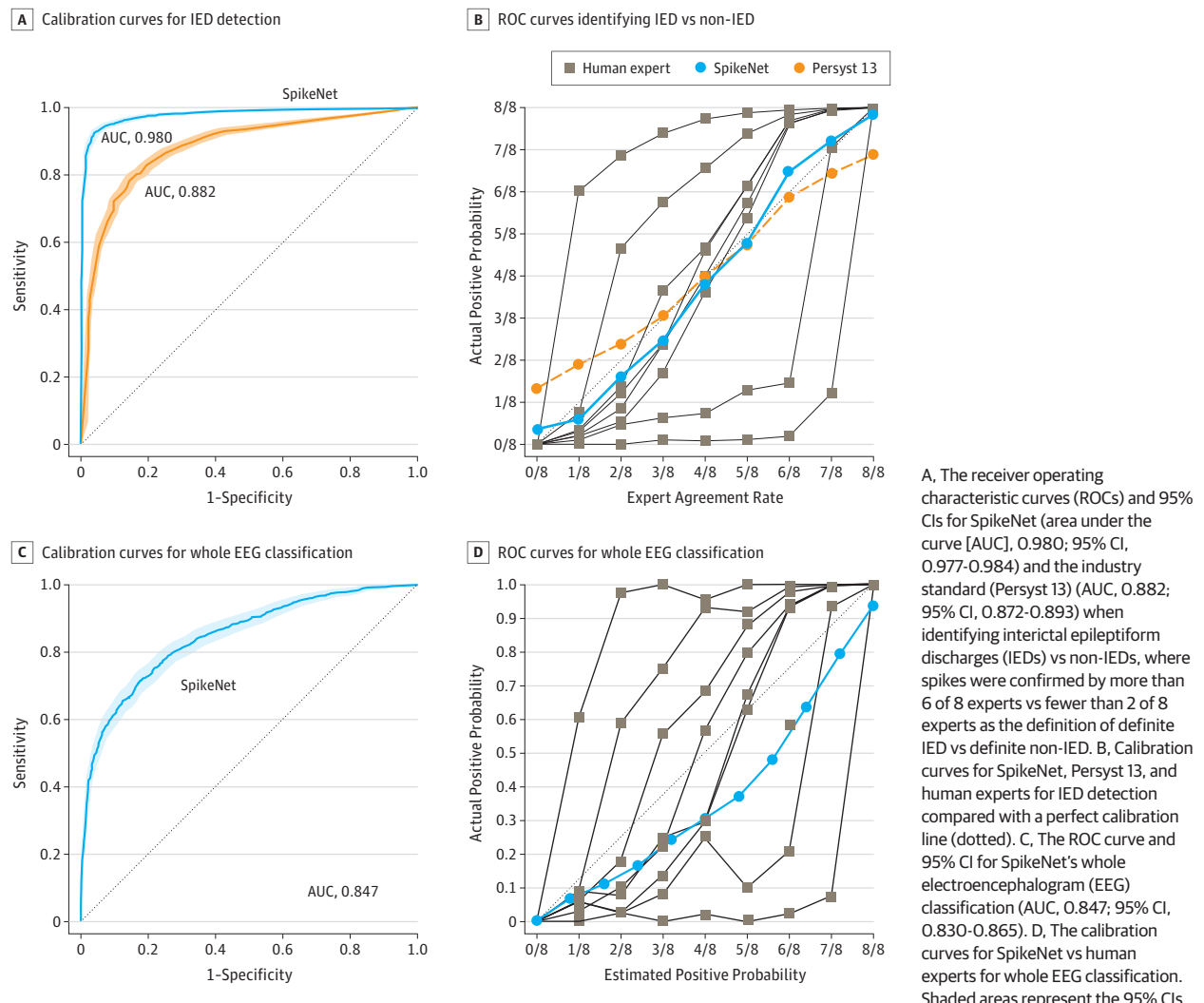
Massachusetts General Hospital Institutional Review Board with waiver of informed consent. The EEGs were filtered (60-Hz notch, 0.5-Hz high-pass) and resampled to 128 Hz. The process used to annotate candidate IEDs is described by Jing et al¹³ in this issue of *JAMA Neurology* and in the eMethods and eFigures 1-3 in the [Supplement](#).

We trained a convolutional neural network (CNN) to perform IED detection (eFigure 4 in the [Supplement](#)), using soft (nonbinary) labels between 0/8 and 8/8 (ie, the proportion of 8 experts voting yes for a given IED candidate) (eMethods in the [Supplement](#)). We used 10-fold, patientwise cross-validation, in which samples from the 1051 annotated EEGs were partitioned into training, validation, and testing sets. Training and validation sets were used for parameter optimization, and the testing set to measure algorithm performance. Training was conducted in 2 steps. In step 1, we

randomly sampled non-IED examples from the 8520 control EEGs to better balance the training set, and trained the model using these plus candidate IEDs from the 1051 annotated EEGs. In step 2, we used the 8520 control EEGs (Table) to identify 6 258 483 difficult samples (eg, benign variants and artifacts that mimic IEDs) assigned by the step 1 CNN a probability greater than 0.43 (the threshold minimizing calibration error in the training set). We also included 13 228 000 examples with probability less than or equal to 0.43 to further increase variety among non-IEDs in the training data. We retrained the CNN on the enriched data using the same methods as in step 1. This final CNN is called SpikeNet. Performance for SpikeNet was measured only on examples annotated by the 8 experts.

We trained SpikeNet to classify an entire EEG as containing IEDs vs no IEDs. This process was done by running SpikeNet

Figure 2. SpikeNet Performance Evaluation



A, The receiver operating characteristic curves (ROCs) and 95% CIs for SpikeNet (area under the curve [AUC], 0.980; 95% CI, 0.977-0.984) and the industry standard (Persyst 13) (AUC, 0.882; 95% CI, 0.872-0.893) when identifying interictal epileptiform discharges (IEDs) vs non-IEDs, where spikes were confirmed by more than 6 of 8 experts vs fewer than 2 of 8 experts as the definition of definite IED vs definite non-IED. B, Calibration curves for SpikeNet, Persyst 13, and human experts for IED detection compared with a perfect calibration line (dotted). C, The ROC curve and 95% CI for SpikeNet's whole electroencephalogram (EEG) classification (AUC, 0.847; 95% CI, 0.830-0.865). D, The calibration curves for SpikeNet vs human experts for whole EEG classification. Shaded areas represent the 95% CIs.

over the entire EEG recording to obtain a time series of IED probabilities, then extracting features that were used for binary classification. Details are in the eMethods and eFigure 5 in the Supplement.

We evaluated SpikeNet on testing data using 2 approaches. First, we evaluated discrimination of definite IEDs vs definite non-IEDs, defined as IED candidates with more than 6 vs less than 2 votes of the 8 experts; performance was measured via area under the receiver operating characteristic curve (AUC). For comparison, we also evaluated the industry-standard commercial IED detection algorithm (Persyst 13), which provides a continuous perception score (waves with higher scores are more likely to be IEDs). Applying various thresholds to the Persyst 13 perception score allows calculation of the Persyst 13 AUC for direct comparison with SpikeNet. Second, we compared calibration between SpikeNet and experts on all labeled IED candidates. An expert's calibration curve shows how likely they are to vote for IEDs (y-axis) vs the fraction of experts who agree (x-axis). An expert's calibration error is the mean absolute deviation from the diagonal line; 0

represents perfect calibration. We also evaluated SpikeNet's ability to classify entire EEG recordings using the same method.

Statistical Analysis

The 95% CIs and P values were obtained via 10 000 rounds of bootstrapping. Findings were considered significant at $P < .05$ with 2-tailed testing. Statistical computations were performed with software written in house.

Results

Detection of IEDs

Example candidate IEDs are shown in Figure 1. Calibration curves for experts, SpikeNet, and the commercial detector are shown in Figure 2. Expert calibration varied: 2 experts were over-callers (above diagonal) and 2 were under-callers (below diagonal) (Figure 2A). The mean calibration error for experts was 0.183 (range, 0.081-0.364), compared with 0.066 (95% CI, 0.060-0.078) for the industry standard. Calibration

error for SpikeNet was 0.041 (95% CI, 0.033-0.049), exceeding the industry standard and mean expert performance ($P < .05$).

Binary classification performance is shown in Figure 2B. SpikeNet achieved an AUC of 0.980 (95% CI, 0.977-0.984), outperforming the industry standard of 0.882 (95% CI, 0.872-0.893) ($P < .05$) (Figure 2B).

Classification of Whole EEGs

Calibration curves for classifying entire EEGs as containing IEDs vs no IEDs are shown in Figure 2C for SpikeNet and human experts. An EEG is considered scored positive by an expert if they voted yes to any candidate IEDs from a given EEG. SpikeNet achieved a calibration error of 0.126 (95% CI, 0.109-0.144). The mean expert performance was 0.197 (range, 0.099-0.372). SpikeNet also performed satisfactory binary classification, with an AUC of 0.847 (95% CI, 0.830-0.865).

Discussion

SpikeNet appears to identify IEDs better than the current standard: neurologists with subspecialty training in clinical neurophysiology. SpikeNet achieved this by training on a large set of candidate IED waveforms independently annotated by multiple experts, learning, in effect, to simulate a committee of independent experts. Better-than-expert performance was achieved not only for categorizing individual waves as IEDs but also for determining whether an entire EEG contained any IEDs. While several studies have attempted to automate IED detection, to our knowledge, general-purpose, expert-level performance has not previously been demonstrated. By automating a diagnostic task previously limited to subspecialists,

SpikeNet opens the path for expanding medical care to a broader range of patients with epilepsy and other disorders in which IED detection is necessary.

Limitations

This study has limitations. One limitation is that some EEG patterns of clinical relevance are not detectable by SpikeNet. Examples include rhythmic patterns (eg, lateralized rhythmic delta activity) and focal slowing patterns with no IEDs. Similarly, in critical care, waveforms with blunted morphologic characteristics that are not considered IEDs may still be clinically significant when they occur as a train of periodic waves. Similarly, although seizures may include IEDs, some do not; SpikeNet does not solve the problem of seizure detection. SpikeNet is thus not a complete solution for all EEG applications.

Future work should address how automated IED detection can be integrated into practice. Initial work will likely focus on systems that amplify clinical neurophysiologists' capabilities. Subsequent work should focus on removing the need for clinicians to be involved in low-level, pattern-recognition aspects of EEG interpretation, such as IED detection.

Conclusions

SpikeNet appears to identify interictal epileptiform discharges more accurately, on average, than clinical experts with subspecialty training in clinical neurophysiology. This conclusion is suggested by the results of testing in a large set of clinical EEG recordings spanning the clinical range. SpikeNet may thus be a useful tool for accelerated review of EEGs.

ARTICLE INFORMATION

Accepted for Publication: August 11, 2019.

Published Online: October 21, 2019.

doi:10.1001/jamaneurol.2019.3485

Author Affiliations: Department of Neurology, Massachusetts General Hospital, Boston, Massachusetts (Jing, Sun, Kim, Maus, Chan, Dolatshahi, Muniz, Chu, Ge, Lam, Cole, Cash, Westover); School of Electrical & Electronic Engineering, Nanyang Technological University, Singapore (Jing, Dauwels); Department of Neurology, Yale University School of Medicine, New Haven, Connecticut (Herlopian); Department of Neurology, Emory University School of Medicine, Atlanta, Georgia (Karakis); Department of Neurology, University of Manitoba, Winnipeg, Manitoba, Canada (Ng); Department of Neurology, Medical University of South Carolina, Charleston, South Carolina (Halford); Department of Medical and Surgical Sciences, University Magna Graecia, Catanzaro, Italy (Sacca); Department of Neurology, University of Pennsylvania General Hospital, Boston, Massachusetts (Pathmanathan).

Author Contributions: Drs Jing, Sun, and Kim contributed equally as co-first authors. Dr Westover had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Concept and design: Jing, Herlopian, Dauwels, Cash, Westover.

Acquisition, analysis, or interpretation of data: Jing, Sun, Kim, Herlopian, Karakis, Ng, Halford, Maus, Chan, Dolatshahi, Muniz, Chu, Sacca, Pathmanathan, Ge, Lam, Cole, Cash, Westover.

Drafting of the manuscript: Jing, Sun, Kim, Herlopian, Cash, Westover.

Critical revision of the manuscript for important intellectual content: Jing, Sun, Kim, Herlopian, Karakis, Ng, Halford, Maus, Chan, Dolatshahi, Muniz, Chu, Sacca, Pathmanathan, Ge, Dauwels, Lam, Cole, Westover.

Statistical analysis: Jing, Sun, Herlopian, Ge, Westover.

Obtained funding: Cole, Westover.

Administrative, technical, or material support: Jing, Sun, Kim, Karakis, Ng, Halford, Maus, Sacca, Dauwels, Westover.

Supervision: Jing, Kim, Chu, Cole, Cash, Westover.

Conflict of Interest Disclosures: Dr Muniz reported having a patent to 10,349,888 issued. Dr Chu reported receiving grants from the National Institutes of Health (NIH) during the conduct of the study and personal fees from Alliance Family of Companies, SleepMed, and Biogen outside the submitted work. Dr Lam reported receiving funding from Empatica outside the submitted work. Dr Westover reported receiving grants from NIH

during the conduct of the study. No other disclosures were reported.

Funding/Support: Dr Westover was supported by research grants 1K23NS090900, 1R01NS102190, 1R01NS102574, and 1R01NS107291 from the NIH National Institute of Neurological Disorders and Stroke (NINDS), and by grants from the Glenn Foundation for Medical Research. Dr Lam was funded by NIH NINDS grants K23 NS101037, R25 NS065743 and grants from the American Academy of Neurology. Dr Cash was funded by NIH NINDS grants R01 NS062092 and K24 NS088568. Dr Jing received support from Sage Therapeutics. Dr Sun reported receiving support from NIH NINDS grant 1K23NS090900. Dr Kim reported receiving support from NIH NINDS grant R25NS06574, the American Heart Association, and the Bee Foundation.

Role of the Funder/Sponsor: The funding organizations had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

REFERENCES

1. Seidel S, Pablik E, Aull-Watschinger S, Seidl B, Pataria E. Incidental epileptiform discharges in

- patients of a tertiary centre. *Clin Neurophysiol*. 2016;127(1):102-107. doi:10.1016/j.clinph.2015.02.056
2. van Donselaar CA, Schimsheimer RJ, Geerts AT, Declerck AC. Value of the electroencephalogram in adult patients with untreated idiopathic first seizures. *Arch Neurol*. 1992;49(3):231-237. doi:10.1001/archneur.1992.00530270045017
3. Fountain NB, Freeman JM. EEG is an essential clinical tool: pro and con. *Epilepsia*. 2006;47(suppl 1):23-25. doi:10.1111/j.1528-1167.2006.00655.x
4. Gavvala J, Abend N, LaRoche S, et al; Critical Care EEG Monitoring Research Consortium (CCEMRC). Continuous EEG monitoring: a survey of neurophysiologists and neurointensivists. *Epilepsia*. 2014;55(11):1864-1871. doi:10.1111/epi.12809
5. Westover MB, Shafi MM, Bianchi MT, et al. The probability of seizures during EEG monitoring in critically ill adults. *Clin Neurophysiol*. 2015;126(3):463-471. doi:10.1016/j.clinph.2014.05.037
6. Kim JA, Boyle EJ, Wu AC, et al. Epileptiform activity in traumatic brain injury predicts post-traumatic epilepsy. *Ann Neurol*. 2018;83(4):858-862. doi:10.1002/ana.25211
7. Kim JA, Rosenthal ES, Biswal S, et al. Epileptiform abnormalities predict delayed cerebral ischemia in subarachnoid hemorrhage. *Clin Neurophysiol*. 2017;128(6):1091-1099. doi:10.1016/j.clinph.2017.01.016
8. Lam AD, Deck G, Goldman A, Eskandar EN, Noebels J, Cole AJ. Silent hippocampal seizures and spikes identified by foramen ovale electrodes in Alzheimer's disease. *Nat Med*. 2017;23(6):678-680. doi:10.1038/nm.4330
9. Gotman J, Ives JR, Gloor P. Automatic recognition of inter-ictal epileptic activity in prolonged EEG recordings. *Electroencephalogr Clin Neurophysiol*. 1979;46(5):510-520. doi:10.1016/0013-4694(79)90004-X
10. Scheuer ML, Bagic A, Wilson SB. Spike detection: Inter-reader agreement and a statistical Turing test on a large data set. *Clin Neurophysiol*. 2017;128(1):243-250. doi:10.1016/j.clinph.2016.11.005
11. Jing J, Dauwels J, Rakthanmanon T, Keogh E, Cash SS, Westover MB. Rapid annotation of interictal epileptiform discharges via template matching under dynamic time warping. *J Neurosci Methods*. 2016;274:179-190. doi:10.1016/j.jneumeth.2016.02.025
12. Westover MB, Halford JJ, Bianchi MT. What it should mean for an algorithm to pass a statistical Turing test for detection of epileptiform discharges. *Clin Neurophysiol*. 2017;128(7):1406-1407. doi:10.1016/j.clinph.2017.02.026
13. Jing J, Herlopian A, Karakis I, et al. Interrater reliability of experts in identifying interictal epileptiform discharges in electroencephalograms [published online October 21, 2019]. *JAMA Neurol*. doi:10.1001/jamaneurol.2019.3531