JAMA Neurology | Original Investigation

Noninvasive Detection of Hippocampal Epileptiform Activity on Scalp Electroencephalogram

Maurice Abou Jaoude, MS; Claire S. Jacobs, MD, PhD; Rani A. Sarkis, MD, MSc; Jin Jing, PhD; Kyle R. Pellerin, BS; Andrew J. Cole, MD; Sydney S. Cash, MD, PhD; M. Brandon Westover, MD, PhD; Alice D. Lam, MD, PhD

IMPORTANCE The hippocampus is a highly epileptogenic brain region, yet over 90% of hippocampal epileptiform activity (HEA) cannot be identified on scalp electroencephalogram (EEG) by human experts. Currently, detection of HEA requires intracranial electrodes, which limits our understanding of the role of HEA in brain diseases.

OBJECTIVE To develop and validate a machine learning algorithm that accurately detects HEA from a standard scalp EEG, without the need for intracranial electrodes.

DESIGN, SETTING, AND PARTICIPANTS In this diagnostic study, conducted from 2008 to 2021, EEG data were used from patients with temporal lobe epilepsy (TLE) and healthy controls (HCs) to train and validate a deep neural network, HEAnet, to detect HEA on scalp EEG. Participants were evaluated at tertiary-level epilepsy centers at 2 academic hospitals: Massachusetts General Hospital (MGH) or Brigham and Women's Hospital (BWH). Included in the study were patients aged 12 to 78 years with a clinical diagnosis of TLE and HCs without epilepsy. Patients with TLE and HCs with a history of intracranial surgery were excluded from the study.

EXPOSURES Simultaneous intracranial EEG and/or scalp EEG.

MAIN OUTCOMES AND MEASURES Performance was assessed using cross-validated areas under the receiver operating characteristic curve (AUC ROC) and precision-recall curve (AUC PR) and additional clinically relevant metrics.

RESULTS HEAnet was trained and validated using data sets that were derived from a convenience sample of 141 eligible participants (97 with TLE and 44 HCs without epilepsy) whose retrospective EEG data were readily available. Data set 1 included the simultaneous scalp EEG and intracranial electrode recordings of 51 patients with TLE (mean [SD] age, 40.7 [15.9] years; 30 men [59%]) at MGH. An automatically generated training data set with 972 095 positive HEA examples was created, in addition to a held-out expert-annotated testing data set with 22 762 positive HEA examples. HEAnet's performance was validated on 2 independent scalp EEG data sets: (1) data set 2 (at MGH; 24 patients with TLE and 20 HCs; mean [SD] age, 42.3 [16.2] years; 17 men [39%]) and (2) data set 3 (at BWH; 22 patients with TLE and 24 HCs; mean [SD] age, 43.0 [14.4] years; 20 men [43%]). For single-event detection of HEA on data set 1, HEAnet achieved a mean (SD) AUC ROC of 0.89 (0.01) and a mean (SD) AUC PR of 0.39 (0.03). On external validation with data sets 2 and 3, HEAnet accurately distinguished TLE from HC (AUC ROC of 0.88 and 0.95, respectively) and predicted epilepsy lateralization with 100% and 92% accuracy, respectively. HEAnet tracked dynamic changes in HEA in response to seizure medication adjustments and performed comparably with human experts in diagnosing TLE from 1-hour scalp EEG recordings, diagnosing TLE in several individuals that experts missed. Without reducing specificity, addition of HEAnet to human expert EEG review increased sensitivity for diagnosing TLE in humans from 50% to 58% to 63% to 67%.

CONCLUSIONS AND RELEVANCE Results of this diagnostic study suggest that HEAnet provides a novel, noninvasive, quantitative, and clinically relevant biomarker of hippocampal hyperexcitability in humans.

JAMA Neurol. 2022;79(6):614-622. doi:10.1001/jamaneurol.2022.0888
Published online May 2, 2022.

Supplemental content

Author Affiliations: Department of Neurology, Massachusetts General Hospital, Boston (Abou Jaoude, Jacobs, Jing, Pellerin, Cole, Cash, Westover, Lam); Harvard Medical School, Boston, Massachusetts (Jacobs, Sarkis, Cole, Cash, Westover, Lam); Department of Neurology, Brigham and Women's Hospital, Boston, Massachusetts (Sarkis).

Corresponding Author: Alice D. Lam, MD, PhD, Department of Neurology, Massachusetts General Hospital, 55 Fruit St, Wang Ambulatory Care Center 735, Boston, MA 02114 (lam.alice@mgh.harvard.edu).

There are no easily available methods to evaluate the electrical activity arising from deep brain structures. Scalp electroencephalogram (EEG) is the primary tool used to assess the brain's electrical activity and captures electrical activity from the brain's cortical surface with high temporal resolution. However, when adjudicated visually, scalp EEG is relatively insensitive to electrical activity arising from deep brain regions, including the hippocampus.

The hippocampus is essential for memory formation and plays an important role in many neurologic disorders, including temporal lobe epilepsy and Alzheimer disease. In the diseased state, the hippocampus is highly prone to generating abnormal spiking activity, or hippocampal epileptiform activity (HEA), which can cause memory impairments, psychiatric disturbances, and can signal impending seizures.[1-5] Up to 95% of HEA cannot be detected using standard clinical interpretation of scalp EEG.[5-10] Currently, detection of HEA requires surgical placement of intracranial electrodes, an invasive procedure in which electrodes are inserted into or adjacent to the hippocampus, where they can directly record its electrical activity.[11,12] As intracranial recordings are costly, require specialized epilepsy centers with surgical expertise, and may carry morbidity for the patient, only a minority of patients with epilepsy will undergo intracranial electrode recordings. However, there are many clinical and research applications where the ability to noninvasively monitor HEA could have an important impact, not only on clinical care, but also on our fundamental understanding of the neurophysiology of brain diseases.

Here, we developed and validated a deep learning algorithm, HEAnet, to accurately and noninvasively detect HEA using only information extracted from a standard scalp EEG recording.

Key Points

Question Can a computer algorithm be trained to detect hippocampal epileptiform activity (HEA) on scalp electroencephalogram (EEG) when human experts cannot identify this activity?

Findings In this diagnostic study of 141 eligible participants, a deep learning algorithm was trained to detect HEA on scalp EEG, using a data set of combined scalp EEG and intracranial recordings. The algorithm accurately detected HEA at the single-event level and performed well on clinically relevant metrics, including quantification and lateralization of HEA and diagnosis of temporal lobe epilepsy.

Meaning Results of this diagnostic study suggest that a computational algorithm can noninvasively detect HEA from scalp EEG and may improve diagnosis and treatment of epilepsy and other brain diseases.

Methods

Clinical Data Sets

In this diagnostic study, clinical data were analyzed retrospectively under a research protocol approved by MassGeneral Brigham, the single institutional review board shared by Massachusetts General Hospital (MGH) and Brigham and Women's Hospital (BWH). As the research involved secondary analysis of existing clinical data and incurred no more than minimal risk, the requirement for informed consent was waived. Data set 1 consisted of 8395 hours of simultaneous foramen ovale (FO) electrode and scalp EEG recordings in patients with temporal lobe epilepsy (TLE) who were monitored in the MGH epilepsy monitoring unit (EMU) from 2008 to 2019. Data set 2 consisted of 4433 hours of scalp EEG recordings from patients with TLE, unique from data set 1, and healthy controls (HCs) without epilepsy monitored in the MGH EMU between 2014 and 2020. Data set 3 consisted of 2137 hours of scalp EEG recordings from patients with TLE and HCs monitored in the Brigham and Women's Hospital (BWH) EMU between 2016 and 2021. Information on patient race and ethnicity was not readily available for these retrospective convenience data sets and, therefore, was not used in this investigation. This study followed the Standards for Reporting of Diagnostic Accuracy (STARD) reporting guidelines.

Clinical diagnosis of TLE was based on seizure semiology, neurophysiologic findings, and neuroimaging. Lateralization of epilepsy (ie, seizure-onset zone) was determined by board-certified epileptologists (A.D.L., R.A.S.), taking into account seizure onsets captured in the EMU, imaging findings, and other relevant clinical details. Patients and HCs with a history of intracranial surgery or multifocal epilepsy were excluded. HCs underwent EMU evaluation and were determined not to have epilepsy, based on the absence of interictal epileptiform abnormalities, capture of typical spells with absence of ictal EEG correlate, and lack of clinical suspicion for epilepsy on follow-up assessment by a board-certified epileptologist (eTables 1, 2, and 3 in the Supplement).

General Approach for Developing HEAnet

HEAnet was developed using data set 1 (simultaneous scalp EEG and FO electrode recordings). FO electrodes are intracranial electrodes that are inserted through the cheek and fluoroscopically guided into the cranium through the FO.[13,14] The FO electrode contacts are positioned adjacent to the mesial temporal lobe and record hippocampal activity with high fidelity. Although most HEA cannot be seen on scalp EEG, it is easily identified on FO recordings. We used FO recordings to obtain ground-truth timing information for when HEA occurred and trained convolutional neural networks (CNNs) to learn the corresponding HEA signature on the simultaneously recorded scalp EEG.

Design Considerations for HEAnet

Prior studies using scalp EEG averaging demonstrated that HEA generates a scalp EEG correlate but with a low signal-to-noise ratio that precludes visual detection at the single-event level.[6,7] We made several design choices to optimize our ability to detect HEA at the single-event level.

First, to maximize the signal-to-noise ratio of HEA on scalp EEG, we designed HEAnet to detect HEA occurring during sleep. Large-amplitude artifacts from movement, myogenic activity, and eye blinks are minimized during sleep, and in most patients with TLE, HEA occurs with highest frequency during sleep.[15-18]

Second, we created a massive, automatically generated training data set of labeled HEA and non-HEA examples. Development of most spike detection algorithms depends on human experts to manually annotate spike examples to create a training data set. This sets a practical limitation on the number of training examples that can be annotated and limits the complexity of machine learning algorithms that can be applied. Here, we used previously developed algorithms to screen the entirety of data set 1, automatically extracting all sleep periods based on the scalp EEG,[19] and automatically labeling HEA (positive) and non-HEA (negative) training examples based on the FO recordings using FOnet,[20] a deep learning algorithm that detects HEA on FO recordings (eMethods in the Supplement). This generated a training data set with more than 2 million labeled scalp EEG examples of HEA during sleep, allowing us to leverage deep learning algorithms to extract salient features of HEA from raw scalp EEG.

Third, although we trained HEAnet using an automatically generated training data set, we assessed its performance on a held-out, criterion-standard, expert-annotated testing data set. The testing data set consisted of 51 hours of recordings (1 hour per patient) that were independently annotated by 2 board-certified epileptologists (A.D.L., C.S.J.) and contained over 20 000 positive HEA examples, based on expert consensus (eMethods in the Supplement). The training data set for HEAnet was a balanced data set (equal number of HEA and non-HEA examples), whereas the testing data set consisted of continuous EEG recordings (imbalanced, with many more non-HEA than HEA examples). The balanced training data set allowed HEAnet to best learn to distinguish HEA from non-HEA, whereas testing on continuous EEG recordings more accurately represents HEAnet's performance in real-world settings.

CNN Training and Evaluation

We first trained individual CNNs to detect HEA from scalp EEG. The input to each CNN was a 25 × 128 matrix representing a 500-millisecond epoch (128 samples) of preprocessed scalp EEG data from 25 bipolar channels (eMethods in the Supplement). The CNN's output is the probability that HEA occurred in the central 250 millisecond of the epoch. CNNs were evaluated using 5-fold cross-validation, stratified across patients (**Figure 1**). For each fold, training data came from the automatically generated training data set (pooled across patients in the training group), whereas testing data came from the expert-annotated recordings (pooled across patients in the testing group). We report cross-validated performance metrics pooled across all testing folds (eMethods in the Supplement).

HEAnet Architecture

HEAnet is an ensemble of 6 top-performing CNN models (**Figure 2**; eTables 4, 5, 6, and eMethods in the Supplement). All CNNs are given the same 500-millisecond scalp EEG input, and the output of HEAnet is the mean output of all 6 CNNs.

Statistical Analysis

Population statistics are reported as mean (SD). Correlation was assessed using Spearman correlation coefficient. We tested trend across groups using the Cuzick test, differences between groups using the Mann-Whitney *U* test, and paired comparisons using the Wilcoxon signed rank test. Statistical significance was determined as $P < .05$ with 2-tailed testing. Statistical analysis was performed using Matlab 2018 (Mathworks) and the scipy.stats package (Python).

Results

HEAnet was trained and validated using data sets that were derived from a convenience sample of 141 eligible participants (97 with TLE and 44 HCs without epilepsy) whose retrospective EEG data were readily available. Data set 1 included the simultaneous scalp EEG and intracranial electrode recordings of 51 patients with TLE (mean [SD] age, 40.7 [15.9] years; 30 men [59%]; 21 women [41%]) at MGH. We created an automatically generated training data set with 972 095 positive HEA examples and a held-out expert-annotated testing data set with 22 762 positive HEA examples. HEAnet's performance was validated on 2 independent scalp EEG data sets: (1) data set 2 (at MGH; 24 patients with TLE and 20 HCs; mean [SD] age, 42.3 [16.2] years; 17 men [39%]; 27 women [61%]) and (2) data set 3 (at BWH; 22 patients with TLE and 24 HCs; mean [SD] age, 43.0 [14.4] years; 20 men [43%]; 26 women [57%]) (**Table**).

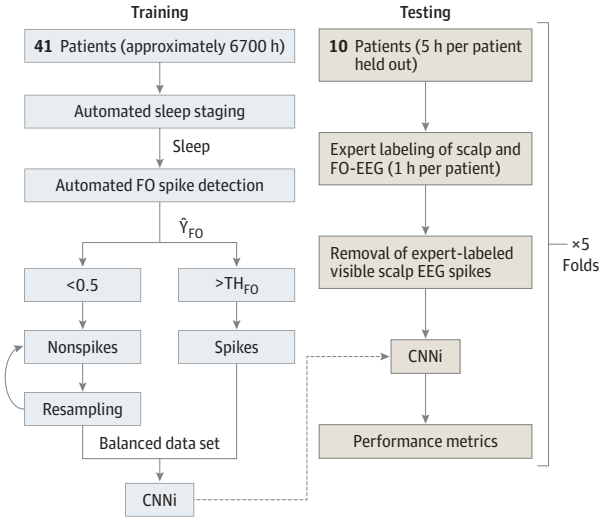
Associations Between HEAnet Output, Scalp EEG Findings, and Features of HEA on FO Electrodes

Figure 3A shows a representative 8-second recording from a patient with TLE, highlighting detections made by HEAnet on scalp EEG. At the single-event level, no obvious scalp EEG signature of HEA was discerned. The mean (SD) scalp EEG signal of all the HEA epochs that were correctly detected by HEAnet revealed a low-amplitude (19.1 [9.4] μV) deflection lasting 150 to 200 milliseconds (eFigure 1 in the Supplement). The median latency between the mean HEA peak on FO electrodes and the peak of the mean deflection on scalp EEG was 0 milliseconds (IQR, −3.91 to 97.66), suggesting that HEAnet may detect a volume-conducted signal (eResults in the Supplement).

We evaluated HEAnet's output ($\hat{Y}_{HEA}$, the probability that HEA occurs in a given epoch, based on scalp EEG input) in association with corresponding HEA features on FO electrodes. $\hat{Y}_{HEA}$ was weakly but significantly correlated with both the amplitude (mean [SD] Spearman ρ [43] = 0.26 [0.25]; 95% CI, 0.19-0.33; $P < .001$) and slope (mean [SD] Spearman ρ [43] = 0.32 [0.27]; 95% CI, 0.24-0.40; $P < .001$) of the corresponding HEA on FO electrodes (eFigure 2 in the Supplement).

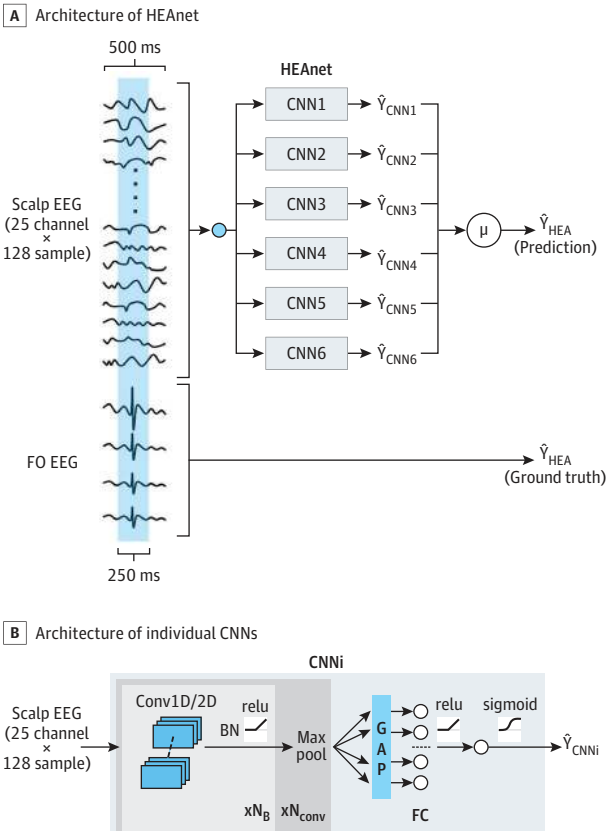
We next labeled testing examples from the expert-annotated data set as negative, equivocal, or positive, depending on whether 0, 1, or 2 experts, respectively, annotated HEA on FO electrodes during the central 250 milliseconds of each example. Figures 3B and 3C show the population distributions of $\hat{Y}_{HEA}$ for all negative, equivocal, or positive HEA testing examples. There was a significant trend of increasing median $\hat{Y}_{HEA}$ going from negative (0.13; IQR, 0.06-0.27), to equivocal (0.34; IQR, 0.14-0.64), to positive (0.58; IQR, 0.26-

Figure 1. Convolutional Neural Network (CNN) Training and Performance Evaluation



For each cross-validation fold, 10 patients are set aside for testing, while the rest of the patients are used for training. Training data were generated automatically from recordings of the training patients, using automated algorithms for sleep staging[19] and intracranial spike detection.[20] The scheme shown is for training without an early stopping set (eMethods in the Supplement). The testing data for each patient consists of a 1-hour expert-annotated recording. Before applying the trained CNN and evaluating performance, all epochs that contained visible epileptiform discharges on the scalp electroencephalogram (EEG) were removed, to provide a rigorous demonstration of HEAnet's performance, independent of scalp-visible epileptiform discharges. CNNi indicates *i*-th CNN; FO, foramen ovale; TH, threshold.

Figure 2. HEAnet Architecture



A, The input to HEAnet is a 25 channel × 128 sample (500msec) scalp electroencephalogram (EEG) segment, which is fed into each of 6 convolutional neural networks (CNNs). The outputs of all CNNs ($\hat{Y}_{CNNi}$) are averaged to yield a final probability ($\hat{Y}_{HEA}$) that the central 250 milliseconds of the input segment contains HEA. $\hat{Y}_{HEA}$ is compared with the ground truth, Y_{HEA}, which is defined based on expert annotation of the foramen ovale (FO) recordings.
B, Architecture of the individual CNNs that comprise HEAnet. The 25 channel × 128 sample EEG is passed through a number of convolutional blocks (N_B) that are comprised of 1-dimensional (D) or 2-D convolutional filters, followed by application of batch normalization (BN) and an activation function (relu). The resulting signal is then passed through a maxpool layer. This sequence of operations (convolution block + maxpool) is repeated for a number (N_{conv}) of times. The resulting signal is passed through a global average pooling (GAP) layer and a fully connected layer (FC) before a final classification is done by a logistic regression unit (eMethods in the Supplement). HEA indicates hippocampal epileptiform activity.

0.92) HEA examples (Cuzick test $P < .001$). The population distribution of $\hat{Y}_{HEA}$ showed a peak of 0.03 (approximately 0) for negative HEA examples and a peak of 0.99 (approximately 1) for positive HEA examples.

Single-Event Level Performance of HEAnet

To evaluate HEAnet's ability to classify positive from negative HEA examples, we plotted receiver operating characteristic (ROC) and precision-recall (PR) curves (Figures 3D and 3E). HEAnet had a mean (SD) area under the curve (AUC) ROC of 0.89 (0.01) (AUC ROC for chance prediction is 0.5 and for perfect prediction is 1), and a mean (SD) AUC PR of 0.39 (0.03) (AUC PR for chance prediction is 0.016 [frequency of positive HEA examples in the testing set] and for perfect prediction is 1). Thus, HEAnet accurately detected HEA from scalp EEG at the single-event level, a task that human experts cannot perform.

At a classification threshold that yields a positive predictive value (PPV) of approximately 0.7, HEAnet had a mean (SD) specificity of 0.996 (0.002), a false-positive rate of 0.86 (0.34) per minute, and a sensitivity of 0.25 (0.08). At a PPV of approximately 0.9, HEAnet had a mean (SD) specificity of 0.999 (0.001), a false-positive rate of 0.11 (0.06) per minute, and a sensitivity of 0.13 (0.05). Notably, visual analysis of false positive detections by HEAnet (eResults, eFigure 3, and eTable 7 in the Supplement) revealed that 43% were false false-positive detections, ie, HEA that had been missed during expert annotation, which suggests that HEAnet's performance is actually higher than reported above.

At high specificity, the sensitivity of HEAnet at the single-event level may initially seem suboptimal. Two considerations place this performance into better clinical context. First, in many patients with TLE, HEA occurs abundantly, which can offset HEAnet's low sensitivity. On our expert-annotated recordings, HEA occurred on FO electrodes at a frequency approximately 12-fold higher than visible epileptiform discharges on scalp EEG. In most patients, the number of HEAnet detections (at PPV of approximately 0.9) was comparable with or exceeded the number of visible discharges on scalp EEG (eTable 1 in the Supplement). Second, most clinical applications of HEAnet do not require detection of every HEA event

Table. Data Sets for Training, Testing, and External Validation of HEAnet

Variable	Data set 1 (n = 51)	Data set 2 (n = 44)	Data set 3 (n = 46)
Institution	MGH	MGH	BWH
Type of recording	Scalp EEG and FO electrodes	Scalp EEG (full-length recordings)	Scalp EEG (full-length and clipped recordings)
Use for HEAnet	Training and testing	Independent validation (within same institution)	Independent validation (from external institution)
Patient, No. (%)			
TLE	51 (100)	24 (54.5)	22 (47.8)
HC	0	20 (45.5)	24 (52.2)
Age, mean (SD), y			
TLE	40.7 (15.9)	42.1 (15.1)	47.7 (14.4)
HC	0	42.6 (17.1)	38.7 (13)
Men, No. (%)			
TLE	30 (59)	12 (50)	12 (55)
HC	0	5 (25)	8 (33)
Women, No. (%)			
TLE	21 (41)	12 (50)	10 (45.5)
HC	0	15 (75)	16 (67)
EEG data per patient, mean (SD), h			
TLE	164.6 (102.9)	120.6 (40.1)	63.3 (57.3)
HC	0	76.9 (58.7)	31.1 (41.6)
Seizure onset location (L/R/B/I), patient No. (%)			
TLE	15 (29)/8 (16)/20 (39)/8 (16)	14 (58)/6 (25)/4 (17)/0	12 (55)/6 (27)/4 (18)/0
HC	NA	NA	NA

Abbreviations: B, bitemporal; BWH, Brigham and Women's Hospital; EEG, electroencephalogram; FO, foramen ovale; HC, healthy control; I, indeterminate; L, left temporal; MGH, Massachusetts General Hospital; R, right temporal; NA, not applicable; TLE, temporal lobe epilepsy.

that occurs. Rather, for a given patient, the ability to determine whether HEA occurs at all, assess the laterality of HEA, and quantify changes in HEA over time would be highly informative. We next evaluated HEAnet's performance on these more clinically relevant metrics.

Correlation of Cumulative Detections by HEAnet Over Time With HEA Rate and Laterality

We first examined HEAnet's ability to quantify HEA frequency over longer recording periods. For testing purposes, we held out 5 hours of sleep recording from each patient, of which 1 hour was annotated by experts. Applying HEAnet (operating at PPV 0.7) to the expert-annotated 1-hour recordings, we found a moderate correlation between the number of detections made by HEAnet on scalp EEG, and the number of positive HEA examples labeled by experts on FO electrodes (Spearman ρ [48] = 0.74; $P < .001$) (Figure 3F). We also assessed HEAnet's performance over the 5-hour held-out recordings. As only 1 hour was annotated, we used previously published algorithms to annotate the remaining 4 hours, applying FOnet[20] to annotate HEA on FO recordings, and SpikeNet,[21] a previously developed deep learning algorithm that detects visible epileptiform discharges on scalp EEG, to annotate visible epileptiform discharges on scalp EEG (eMethods in the Supplement). FOnet detections coinciding with SpikeNet detections were removed from analysis. On the 5-hour recordings, the number of detections made by HEAnet on scalp EEG correlated closely with the number of HEA on FO electrodes (Spearman ρ [48] = 0.78; $P < .001$) (Figure 3G; eResults and eFigure 4 in the Supplement).

We also tested whether left-right asymmetries in HEAnet detections could be useful for predicting epilepsy lateralization. Using the 1-hour, expert-annotated recordings, we calculated asymmetry indices using HEAnet detections on scalp EEG (asymmetry HEA), as well as expert annotations of HEA on FO electrodes (asymmetry EXP) (eMethods in the Supplement). Asymmetry HEA and asymmetry EXP were moderately correlated (Spearman ρ [48] = 0.66; $P < .001$) (eFigure 5 in the Supplement).

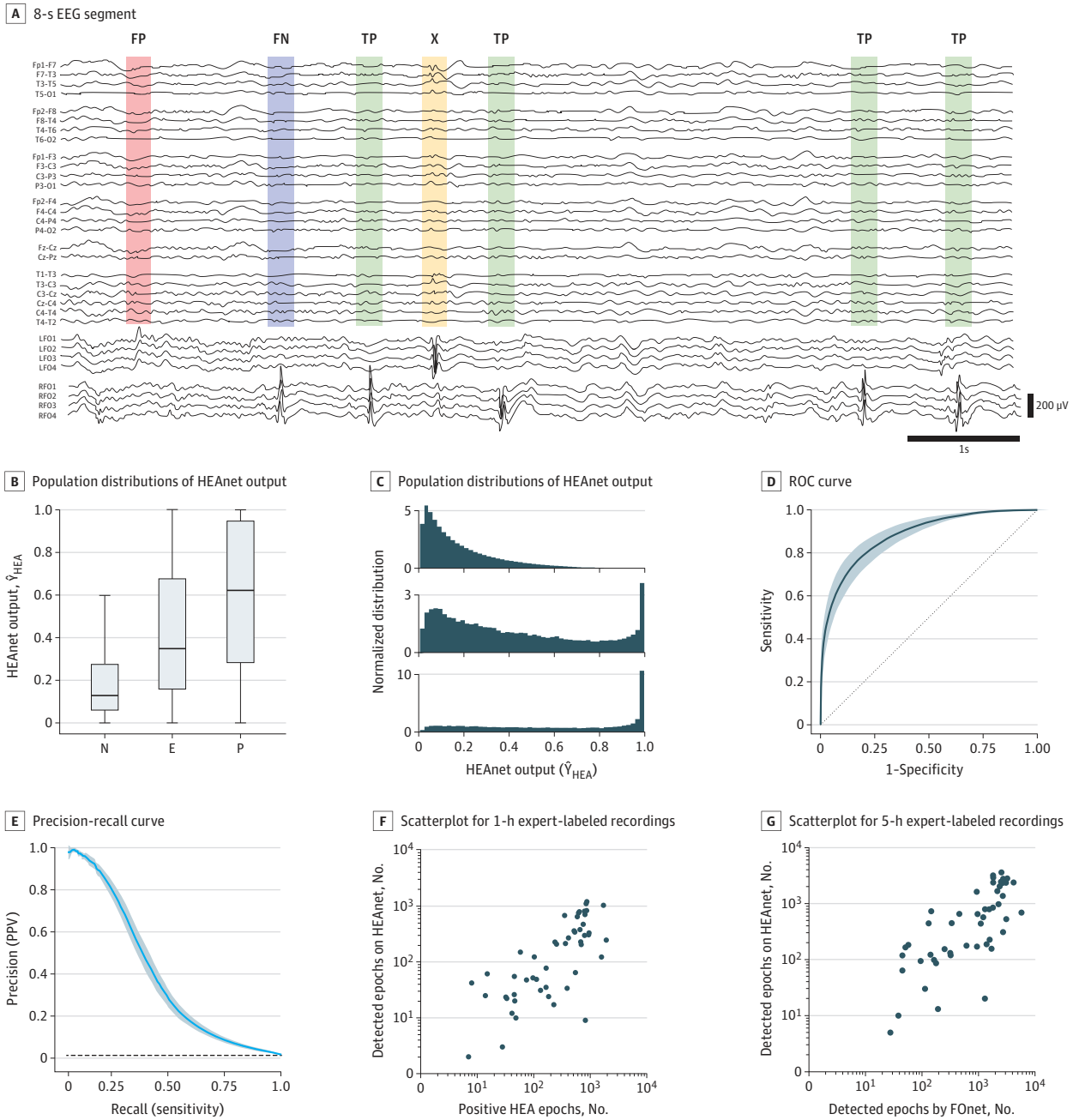
Performance of HEAnet Across Sleep and Awake States

Although HEAnet was optimized for use on sleep EEG data, we found that when applying HEAnet to data from the awake state, its specificity remained unchanged, although sensitivity was reduced by 30% (eTable 8 in the Supplement). Thus, HEAnet can be applied to data that include the awake state, without a substantial increase in false positive detections. We found no significant differences in HEAnet's performance (AUC ROC, AUC PR) across different sleep stages (eResults, eTable 8 in the Supplement).

Validation of HEAnet on an Independent Data Set From the Same Institution

To evaluate how HEAnet's performance generalizes on an independent data set, we applied HEAnet (operating at PPV 0.9) to data set 2 (eTable 2 in the Supplement). We found that the HEAnet detection rate (mean [SD] total number of HEAnet detections [left and right] made per hour of nonrapid eye movement [NREM] sleep) was significantly higher for patients with TLE than for HCs (268 [300] vs 15 [39],

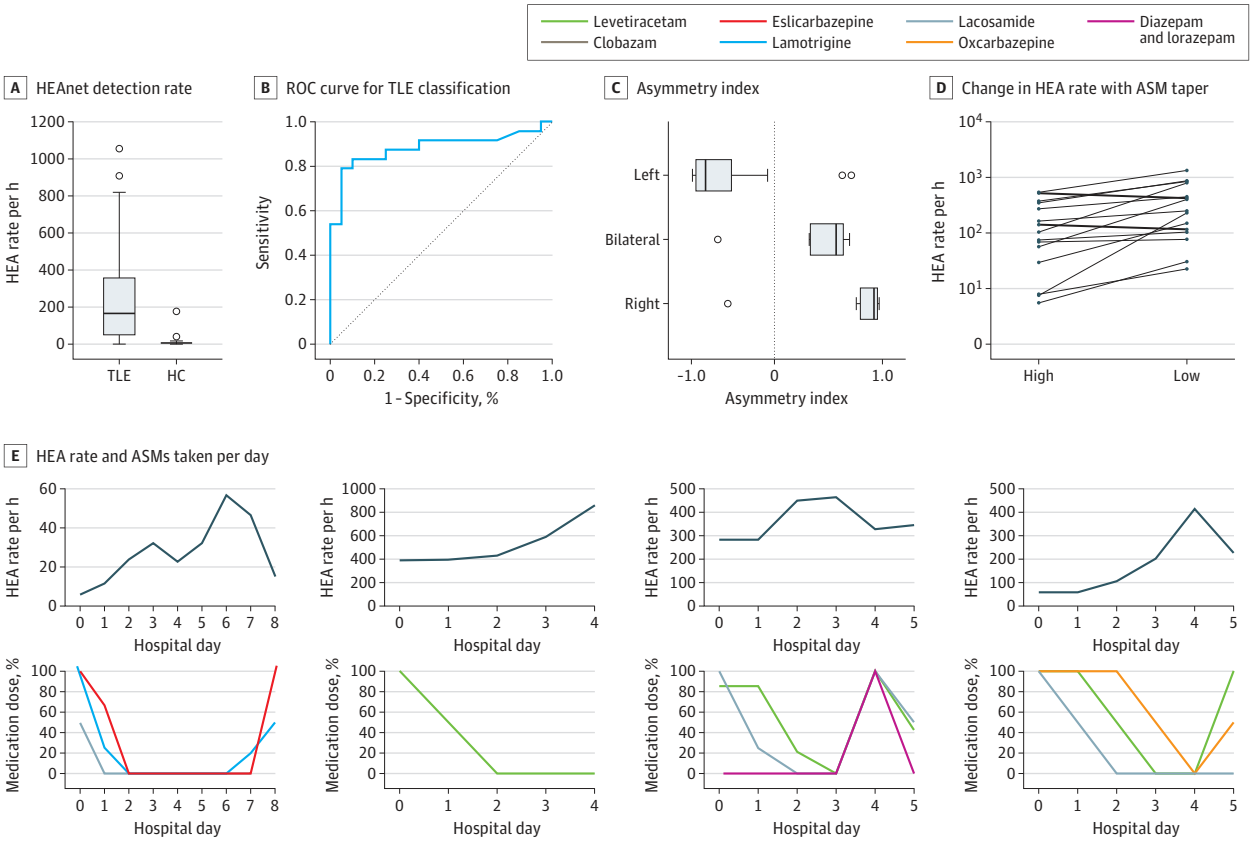
Figure 3. HEAnet Performance at the Single-Event Level



A, An 8-second electroencephalogram (EEG) segment from a representative patient from data set 1, showing examples of false positive (FP, red), false negative (FN, blue), and true positive (TP, green) detections from HEAnet. Epileptiform discharges visible on scalp EEG (X, yellow) were excluded from analysis. Note that the FP was detected on the left side and represents a "false" FP corresponding to hippocampal epileptiform activity (HEA) that was missed on expert annotation, likely due to its less robust morphology. B and C, Population distributions of HEAnet output ($\hat{Y}_{HEA}$) for negative (N), equivocal (E), and positive (P) HEA epochs, shown as box and whiskers plots and normalized histograms (density plots). D and E, Receiver operating characteristic (ROC) curve and precision-recall curve for classification of HEA epochs at the single-event level, respectively. The solid line and shaded regions correspond to the mean and SD of each curve, across all cross-validation folds. The black dashed line shows the expected performance for a random classifier. F, Scatterplot showing correlation between the number of HEA epochs detected by HEAnet on scalp EEG and the number of positive HEA epochs on foramen ovale (FO) electrodes, for 1-hour expert-labeled recordings. Each dot represents a unique patient. G, Scatterplot showing correlation between the number of HEA epochs detected by HEAnet on scalp EEG and the number of HEA epochs detected by FOnet on FO electrodes, for 5-hour held-out recordings. Each dot represents a unique patient. PPV indicates positive predictive value.

respectively; Mann-Whitney $U = 62$; $P < .001$) (**Figure 4**A). To assess whether the HEAnet detection rate could be used to distinguish patients with TLE from HCs, we plotted an ROC curve, where patients were classified as having TLE if their HEAnet detection rate on the left or right exceeded a certain threshold, which we varied to generate the plot. HEAnet dis-

Figure 4. HEAnet on Data Set 2



A, HEAnet detection rate (combined left- and right-sided detections, per hour of nonrapid eye movement [NREM] sleep), for patients with temporal lobe epilepsy (TLE) and healthy controls (HC) without epilepsy. B, Receiver operating characteristic (ROC) curve for classification of TLE vs HC based on HEAnet detection rate. C, Asymmetry index calculated based on HEAnet detections, for patients with left TLE, bitemporal epilepsy (bilateral), and right TLE. The Bilat group had clinical features suggesting right TLE predominance (eTable 2 in the Supplement). D, Line plots showing the detected hippocampal epileptiform activity (HEA) rate (per hour of NREM sleep) on the days that each patient was taking the highest amount of antiseizure medication (ASM; high) and the lowest amount of ASM (low). Each line represents a different patient who experienced an increase or decrease in HEA rate on withdrawal of medication. Patients were excluded if they did not undergo medication taper, or if they had an insufficient HEA rate (high HEA and low HEA each <10 detections per hour NREM) or insufficient NREM sleep (<2.5 hours per day) for accurate comparison. E, Relationship between detected HEA rate (per hour of NREM sleep, calculated over 24-hour intervals) and the amount of ASMs taken each day, over the course of an epilepsy monitoring unit stay for 4 representative patients with TLE. Medication doses are shown as a percentage of the maximal dose for each patient.

tinguished patients with TLE from HCs with an AUC ROC of 0.88 (Figure 4B).

We next evaluated HEAnet's ability to predict the laterality of TLE. We calculated the asymmetry index of HEAnet detections (asymmetry HEA) for all patients with TLE (eTable 2 in the Supplement). A total of 86% of patients with left TLE had a negative asymmetry HEA, and 83% of patients with right TLE had a positive asymmetry HEA (Figure 4C). Defining a threshold of |asymmetry HEA| greater than 0.75, for HEAnet to predict epilepsy laterality (ie, asymmetry HEA <−0.75 = left TLE and >0.75 = right TLE), HEAnet predicted epilepsy laterality in 58% of patients with TLE, with an accuracy of 100%.

Use of HEAnet on 1-Hour Sleep Recordings to Improve Epilepsy Diagnosis

We assessed the diagnostic capability of HEAnet on 1-hour sleep recordings, which can be reasonably attained in an outpatient setting. Using the first contiguous hour of recording with greater than 90% sleep for each patient in data set 2, we evaluated how well HEAnet distinguished patients with TLE from HCs, based on HEAnet detection rate. HEAnet distinguished patients with TLE from HCs with an AUC ROC of 0.823. We defined a threshold for diagnosis of TLE, as HEAnet detection rate more than 33 HEAnet detections per hour of NREM sleep on either the left or right side, which maximized the diagnostic specificity. Applying this threshold to the 1-hour recordings resulted in a diagnostic specificity of 1 and a sensitivity of 0.58. To directly compare this with human performance, 2 experts (A.D.L., C.S.J.) independently annotated the same 1-hour recordings, blinded to diagnosis, and determined whether the EEG was diagnostic of TLE (eTable 2 in the Supplement). One expert diagnosed TLE with a specificity of 1 and sensitivity of 0.5, whereas the other expert had a specificity of 0.75 and sensitivity of 0.58. A total of 13% of patients with TLE (3 of 24) were successfully diagnosed by HEAnet but missed by at least 1 expert. HEAnet correctly classified 2 patients with TLE that both

experts missed, and 1 patient with TLE that 1 expert missed (eTable 2 in the Supplement). Combining human expert review with HEAnet increased the diagnostic sensitivity for TLE from 0.50 to 0.58, for individual experts, to 0.63 to 0.67, without reducing diagnostic specificity.

Monitoring Changes in HEA in Response to Medication Adjustments

We next evaluated HEAnet's ability to monitor changes in HEA in response to changes in antiseizure medications (ASMs). Patients with epilepsy admitted to the EMU are typically tapered off ASMs at the beginning of the hospitalization, to increase likelihood of capturing seizures. Later, in preparation for discharge home, ASMs are restarted. We compared HEAnet detection rates for each patient, on days that they were receiving the highest and lowest amounts of ASM (HEA high and HEA low, respectively) (Figure 4D). In 89% of patients (16 of 18) with TLE, the HEAnet detection rate increased with reduction of ASMs. HEA low was significantly greater than HEA high (mean [SD], 334 [368] vs 157 [182] detections per hour, respectively; $P = .005$). Figure 4E illustrates changes in HEA detection rate for 4 representative patients. The HEA detection rate increased as ASMs were tapered and decreased with reintroduction of ASMs.

Validation of HEAnet on an Independent Data Set From an External Institution

As a final validation of HEAnet's performance generalizability, we applied HEAnet (operating at PPV 0.9) to a data set from an outside institution (data set 3, BWH) (eTable 3 in the Supplement). HEAnet's performance metrics on data set 3 were comparable with those in data set 2, distinguishing patients with TLE from HCs with an AUC ROC of 0.95 (eFigure 6 in the Supplement). Applying the threshold previously defined in data set 2 for diagnosis of TLE (HEAnet detection rate on either side >33 HEAnet detections per hour of NREM sleep), HEAnet correctly diagnosed TLE in 68% of patients with TLE (15 of 22), but incorrectly diagnosed TLE in 4% of HCs (1 of 24), yielding a PPV for TLE diagnosis of 93.7%. Applying the threshold previously defined in data set 2 for prediction of TLE laterality (|asymmetry HEA| >0.75), HEAnet predicted laterality in 59% of patients (13 of 22) with TLE, with an accuracy of 92%. Among patients with unilateral TLE, who were correctly diagnosed by HEAnet as having TLE (n = 12), HEAnet provided laterality predictions for 75% (n = 9), with 100% accuracy (eResults in the Supplement).

Discussion

Results of this diagnostic study suggest that HEAnet is an algorithm that noninvasively detects HEA using only information from a standard scalp EEG. Although the goal of most spike-detection algorithms is to automate the detection of epileptiform activity that is easily identified by human experts, HEAnet is unique in that it detects hippocampal epileptiform activity that cannot be identified by human experts.

One prior study attempted to develop an HEA detection algorithm on scalp EEG, using a small data set of 20-minute combined scalp EEG and FO recordings from 18 patients with TLE (total data set length, approximately 6 hours), with approximately 6100 HEA examples.[22] Using logistic regression with hand-crafted features, their optimized model achieved an AUC ROC of 0.67 at the single-event level.[22] Here, we developed HEAnet using a much larger data set (51 patients with TLE with 972 095 HEA examples) and applied deep learning algorithms to achieve an AUC ROC of 0.89 and an AUC PR of 0.39 at the single-event level. HEAnet's performance generalized well on 2 external data sets, where its output accurately distinguished patients with TLE from HCs, provided lateralizing information, and monitored changes in HEA in response to changes in ASMs. HEAnet has several applications that could fill important gaps in the diagnosis and treatment of TLE. First, HEAnet can improve the sensitivity of scalp EEG for diagnosing TLE, particularly when clinical interpretation is normal or questionably abnormal. In data set 2, 13% of patients with TLE (3 of 24) were successfully diagnosed by HEAnet but missed by at least 1 expert on scalp EEG review. Combining expert review with HEAnet increased the sensitivity of diagnosing TLE from 0.50 to 0.58 for human experts, to 0.63 to 0.67, without reducing specificity. Second, a common disabling feature of TLE is cognitive impairment.[23] The ability to noninvasively monitor HEA would allow clinicians to assess whether cognitive impairment in a patient with TLE is related to HEA and to assess whether reducing HEA with medications improves their cognitive function. Third, in patients with TLE undergoing evaluation for epilepsy surgery, HEAnet could serve as a complementary, noninvasive biomarker, independent of scalp EEG epileptiform discharges, to guide surgical decision-making.

Limitations

Our study had several limitations. First, HEAnet was specifically trained to detect HEA occurring during sleep. Nevertheless, HEAnet performed reasonably well on awake EEG data, maintaining excellent specificity, though with reduced sensitivity. We also found that 1-hour of sleep EEG, achievable in the outpatient setting, was sufficient for HEAnet to improve the diagnostic sensitivity of scalp EEG beyond that of human expert performance. Second, our study used FO recordings as the ground truth for HEA. As such, HEAnet might detect epileptiform activity arising not only from the hippocampus but also from other temporal lobe and extratemporal regions (eg, amygdala, parahippocampal gyrus, orbitofrontal lobe, insula). Future studies will better define the spatial specificity of HEAnet. Third, HEAnet used a simple CNN architecture with raw scalp EEG as input. Addition of hand-crafted features (eg, time-frequency,[22] connectivity,[10] zero-crossing patterns[24]), and more complex architectures[21,25] will be explored in future work. Fourth, although we validated HEAnet's performance and prediction thresholds using 2 external data sets, both data sets were relatively small and lacked ground truth information on HEA. Validation on larger data sets that include intracranial electrodes and that span a wider range of ages and diseases

will better define appropriate-use cases for HEAnet. Finally, a unique concern in applying HEAnet to clinical data is that visual confirmation of HEAnet's detections (scalp-negative spikes) is not possible; as such, operating HEAnet at thresholds that minimize false-positive detections and setting thresholds for clinical decision-making that are significantly higher than HEAnet's false positive detection rate will be essential for minimizing unnecessary diagnostic tests and treatments that could result from application of HEAnet.

Conclusions

Results of this diagnostic study suggest that HEAnet provides a novel and noninvasive biomarker of hippocampal hyperexcitability on scalp EEG. Although most machine learning applications in medicine aim to match human expert-level performance for a chosen problem, HEAnet performs a task that human experts cannot.

ARTICLE INFORMATION

Accepted for Publication: March 9, 2022.

Published Online: May 2, 2022. doi:10.1001/jamaneurol.2022.0888

Author Contributions: Mr Abou Jaoude and Dr Lam had full access to all of the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.
Concept and design: Jing, Cash, Lam.
Acquisition, analysis, or interpretation of data: Abou Jaoude, Jacobs, Sarkis, Pellerin, Cole, Cash, Westover, Lam.
Drafting of the manuscript: Abou Jaoude, Lam.
Critical revision of the manuscript for important intellectual content: All authors.
Statistical analysis: Abou Jaoude, Jing, Westover, Lam.
Obtained funding: Cole, Lam.
Administrative, technical, or material support: Abou Jaoude, Jacobs, Pellerin, Cole, Cash, Westover, Lam.
Supervision: Cole, Cash, Westover, Lam.

Conflict of Interest Disclosures: Mr Abou Jaoude reported being currently employed by InteraXon Inc, but neither this work nor this manuscript was performed in collaboration with or funded by InteraXon Inc. Dr Jacobs reported receiving grants from Harvard Catalyst and performing medical monitoring for Pfizer for an unrelated clinical trial outside the submitted work. Dr Sarkis reported receiving grants from the National Institutes of Health and Biogen outside the submitted work. Dr Cash reported receiving grants from the National Institutes of Health, being a founder of Beacon Biosignals, and having a patent for Detection of Inter-Ictal Discharges pending. Dr Westover reported receiving grants from the National Institutes of Health and being a cofounder of Beacon Biosignals. Dr Lam reported receiving grants from the National Institutes of Health, the American Academy of Neurology Institute, and Sage Therapeutics; and personal fees from Sage Therapeutics, Neurona Therapeutics, and Cognito Therapeutics outside the submitted work. No other disclosures were reported.

Funding/Support: The study was supported in part by grant R01NS062092 (Dr Cash), R01NS102190 (Dr Westover), R01NS107291 (Dr Westover), and K23NS101037 (Dr Lam) from the National Institutes of Health, and by the American Academy of Neurology Institute (Dr Lam).

Role of the Funder/Sponsor: The funders 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. Horak PC, Meisenhelter S, Song Y, et al. Interictal epileptiform discharges impair word recall in multiple brain areas. *Epilepsia*. 2017;58(3):373-380. doi:10.1111/epi.13633

2. Kleen JK, Scott RC, Holmes GL, et al. Hippocampal interictal epileptiform activity disrupts cognition in humans. *Neurology*. 2013;81(1):18-24. doi:10.1212/WNL.0b013e318297ee50

3. Inui K, Motomura E, Okushima R, Kaige H, Inoue K, Nomura J. Electroencephalographic findings in patients with DSM-IV mood disorder, schizophrenia, and other psychotic disorders. *Biol Psychiatry*. 1998;43(1):69-75. doi:10.1016/S0006-3223(97)00224-2

4. Lambert I, Tramoni-Negre E, Lagarde S, et al. Accelerated long-term forgetting in focal epilepsy: do interictal spikes during sleep matter? *Epilepsia*. 2021;62(3):563-569. doi:10.1111/epi.16823

5. Lam AD, Deck G, Goldman A, Eskandar EN, Noebels J, Cole AJ. Silent hippocampal seizures and spikes identified by foramen ovale electrodes in Alzheimer disease. *Nat Med*. 2017;23(6):678-680. doi:10.1038/nm.4330

6. Nayak D, Valentín A, Alarcón G, et al. Characteristics of scalp electrical fields associated with deep medial temporal epileptiform discharges. *Clin Neurophysiol*. 2004;115(6):1423-1435. doi:10.1016/j.clinph.2004.01.009

7. Koessler L, Cecchin T, Colnat-Coulbois S, et al. Catching the invisible: mesial temporal source contribution to simultaneous EEG and SEEG recordings. *Brain Topogr*. 2015;28(1):5-20. doi:10.1007/s10548-014-0417-z

8. Tao JX, Ray A, Hawes-Ebersole S, Ebersole JS. Intracranial EEG substrates of scalp EEG interictal spikes. *Epilepsia*. 2005;46(5):669-676. doi:10.1111/j.1528-1167.2005.11404.x

9. Fernández Torre JL, Alarcón G, Binnie CD, Polkey CE. Comparison of sphenoidal, foramen ovale and anterior temporal placements for detecting interictal epileptiform discharges in presurgical assessment for temporal lobe epilepsy. *Clin Neurophysiol*. 1999;110(5):895-904. doi:10.1016/S1388-2457(99)00039-5

10. Maharathi B, Patton J, Serafini A, Slavin K, Loeb JA. Highly consistent temporal lobe interictal spike networks revealed from foramen ovale electrodes. *Clin Neurophysiol*. 2021;132(9):2065-2074. doi:10.1016/j.clinph.2021.06.013

11. Chauvel P, Gonzalez-Martinez J, Bulacio J. Presurgical intracranial investigations in epilepsy surgery. *Handb Clin Neurol*. 2019;161:45-71. doi:10.1016/B978-0-444-64142-7.00040-0

12. Kovac S, Vakharia VN, Scott C, Diehl B. Invasive epilepsy surgery evaluation. *Seizure*. 2017;44:125-136. doi:10.1016/j.seizure.2016.10.016

13. Wieser HG, Elger CE, Stodieck SRG. The foramen ovale electrode: a new recording method for the preoperative evaluation of patients suffering from mesiobasal temporal lobe epilepsy. *Electroencephalogr Clin Neurophysiol*. 1985;61(4):314-322. doi:10.1016/0013-4694(85)91098-3

14. Sheth SA, Aronson JP, Shafi MM, et al. Utility of foramen ovale electrodes in mesial temporal lobe epilepsy. *Epilepsia*. 2014;55(5):713-724. doi:10.1111/epi.12571

15. Goncharova II, Alkawadri R, Gaspard N, et al. The relationship between seizures, interictal spikes and antiepileptic drugs. *Clin Neurophysiol*. 2016;127(9):3180-3186. doi:10.1016/j.clinph.2016.05.014

16. Malow BA, Lin X, Kushwaha R, Aldrich MS. Interictal spiking increases with sleep depth in temporal lobe epilepsy. *Epilepsia*. 1998;39(12):1309-1316. doi:10.1111/j.1528-1157.1998.tb01329.x

17. Goncharova II, Zaveri HP, Duckrow RB, Novotny EJ, Spencer SS. Spatial distribution of intracranially recorded spikes in medial and lateral temporal epilepsies. *Epilepsia*. 2009;50(12):2575-2585. doi:10.1111/j.1528-1167.2009.02258.x

18. Sammaritano M, Gigli GL, Gotman J. Interictal spiking during wakefulness and sleep and the localization of foci in temporal lobe epilepsy. *Neurology*. 1991;41(2 (pt 1)):290-297. doi:10.1212/WNL.41.2_Part_1.290

19. Abou Jaoude M, Sun H, Pellerin KR, et al. Expert-level automated sleep staging of long-term scalp electroencephalography recordings using deep learning. *Sleep*. 2020;43(11):zsaa112. doi:10.1093/sleep/zsaa112

20. Abou Jaoude M, Jing J, Sun H, et al. Detection of mesial temporal lobe epileptiform discharges on intracranial electrodes using deep learning. *Clin Neurophysiol*. 2020;131(1):133-141. doi:10.1016/j.clinph.2019.09.031

21. Jing J, Sun H, Kim JA, et al. Development of expert-level automated detection of epileptiform discharges during electroencephalogram interpretation. *JAMA Neurol*. 2020;77(1):103-108. doi:10.1001/jamaneurol.2019.3485

22. Spyrou L, Martín-Lopez D, Valentín A, Alarcón G, Sanei S. Detection of intracranial signatures of interictal epileptiform discharges from concurrent scalp EEG. *Int J Neural Syst*. 2016;26(4):1650016. doi:10.1142/S0129065716500167

23. Fisher RS, Vickrey BG, Gibson P, et al. The impact of epilepsy from the patient's perspective I. Descriptions and subjective perceptions. *Epilepsy Res*. 2000;41(1):39-51. doi:10.1016/S0920-1211(00)00126-1

24. Pyrzowski J, Le Douget J-E, Fouad A, Siemiński M, Jędrzejczak J, Le Van Quyen M. Zero-crossing patterns reveal subtle epileptiform discharges in the scalp EEG. *Sci Rep*. 2021;11(1):4128. doi:10.1038/s41598-021-83337-3

25. Hannun AY, Rajpurkar P, Haghpanahi M, et al. Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. *Nat Med*. 2019;25(1):65-69. doi:10.1038/s41591-018-0268-3