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. 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).

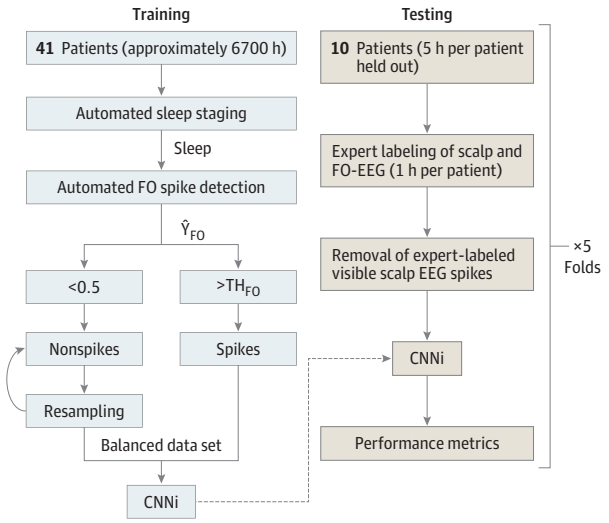
HEAnet Architecture

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).

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

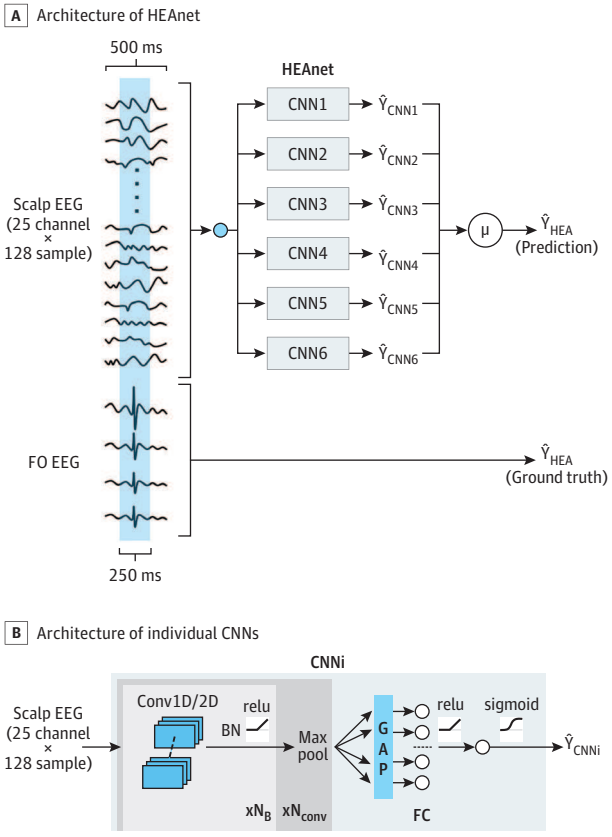


For each cross-validation fold, 10 patients are set aside for testing, while the rest of the patients are used for training. Training data were generated automatically from recordings of the training patients, using automated algorithms for sleep staging[19] and intracranial spike detection.[20] The 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.

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**).

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.

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, 0.26-0.92) 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 that occurs. Rather, for a given patient, the ability to determine whether HEA occurs at all, assess the laterality of HEA,

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.

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).

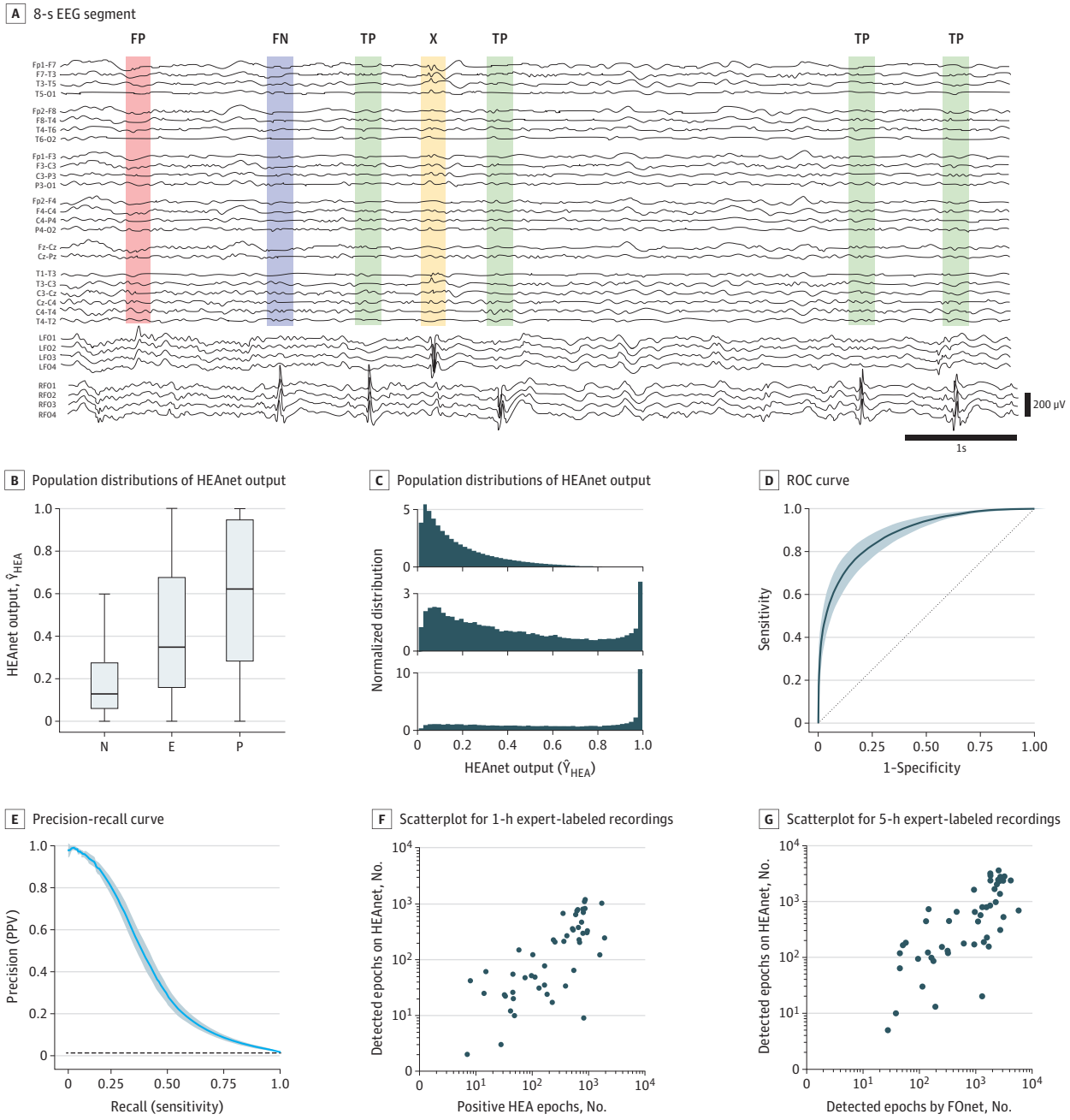
Performance of HEAnet Across Sleep and Awake States

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 move-

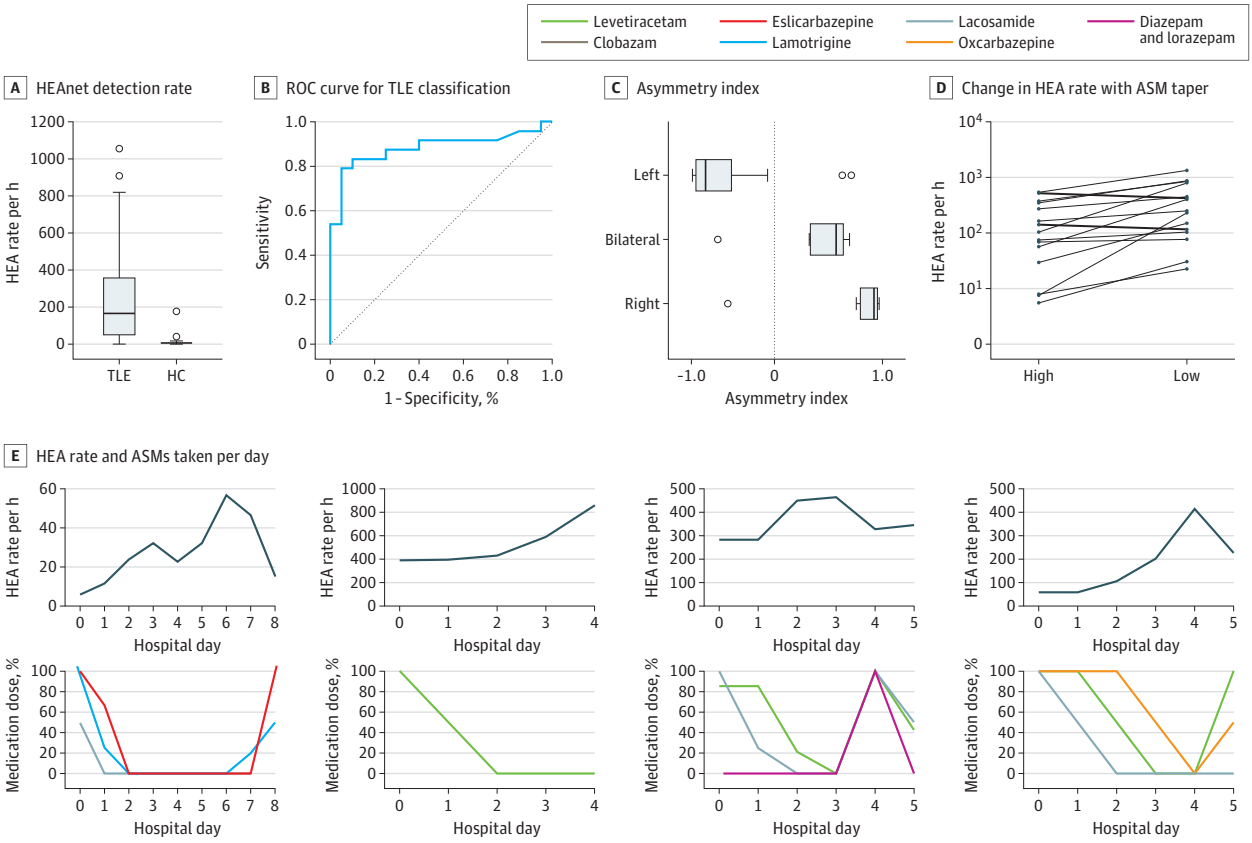
Figure 3. HEAnet Performance at the Single-Event Level



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.

ment [NREM] sleep) was significantly higher for patients with TLE than for HCs (268 [300] vs 15 [39], 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

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.

the left or right exceeded a certain threshold, which we varied to generate the plot. HEAnet distinguished 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

Supplementary Online Content

Abou Jaoude M, Jacobs CS, Sarkis RA, et al. Noninvasive detection of hippocampal epileptiform activity on scalp electroencephalogram. *JAMA Neurol*. Published online May 2, 2022. doi:10.1001/jamaneurol.2022.0888

This supplementary material has been provided by the authors to give readers additional information about their work.

eMETHODS.

EEG acquisition

EEG recordings were acquired using XLTEK hardware (Natus Medical Inc., Pleasanton CA) with data sampled at 1024 Hz. Scalp electrodes were placed using the International 10–20 system with anterior temporal electrodes (T1, T2). Four-contact FO electrodes (Ad-Tech, Racine, WI) were placed bilaterally, as described previously[1,2].

Scalp EEG pre-processing

Scalp EEG recordings were down-sampled to 256 Hz, bandpass filtered from 0.5 to 70 Hz with a Butterworth third order filter, and notch filtered at 60 Hz with a Butterworth fourth-order filter. A longitudinal bipolar montage with a coronal ring was used, resulting in 25 scalp EEG bipolar channels. Each bipolar channel was normalized to zero-median, unit interquartile range.

We applied automated artifact rejection to remove noisy segments from all recordings. For FO recordings, artifact rejection was performed as previously described[3]. For scalp EEG recordings, artifacts were identified as follows: Scalp EEG recordings were divided into 1-second non-overlapping epochs. For each epoch, the maximum amplitude across all channels (amp_{max}) was calculated. For each patient, all epochs with amp_{max} greater than 3 standard deviations above the mean amp_{max} for that patient were considered artifactual and removed from further analysis.

Automated sleep staging

Automated sleep staging was performed on all scalp EEG recordings, using a publicly available deep learning algorithm that we previously developed for use on long-term scalp EEG recordings, and that performs at the level of human sleep experts (https://github.com/mauriceaj/CRNNeeg-sleep)[4]. We extracted all sleep epochs (N1, N2, N3, and REM) from each recording for analysis and discarded awake portions of the recordings.

Automated generation of a training dataset of HEA examples on scalp EEG

FO electrodes record HEA with high fidelity and thus provide the ground truth for labeling HEA in Dataset #1. To automatically generate a massive dataset of HEA examples, we used a deep learning algorithm (referred to here as FOnet) that accurately detects HEA from the FO recordings[3]. FOnet was independently applied to the left and right FO recordings from Dataset #1, to automatically label positive and negative HEA examples from the left and right hippocampi, respectively. All 250ms epochs in which the output of FOnet exceeded a threshold, th_{FO}, were labeled as *positive HEA training examples*. We tested three th_{FO} values (0.870, 0.923, and 0.966), which corresponded to a positive predictive value (PPV) of HEA detection of 0.7, 0.8, and 0.9, respectively. Epochs in which FOnet's output was less than 0.5 were labelled as negative HEA epochs. We extracted ~ 2.12 million, 1.58 million, and 0.97 million positive HEA training examples for PPVs of 0.7, 0.8, and 0.9, respectively. We extracted 18.9 million negative HEA examples. We used the temporal information from the positive and negative HEA examples to extract the corresponding scalp EEG data for training.

As epileptiform discharges are relatively infrequent events, the number of negative HEA examples greatly exceeded the number of positive HEA examples. To create a balanced training dataset, we chose the number of negative HEA examples to match the number of positive HEA examples. Negative HEA examples for the training dataset were randomly sampled from all negative HEA epochs and stratified uniformly across all training patients and across left and right sides. To ensure that negative HEA examples were truly negative, we excluded negative HEA examples from FO recordings in which HEA comprised >20% of all epochs on a given hemisphere, or in which HEA on both hemispheres additively comprised > 30% of all epochs.

Training HEAnet to detect the laterality of HEA

We chose to develop a single HEA detection algorithm that is agnostic to laterality and that can be applied to independently detect HEA from either the left or right hippocampus. This approach, which we employed previously[5,6], assumes that the scalp EEG signatures for left and right HEA are similar, but simple mirror images of one another. We designated the left side as the default and kept the scalp EEG data corresponding to all left-sided HEA training examples intact. For all right-sided HEA training examples, we performed a right-left reflection of the scalp EEG data. For example, we swapped the labels for Fp2-F8 (right side) and Fp1-F7 (its left-sided correlate) and performed similar right-left label swaps for all channels on the scalp EEG, so that all right-sided HEA examples now appeared as left-sided HEA. As such, all scalp EEG examples of HEA in the training dataset appear to arise from the left. The HEA detection algorithm is thus trained to detect left-sided HEA as a default but can also be used to detect right-sided HEA. To detect left and right-sided HEA, HEAnet processes each scalp EEG input segment twice, first using the original scalp EEG data (to detect left-sided HEA), and then using right-left reflected scalp EEG data (to detect right-sided HEA)

Expert annotation of testing dataset for HEAnet

5 hours of asleep EEG recording from each patient in Dataset #1 were held out for testing algorithm performance. 1 hour of the held-out data from each patient was independently annotated by two board-certified epileptologists (ADL, CSJ), using a custom-made graphical user interface, which allowed the experts to view the FO and scalp EEG data, switch between different montages (longitudinal bipolar, referential, and average), and adjust gain and filter settings as they would for clinical EEG interpretation. Expert annotation was performed in two stages. In the first stage, the epileptologists reviewed only scalp EEG data (blinded to FO recordings) and marked all epileptiform discharges visible on scalp EEG. In the second stage, the experts viewed the combined FO and scalp EEG recordings and annotated all HEA visible on the FO recordings.

Testing examples were generated for each patient by dividing the 1-hour, expert-annotated recording into non-overlapping 250ms epochs. Epochs were labeled as positive, equivocal, or negative HEA examples if HEA was annotated within the epoch by both, either, or none of the experts, respectively. All epochs in which either expert annotated a visible scalp EEG epileptiform discharge were discarded from the testing dataset, to ensure that HEAnet did not simply learn to detect visible epileptiform discharges on scalp EEG, and to provide a rigorous evaluation of HEAnet's performance in detecting HEA that cannot already be easily detected by experts.

Altogether, the experts labeled a total of 22,762 positive HEA examples and 26,770 equivocal HEA examples. Cohen's kappa for inter-rater agreement between experts was 0.62, indicating moderate agreement. Each patient had, on average, 455 ± 448 positive HEA examples and 535 ± 516 equivocal HEA examples. 12% of positive HEA examples, and 10% of equivocal HEA examples coincided with an annotation of a visible epileptiform discharge on scalp EEG and were removed from the testing dataset. 29.4% of annotations of visible epileptiform discharges on scalp EEG did not coincide with a positive or equivocal HEA example on FO electrodes. We examined the performance of FOnet on the 1-hour held-out recordings, using the expert annotations as ground truth, and found that FOnet had a sensitivity of 0.75 and PPV of 0.996.

The two experts also annotated 1 hour of asleep scalp EEG recording for each patient from Dataset #2, using the same interface as above, and blinded to diagnosis. They annotated all epileptiform discharges visible on the scalp EEG, and at the end of each recording, they determined whether the EEG was diagnostic for TLE or not.

CNN training and cross-validation (detailed)

We initially trained individual CNN models to detect HEA, based on a scalp EEG input. The input to each CNN is a 25 x 128 matrix that represents a 500ms segment (128 samples) of pre-processed scalp EEG data from 25 bipolar channels, ordered as: T3-C3, C3-Cz, F3-C3, C3-P3, P3-O1, Fp1-F7, F7-T3, T3-T5, T5-O1, T1-T3, Fp1-F3, T1-T2, Fz-Cz, Cz-Pz, T4-C4, C4-Cz, F4-C4, C4-P4, P4-O2, Fp2-F8, F8-T4, T4-T6, T6-O2, T2-T4, Fp2-F4. The output of each CNN is the probability that HEA occurred in the central 250ms of the scalp EEG input.

We used 5-fold cross validation, in which patients were divided into 5 groups of 10 patients each, and data from each patient was used for either training or testing, depending on the fold. Patients were stratified into groups such that the number of definite and equivocal HEA testing examples (from the 1-hour expert-labelled recordings) was similar across groups.

For each fold, one group of patients was designated the testing set, one group was designated the early-stopping set (ES), and the remaining three groups were pooled to form the optimization set (Opt). Each group was used at most once for the testing set, and at most once for the ES set. One patient's data was used only in the Opt set, as the 1-hour expert annotated recording for this patient did not contain any definite HEA examples. Data from the Opt set was comprised of automatically extracted HEA training examples, pooled across all patients in the Opt set. Data from the ES and testing sets was comprised of HEA testing examples from the 1-hour expert-labelled recordings, pooled across all patients in the ES or testing sets, respectively.

CNNs were trained using the Adaptive Moment Estimation (Adam) optimization algorithm with a batch-size of b_s, and parameters b_1, b_2 and e set to 0.9, 0.999, and 10^{-8}, respectively. We used the log (cross-entropy) loss function. A training cycle was defined as a complete iteration over all training examples in the Opt set. Negative HEA examples for training were randomly re-sampled after every C_R training cycles. Positive HEA examples were shifted by a randomly chosen time jitter of $s \in [-125, 125]$ ms every J training cycles.

After each training cycle, we calculated the performance of the CNN model on the ES set, based on the area under the precision-recall curve (AUC_{PR}). We used a counter (ES_{count}) which was incremented each time the AUC_{PR} for the ES set did not improve for 5 consecutive training cycles. Each time ES_{count} reached a new multiple of 5, the learning rate was reduced by 50%. Training was stopped when ES_{count} reached 25. The optimal model weights were taken from the training cycle with the highest corresponding AUC_{PR} on the ES set.

We also trained CNNs without early stopping. For each fold, 1 group of patients was designated the testing set, and the remaining four groups were pooled to form the Opt set. Each model was trained for a fixed number of training cycles. Every C_L training cycles, the learning rate was reduced by L_{Red}. The optimal model weights were taken from the final training cycle.

After training was complete, the performance of the optimal model was evaluated on the testing set. After repeating the training procedure for all folds, the performance metrics computed on each testing set were averaged across folds, to yield a final estimation of model performance. We repeated the cross-validation procedure to test multiple CNN models, each with different combinations of hyperparameters. The hyperparameters and range of values evaluated are shown in eTable 5.

CNN architecture

The Keras library running on top of Tensorflow was used to configure the CNN models, which were trained on two CUDA-enabled NVIDIA GPUs, running on CentOS 7. The architecture of each CNN is shown in Figure 2B. For each CNN, the scalp EEG input is passed through a sequence of N_{conv} convolutional layers, each consisting of N_B convolutional blocks. Each convolutional block of convolutional layer i contains f_i filters $f_i = i \times f_0$, where f_0 is a hyperparameter corresponding to the number of filters in the first convolutional layer. We tested both 1D and 2D convolutions. For 1D convolutions, each filter slides across the temporal dimension only and is of size $C_i \times K_{i,1D}$, where C_i is the channel dimension and $K_{i,1D}$ is a hyperparameter. For 2D convolutions, each filter slides across time and channels, and is of size $K_{i,2D} \times K_{i,2D}$, where $K_{i,2D}$ is a hyperparameter. After hyperparameter optimization, $K_{i,1D}$ and $K_{i,2D}$ were set as in eTable 4 for all CNNs in HEAnet. Following the N_B blocks is a max-pooling layer of size m_i and stride = 1, and a dropout regularization step[7]. Following all N_{conv} convolutional layers is a fully connected layer of size 1024, followed by a logistic regression unit for final classification. We use the rectified linear unit (relu) function[7] as the activation function for the convolutional and fully connected layers, and the sigmoid function as the activation function for the final output.

Performance metrics

To test each individual CNN model, we applied it to the scalp EEG data from the appropriate testing set and recorded all detections made by the model. Positive HEA examples were classified as True Positive (TP) detections if a detection was made within 250ms of the example, with the correct laterality. False Positive (FP) detections were defined as any detection that did not occur within 250ms of a positive or equivocal HEA example on the same side. True Negative (TN) detections were defined as equivocal or negative HEA examples on which no detections were made on the same side. Sensitivity was defined as the number of TP detections divided by the number of positive HEA examples. Precision, also known as positive predictive value (PPV), was defined as the number of TP detections divided by the sum of TP and FP detections. We defined specificity as the number of TN detections divided by the sum of negative HEA examples and equivocal HEA examples without detections.

HEAnet architecture

HEAnet is an ensemble of 6 CNNs, comprised of the top 3 performing 1D-CNNs and the top 3 performing 2D-CNNs based on AUC_{PR}. The parameters and individual performance of each CNN is shown in eTable 2. The output of HEAnet is the averaged output from all 6 CNNs. We found that an ensemble of 6 CNNs provided a balance between increasing the stability of the algorithm's output and setting a practical limitation on the computational requirements needed to apply the algorithm to longer recordings. The final HEAnet model was trained using the training dataset from all patients in Dataset #1.

Applying SpikeNet for detection of visible epileptiform discharges on scalp EEG

We used the previously published deep learning algorithm, SpikeNet[8], to automatically detect visible epileptiform discharges on scalp EEG. The purpose was to exclude these epochs from datasets used to test the performance of HEAnet, to ensure that HEAnet was not simply detecting epileptiform discharges that were easily visible on the scalp EEG, and to provide a rigorous demonstration of HEAnet's ability to detect HEA that is not readily detected by humans or spike detection algorithms. Specifically, we used SpikeNet to (a) automatically label 4 hours of testing data (scalp EEG) that had not undergone expert annotation, for each patient in Dataset #1; and (b) automatically label scalp EEG recordings from each patient in Dataset #2.

SpikeNet takes as an input a 1-second segment of scalp EEG. The raw EEG is passed through a high-pass filter with 0.5Hz cutoff and a notch filter at 60Hz. The data is downsampled to 128Hz. The input data is formatted as a concatenation of longitudinal bipolar channels (18) and common average reference (19), for a total of 37 channels. SpikeNet is applied on moving 1s windows with step size = 8 samples (62.5ms), and a detection is made when the output of SpikeNet is > 0.42 (threshold that minimized the calibration error in the training set of SpikeNet). Consecutive detections less than 8 samples apart are clustered together into a single event. Finally, events that have a duration smaller than 8 samples are discarded. We compared SpikeNet detections with expert annotations of the 1-hour held-out recordings from Dataset #1 and found that at a threshold of 0.42, SpikeNet had a PPV of 0.54 and a sensitivity of 0.48. All 250ms epochs containing spike detections by SpikeNet were removed from further analysis.

Assessing scalp EEG signatures of HEA, as detected by HEAnet

We set the detection threshold of HEAnet such that PPV ~ 0.8 and sensitivity was ~ 20% and applied it to the testing data from Dataset #1. For each true positive epoch, a 250ms window centered on the HEA peak on FO electrodes (time at peak absolute amplitude on the FO channels) was extracted for both the FO electrodes and scalp EEG. These windows were averaged across all true positive epochs for each patient (intra-patient). The intra-patient averaged waveforms were then averaged across all patients (inter-patient). Patients with < 5 true positive epochs were excluded from this analysis. The same procedure was repeated for false negative epochs. To calculate the latency between the peak of the deflection on the scalp EEG and the peak of the HEA on the FO channels, we determined the time-point at maximum absolute amplitude on the scalp EEG, for each intra-patient averaged waveform. The latency was calculated as that point minus 125ms (the middle of the 250ms window, corresponding to the peak of HEA on the FO channels).

Calculation of Asymmetry Index

Asymmetry index was defined as (R-L) / (R+L), where R and L represent the number of right- and left-sided HEA detections, respectively. A positive asymmetry index indicates more right-sided detections, whereas a negative asymmetry index indicates more left-sided detections.

Applying HEAnet to Datasets #2 and #3

Scalp EEGs for Datasets #2 and #3 underwent the same scalp EEG pre-processing steps as described above. We applied the automated sleep staging algorithm to extract all sleep epochs for testing. We applied SpikeNet to automatically remove epochs with visible epileptiform discharges on scalp EEG from Dataset #2.

Dataset #3 came from an external institution, where a board-certified epileptologist selected TLE and HC patients for the dataset, based on our inclusion/exclusion criteria. As a stringent test of HEAnet's application, we remained blinded to the diagnoses of the patients in this dataset (including how many patients were in each diagnostic group) while performing the initial analysis, i.e., applying HEAnet to predict diagnosis (TLE vs HC) and TLE laterality. To more closely approximate real-world application, we did not use SpikeNet to remove scalp visible epileptiform discharges for this analysis.

eRESULTS.

Scalp EEG signature of HEA learned by HEAnet

To visualize the scalp EEG signature of HEA learned by HEAnet, we calculated the averaged scalp EEG signal across all true positive detections for each patient, synchronized on the HEA peak on FO recordings (eFigure 1A). The averaged scalp EEG signature of detected HEA was a low amplitude (19.1 ± 9.4 μV) negative deflection lasting 150-200ms. For left-sided HEA detections, the voltage field of the scalp EEG signature was broadly distributed, involving most left hemispheric channels and extending to the midline and right frontal and parasagittal channels; the right temporal channels showed the lowest amplitudes. The median latency between the HEA peak on FO electrodes and the peak of the scalp EEG deflection was 0 ms (IQR: [-3.91, 97.66] (n = 37).

We also calculated the averaged scalp EEG and FO electrode signatures across all false negative (missed) detections for each patient (eFigure 1B). The scalp EEG signature of missed HEA showed a lower amplitude deflection (10.5 ± 7.9 μV) than for detected HEA (Wilcoxon Signed Rank Statistic = 16.0, $p < 0.001$). The absolute HEA peak amplitude on FO electrodes was also lower for missed HEA (221 ± 184 μV, n = 6644) compared to detected HEA (241 ± 186μV, n = 3515) (Mann-Whitney $U > 1.0 \times 10^7$, $p < 0.001$). Similarly, the maximal slope for HEA on FO electrodes was significantly lower for missed HEA (189 ± 263 μV/ms, n = 6644) compared to detected HEA (244 ± 229 μV/ms, n = 3515) (Mann-Whitney $U > 9.7 \times 10^6$, $p < 0.001$).

Overall, we suspect that HEAnet detects most HEA through a volume-conducted signal, based on our findings above showing: 1) a median latency of 0ms between the HEA peak on FO electrodes and the peak of the averaged scalp EEG signature of detected HEA; 2) positive correlations between $\hat{y}_{HEA}$ and HEA amplitude and slope on FO electrodes (eFigure 2). It is possible that HEAnet may also detect some HEA propagated from the hippocampus to neocortex, as the IQR for the median latency above [-3.91, 97.66] was skewed in the positive direction.

Analysis of false positive detections by HEAnet

We visually analyzed a random subset of false positive (FP) detections made by HEAnet on the testing dataset for Dataset #1. For each patient, we randomly selected up to 10 FP detections for review; if a patient had fewer than 10 FP detections, we reviewed all FP detections from that patient. Overall, we reviewed a median of 6 FP detections per patient (IQR: [0, 10]), which comprised a total of 302 FP detections reviewed (~ 27% of all FP detections in the testing dataset, with HEAnet operating at a PPV of ~0.8).

Images of FP detections were presented in random order to a board-certified epileptologist, to identify whether specific background abnormalities, artifacts, or other features could be consistently identified as the "cause" for the FP detections. Each image contained 15 seconds of EEG recording, with the FP epoch centered and an indication of the side (left / right) that the FP was detected on (eFigure 3). Both the scalp EEG data (a full anterior-posterior longitudinal bipolar montage) and the foramen ovale electrode data were shown. The results of this analysis are shown in eTable 7, with representative examples in eFigure 3. Notably, we found that 43% of FP detections were actually true positive detections of HEA that had been missed on initial expert annotation. 28% of FPs had no obvious change from background, though in many cases, there were associated sleep spindles, or high-frequency low-amplitude backgrounds. 10% of FPs were associated with artifact on scalp EEG, most commonly low-amplitude "spiky" artifacts such as EKG, electrode pop artifact, or myogenic artifact; importantly, these types of artifact can be readily identified and corrected for.

HEAnet performance across awake and asleep states (from Dataset #1)

For each patient in Dataset #1, we selected 1-hour of clean awake recording (>99% awake, with < 5% of epochs flagged as artifact on the FO electrodes). We applied FOnet (operating at a threshold where PPV was 0.9, and sensitivity was 0.5) to the FO electrodes, to obtain a ground truth on when HEA occurred during these recordings. We then applied HEAnet (operating at a threshold where the PPV was 0.9 for sleep EEG recordings) to the scalp EEG of these awake recordings. We report cross-validated performance metrics in the awake state (eTable 8). At the single-event level, HEAnet applied to awake recordings had an AUC_{ROC} of 0.85 ± 0.02 and an AUC_{PR} of 0.19 ± 0.07 (for chance performance, we would expect an $AUC_{PR} = 0.008$, the frequency of HEA in the awake state). Using the same detection threshold for HEAnet that yielded a PPV of 0.9 during sleep, application of HEAnet during the awake state had a PPV of 0.81 ± 0.10, a specificity of 0.999 ± 0.00, and a sensitivity of 0.06 ± 0.02. Ultimately, this means that, while application of HEAnet to awake EEG will have a lower sensitivity for detecting HEA at the single event level compared to during sleep, we can at least expect that the number of false positive detections during the awake state will also remain low.

We also examined the performance of HEAnet in each different sleep stage (N1, N2, N3, and REM) in the testing dataset (eTable 7). There were no significant differences in HEAnet's performance across different sleep stages for AUC_{ROC} (one-way ANOVA, F=1.23, p=0.33) or AUC_{PR} (one-way ANOVA, F=0.89, p=0.47).

Tracking dynamics of HEA rate over time on Dataset #1

We evaluated whether, for a given individual in Dataset #1, HEAnet's output could be used to monitor changes in HEA rate over time. We divided each patient's 5-hour held-out recording into non-overlapping 30-minute epochs and calculated the number of HEAnet detections on scalp EEG for each 30-minute epoch. To estimate the ground truth HEA rate for each 30-minute epoch, we

calculated the number of FOnet detections on FO electrodes, that were not associated with a visible epileptiform discharge on scalp EEG. eFigure 4 shows the relationship between HEA rates predicted by HEAnet from scalp EEG, and the ground truth HEA rates based on FOnet detections from FO electrodes, for 4 representative patients. We calculated the correlation between HEAnet's predicted HEA rate, and the ground truth HEA rate, across all 30-minute epochs in the 5-hour recording for each patient. Averaging across all patients yielded a mean correlation coefficient of 0.62 ± 0.33 (Wilcoxon Signed test Statistic = 14.0, $p < 0.001$). 50% of patients had a correlation coefficient > 0.7. HEAnet's output thus has a sufficient dynamic range to track changes in HEA rate over time, for individual patients.

Laterality of HEAnet detections correlate with laterality of TLE on Dataset #1

Using the 1-hour, expert-annotated recordings, we calculated an asymmetry index for each patient, defined as (R-L) / (R+L), where R and L represent the number of right- and left-sided HEA detections, respectively. A positive asymmetry index indicates more right-sided detections, whereas a negative asymmetry index indicates more left-sided detections. We calculated an asymmetry index using HEAnet detections on scalp EEG ($ASym_{HEA}$), as well as an asymmetry index using expert annotations on FO electrodes ($ASym_{EXP}$). We found a moderate correlation between $ASym_{HEA}$ and $ASym_{EXP}$ ($R = 0.67$, $p < 0.001$, eFigure 5A). We next divided patients from Dataset #1 into 3 groups, based on whether they were clinically diagnosed as having left TLE, right TLE, or bi-temporal epilepsy. eFigure 5B shows the median $ASym_{HEA}$ for each group. There was a significant trend towards increasing median $ASym_{HEA}$, going from left TLE to bi-temporal TLE to right TLE ($p = 0.011$). 73% of patients with left TLE had a negative $ASym_{HEA}$, and 63% of patients with right TLE had a positive $ASym_{HEA}$.

Performance of HEAnet on Dataset #3

We applied HEAnet (operating at PPV 0.9) to Dataset #3 (from an external institution, BWH), and calculated the HEA detection rate (number of HEAnet detections per hour NREM sleep) for each patient (eTable 3). The HEA detection rate was significantly higher for TLE than controls (99 ± 109 vs. 13 ± 45, respectively; Mann-Whitney $U = 30$, $p < 0.001$; eFigure 6A). The HEAnet detection rate distinguished TLE from controls with an AUC_{ROC} of 0.95 (eFigure 6B). We then calculated the asymmetry index of HEAnet detections ($ASym_{HEA}$) for all TLE patients (eTable 3). 100% of left TLE patients had a negative $ASym_{HEA}$, and 83% of right TLE patients had a positive $ASym_{HEA}$ (eFigure 6C, eTable 3). Among TLE patients, $ASym_{HEA} < -0.75$ had a PPV of 0.91 for predicting left TLE, while an $ASym_{HEA} > +0.75$ had a PPV of 1 for predicting right TLE. Thus, in TLE patients where the magnitude of $ASym_{HEA}$ is high (59% of TLE patients in Dataset #3 had an $|ASym_{HEA}| > 0.75$), HEAnet accurately lateralized the seizure onset zone.

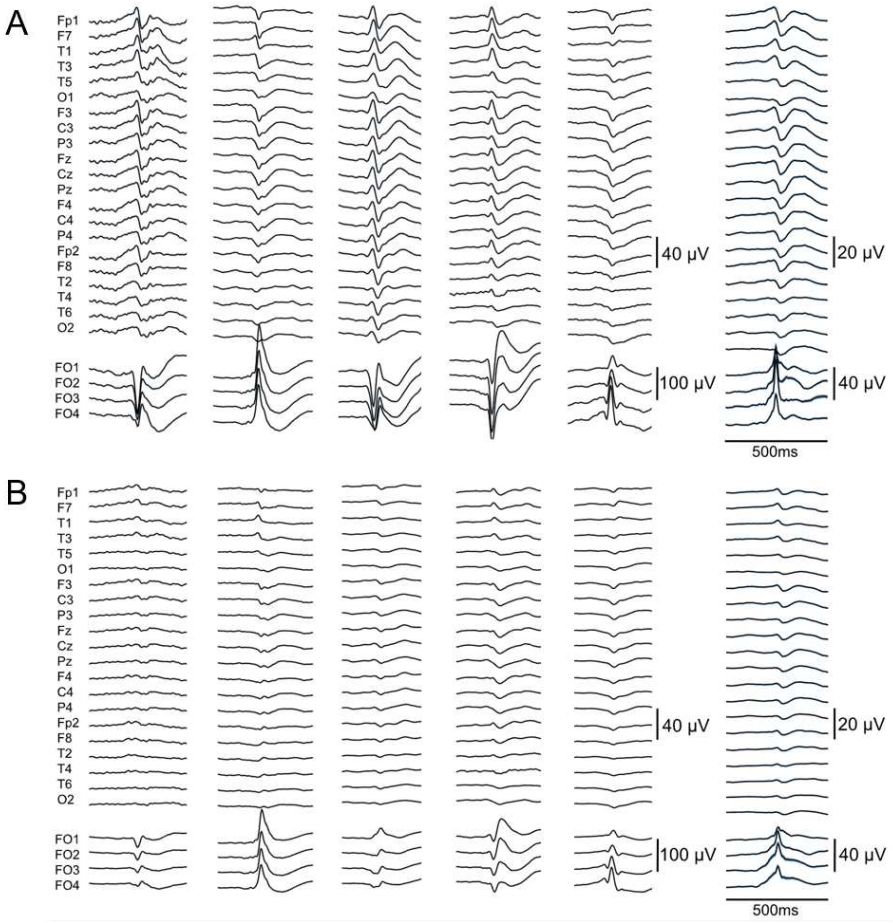
Further analysis of HEAnet's performance in specific patients in Dataset #3 yielded additional insights. As described in the main manuscript, we predicted a diagnosis of TLE based on a threshold that we defined based on Dataset #2 (HEAnet detection rate > 33 per hour NREM sleep, on either the left or right side). On Dataset #3, HEAnet missed a diagnosis on a L TLE patient who had an HEAnet detection rate of 32.5 per hour NREM sleep on the left side, with an $ASym_{HEA} = -0.97$. Similarly, HEAnet missed a diagnosis on a B TLE patient, who had HEAnet detection rates of 29.4 and 25.8 on the left and right sides, respectively, and an $ASym_{HEA} = -0.06$. Thus, while the threshold of 33 detections per hour works reasonably well, further fine-tuning of this threshold may optimize performance. For example, reducing this threshold to 29 detections per hour would have increased sensitivity of diagnosing TLE from 68% to 77%, without changing the specificity of detection.

HEAnet incorrectly diagnosed 1 HC patient as having TLE, with an HEAnet detection rate of 206.3 per hour on the right side, significantly higher than the detection rates in all other HC. Visual inspection of these detections revealed that the vast majority of these detections occurred synchronously with very low amplitude EKG artifact on the scalp EEG. Removing this patient from the analyses above, we found that the HEA detection rate in controls fell to 4 ± 8 per hour (compared to 13 ± 45, with this patient included).

eREFERENCES.

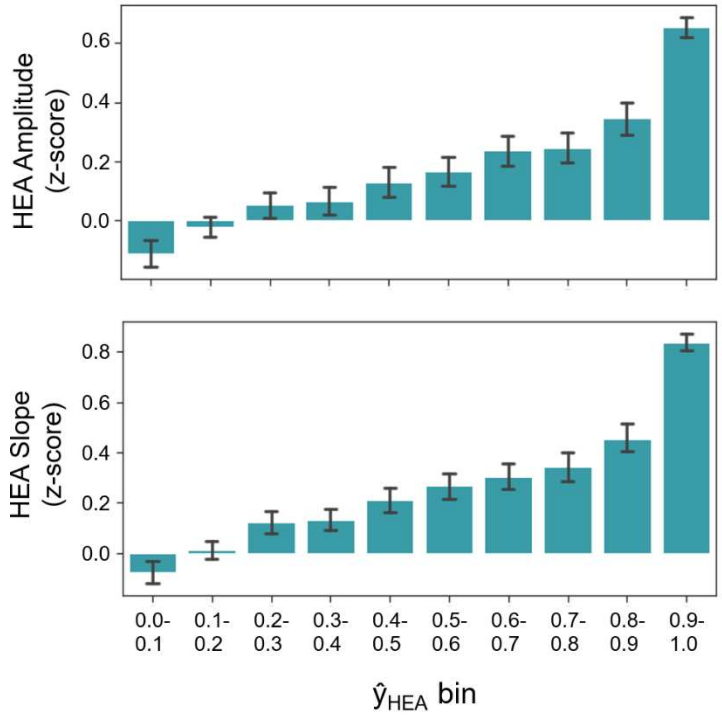
1. Sheth SA, Aronson JP, Shafi MM, et al. Utility of foramen ovale electrodes in mesial temporal lobe epilepsy. *Epilepsia*. 2014;55(5):713-724. doi:10.1111/epi.12571
2. Wieser HG, Elger CE, Stodieck SRG. *The "Foramen Ovale Electrode": A New Recording Method for the Preoperative Evaluation of Patients Suffering from Mesio-Basal Temporal Lobe Epilepsy*. Vol 61.; 1985.
3. 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
4. Abou Jaoude M, Sun H, Pellerin KR, et al. Expert-level automated sleep staging of long-term scalp EEG recordings using deep learning. *Sleep*. 2020. doi:10.1093/sleep/zsaa112
5. Lam AD, Zepeda R, Cole AJ, Cash SS. Widespread changes in network activity allow non-invasive detection of mesial temporal lobe seizures. *Brain*. 2016;139(10):2679-2693. doi:10.1093/brain/aww198
6. Lam AD, Maus D, Zafar SF, Cole AJ, Cash SS. SCOPE-mTL: A non-invasive tool for identifying and lateralizing mesial temporal lobe seizures prior to scalp EEG ictal onset. *Clin Neurophysiol*. 2017;128(9):1647-1655. doi:10.1016/j.clinph.2017.06.040
7. Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R. Dropout: A Simple Way to Prevent Neural Networks from Overfittin. *J Mach Learn Res*. 2014;15:1929-1958. doi:10.1016/0370-2693(93)90272-J
8. 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

eFigure 1. Scalp EEG Signatures of HEA Detected by HEAnet



eFigure 1. Scalp EEG signatures of HEA detected by HEAnet. (A) Averaged scalp EEG and FO signals from all correctly detected HEA epochs, from five representative patients (left), and on the far right, from the population-average across all patients (n=37; we excluded 12 patients with < 5 true positive detections and 2 patients with noisy FO recordings, from the average). Epochs were aligned by the peak of the intracranial spike. Examples were generated with HEAnet operating at a PPV of ~0.8. **(B)** Similar to A, but for all missed HEA epochs (based on a HEAnet detection threshold < 0.5). For (A) and (B), a C2 reference electrode was used, and all HEA occurring on the right side were left-right reflected.

eFigure 2. Relationship Between $\hat{Y}_{HEA}$ and Corresponding HEA Amplitude and Slope on FO Electrodes



eFigure 2. Relationship between $\hat{y}_{HEA}$ and corresponding HEA amplitude and slope on FO electrodes. For each positive HEA example in the testing dataset, we calculated the maximum amplitude and slope on FO electrodes. We then standardized HEA amplitudes and slopes to z-scores within each patient, to capture variability in these measures within a given patient, while also allowing comparison across patients, since the exact positioning of FO electrodes placement and recorded HEA characteristics can differ substantially between patients. Positive HEA examples from all patients (excluding 2 who had noisy FO electrode recordings) were binned by $\hat{y}_{HEA}$, and the corresponding z-scores for amplitude or slope were averaged across all examples in each bin. Bars represent means with 95% confidence interval based on bootstrapping.

eFigure 3. Examples of False-Positive Detections by HEAnet

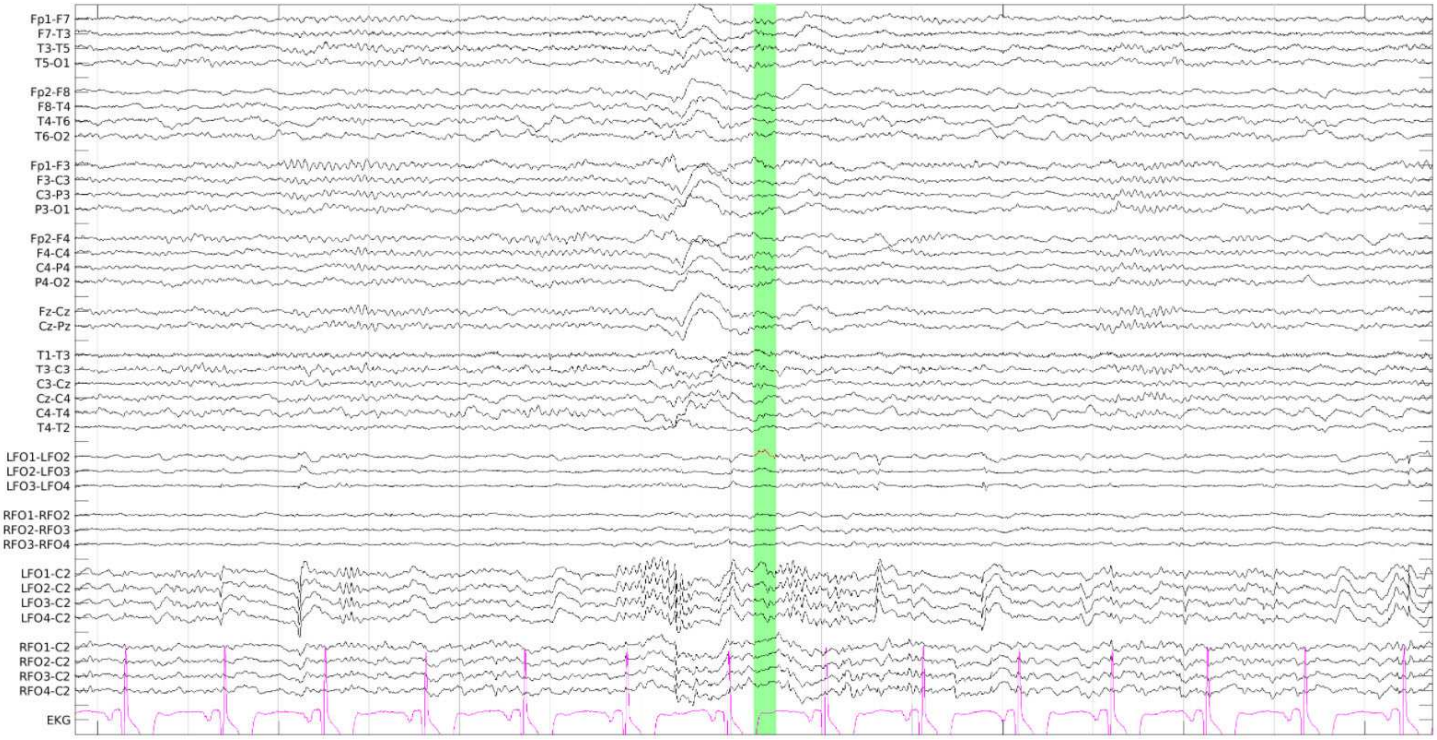
Each panel shows 15 seconds of EEG. The false positive detection epoch is highlighted in green, with the laterality of the detection being on the same side as the red trace shown in either LFO1-2 or RFO1-2. Vertical bars represent 1 second.
A. "False" false positive (true positive) on the RIGHT. Annotation of HEA on RFO electrode was likely missed by experts due to occurrence of HEA on LFO electrode at same time, and minimal deflection on the RFO bipolar montage.
B. "False" false positive (true positive) detection on the LEFT. There is probable HEA on the LFO electrode, given similar morphology though lower amplitude compared to two other examples of L HEA on the same page. Annotation was likely missed by experts due to the low amplitude, in addition to EKG artifact on FO electrodes.
C. False positive detection on the LEFT, likely related to a "forme fruste" epileptiform discharge visible on scalp EEG on the left that was missed on expert annotation.
D. False positive on the LEFT, possibly related to sleep spindles in the background.
E. False positive detection on the LEFT, possibly related to myogenic artifact in the background.
F. False positive detection on the RIGHT, possibly related to electrode pop artifact in the background.
G. False positive detection on the LEFT, possibly related to HEA on RFO electrode, with concurrent sleep spindles in the background.
H. False positive detection on the RIGHT, possibly due to high frequency, low amplitude activity on the right.

A. "False" false positive (true positive) on the RIGHT. Annotation of HEA on RFO electrode was likely missed by experts due to occurrence of HEA on LFO electrode at same time, and minimal deflection on the RFO bipolar montage.

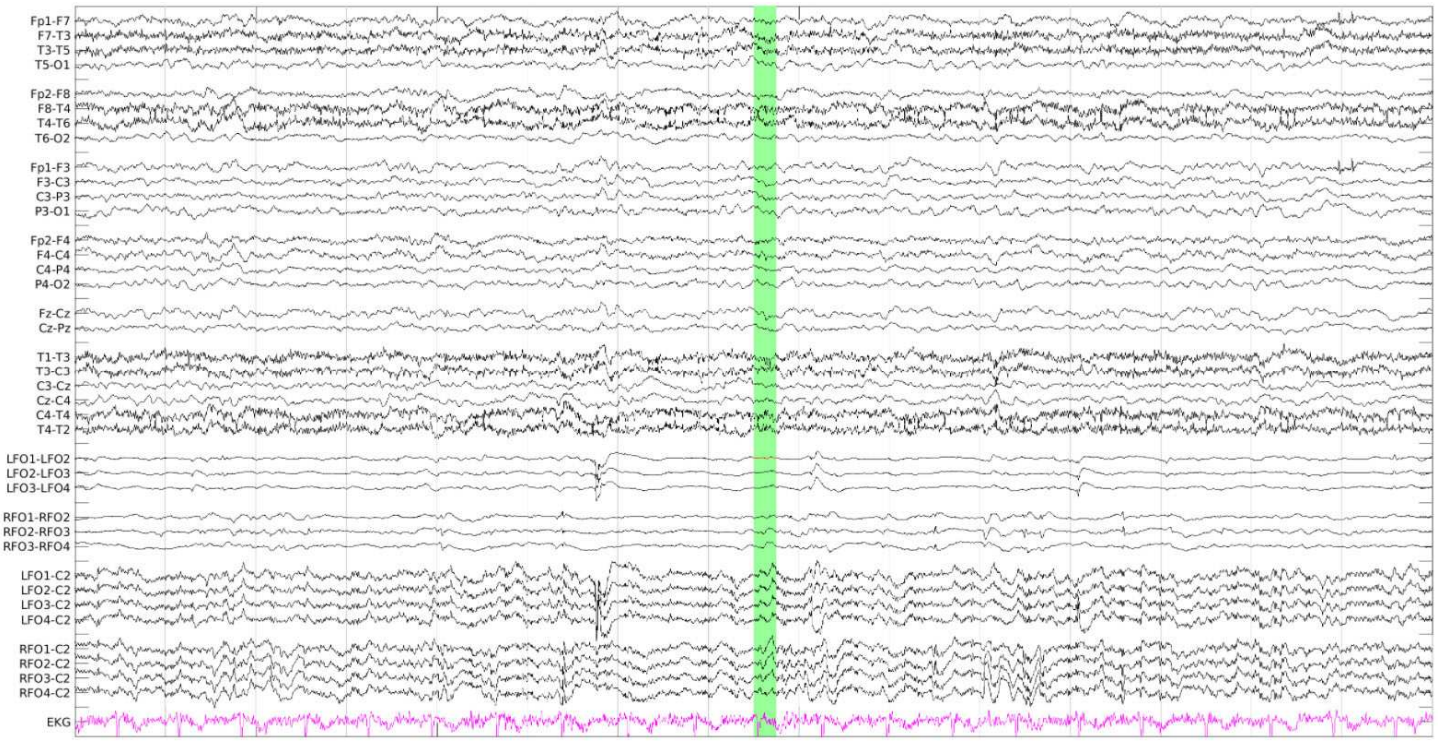
B. “False” false positive (true positive) detection on the LEFT. There is probable HEA on the LFO electrode, given similar morphology though lower amplitude compared to two other examples of L HEA on the same page. Annotation was likely missed by experts due to the low amplitude, in addition to background EKG artifact on FO electrodes.

C. False positive detection on the LEFT, likely related to a “forme fruste” epileptiform discharge visible on scalp EEG on the left that was not marked on expert annotation.

D. False positive on the LEFT, possibly related to sleep spindles in the background.



E. False positive detection on the LEFT, possibly related to myogenic artifact in the background.

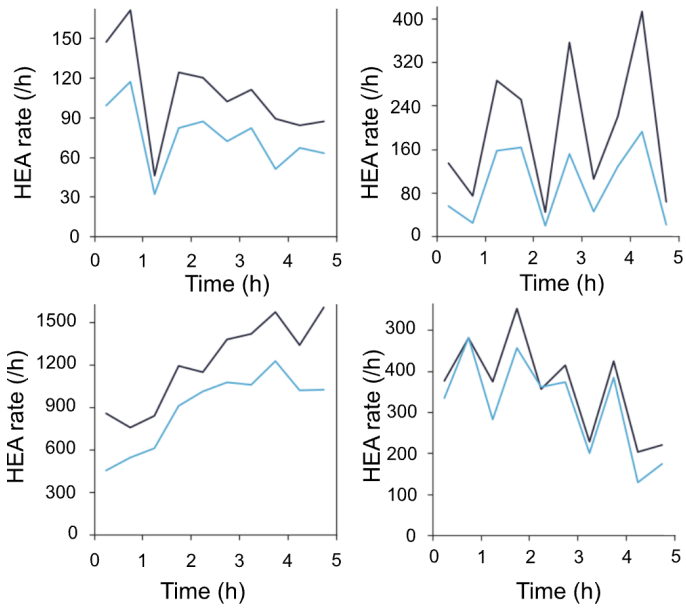


F. False positive detection on the RIGHT, possibly related to electrode pop artifact in the background.

G. False positive detection on the LEFT, possibly related to HEA on RFO electrode, with concurrent sleep spindles in the background.

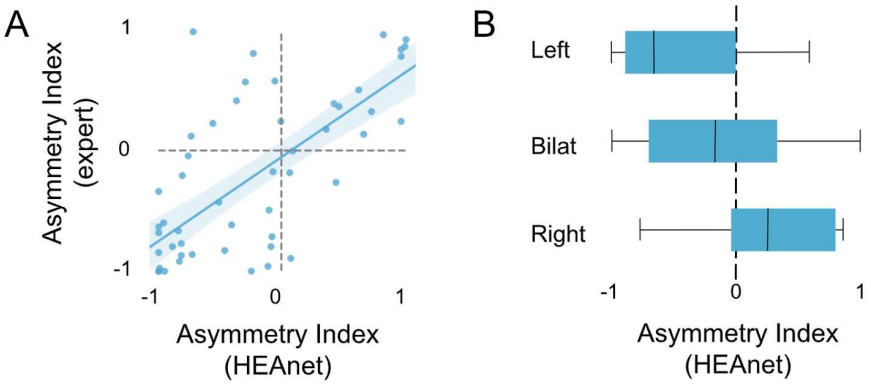
H. False positive detection on the RIGHT, possibly due to high frequency, low amplitude activity on the right.

eFigure 4. Tracking Changes in HEA Rate Over Time in Data Set 1



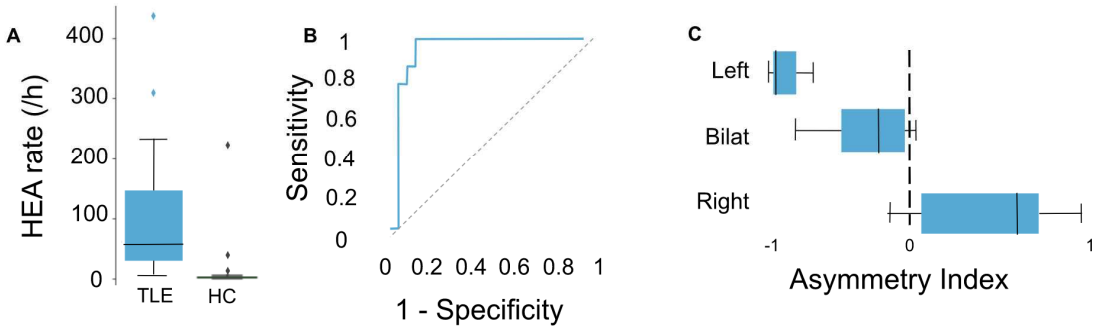
eFigure 4. Tracking changes in HEA rate over time. Data from four representative patients from Dataset #1, showing the computed HEA rate per hour, based on HEAnet detections (blue curve), and FOnet detections (black curve), over the 5h held-out recording. Detection rate was calculated across consecutive, non-overlapping 30-minute windows.

eFigure 5. Assessing Laterality of HEAnet Detections in Data Set 1



eFigure 5. Assessing laterality of HEAnet detections in Dataset #1. **(A)** Scatter plot showing correlation between asymmetry index as calculated based on expert-labeled HEA and based on HEAnet detections. Each dot represents a different patient. Shaded regions correspond to 95% confidence intervals estimated using 1000 bootstrap resamples. **(B)** Box-and-whiskers plots showing the asymmetry index calculated based on HEAnet detections, for patients with left TLE [L, n=15], bitemporal epilepsy [B, n=19], and right TLE [R, n=8]. 8 patients from Dataset #1 were classified as having indeterminate lateralization, as they did not have any seizures recorded. The vertical line at AI = 0 represents the point of symmetry. The threshold for HEAnet was set such that the PPV for HEA detection was ~0.7.

eFigure 6. Validation of HEAnet Performance on Data Set 3



eFigure 6. Validation of HEAnet Performance on Dataset #3. **(A)** HEAnet detection rate (combined left and right sided detections, per hour of NREM sleep), for TLE patients and non-epileptic healthy controls (HC). **(B)** ROC curve for classification of TLE versus HC based on HEAnet detection rate. **(C)** Asymmetry index calculated based on HEAnet detections, for patients with left TLE [n=12], bitemporal epilepsy (Bilat) [n=4], and right TLE [n=7].

ONLINE-ONLY TABLES

eTable 1. Clinical Details and HEAnet Performance for Data Set 1

						Expert-Annotated 1-hour Recordings (Held-Out Testing Data)					
						Expert Annotations				# HEAnet Detections without scalp visible spikes	
Age	Sex	Sz onset	Surgery	Surgical Pathology	Record Length (h)	# Positive HEA	# Scalp visible spikes	# Positive HEA without scalp visible spike	% Positive HEA without scalp visible spike	PPV 0.7	PPV 0.9
32	M	L	L ATL	hippocampal sclerosis	331.7	579	76	507	87.6	511	256
19	F	R	R ATL	dentate hilus gliosis	20.6	848	761	537	63.3	354	155
19	F	BL	none	n/a	146.4	101	8	100	99.0	97	25
26	F	L	L ATL	hippocampal sclerosis	68.8	938	96	901	96.1	261	101
53	M	L	declined surgery	n/a	74.6	748	185	666	89.0	310	117
39	F	R	R ATL	hippocampal gliosis	117.3	359	59	325	90.5	171	113
19	M	L	L ATL	hippocampal sclerosis	160.0	800	117	679	84.9	600	429
16	M	R	R ATL	hippocampal sclerosis	137.3	228	29	226	99.1	8	1
42	F	L	declined surgery	n/a	63.3	656	378	536	81.7	269	130
57	M	R	R ATL	hippocampal sclerosis	190.7	58	26	58	100.0	139	82
40	F	BL	Palliative R ATL	hippocampal sclerosis	146.8	793	21	788	99.4	245	95
46	F	BL	Palliative R ATL	hippocampal sclerosis	114.4	536	252	439	81.9	238	74
33	F	L	L ATL	hippocampal sclerosis	134.0	48	2	48	100.0	10	1
26	M	L	L ATL	hippocampal sclerosis	228.2	826	3	826	100.0	6	1

						Expert-Annotated 1-hour Recordings (Held-Out Testing Data)					
						Expert Annotations				# HEAnet Detections without scalp visible spikes	
Age	Sex	Seizure onset	Surgery	Surgical Pathology	Record Length (h)	# Positive HEA	# Scalp visible spikes	# Positive HEA without scalp visible spike	% Positive HEA without scalp visible spike	PPV 0.7	PPV 0.9
33	M	L	L ATL	sub-ependymal gliosis	47.4	75	5	74	98.7	43	14
65	M	BL	none	n/a	111.4	1896	101	1759	92.8	179	53
26	M	BL	RNS	n/a	329.0	626	217	573	91.5	538	221
53	M	BL	none	n/a	382.3	15	1	15	100.0	53	24
50	M	BL	none	n/a	65.6	933	94	846	90.7	204	65
51	F	R	none	n/a	184.3	33	10	33	100.0	18	10
37	F	BL	RNS	n/a	306.3	0 **	0	0	n/a	n/a	n/a
67	M	N/A*	declined surgery	n/a	107.7	524	51	489	93.3	278	144
35	M	BL	none	n/a	62.3	32	6	32	100.0	20	11
37	F	BL	RNS	n/a	135.8	679	243	554	81.6	128	63
36	M	L	RNS	n/a	88.6	41	7	38	92.7	10	4
23	F	R	R ATL	reactive gliosis	95.8	876	102	800	91.3	653	365
65	M	BL	none	n/a	150.4	695	97	614	88.3	133	47
20	M	BL	RNS	n/a	138.4	1571	34	1559	99.2	104	28
59	M	BL	Palliative R ATL	hippocampal sclerosis	184.8	871	433	569	65.3	620	309
30	M	BL	Palliative R ATL	hippocampal sclerosis	234.9	7	7	6	85.7	1	0
67	F	L	none	n/a	96.3	186	0	186	100.0	18	7
42	F	BL	Palliative L ATL	hippocampal sclerosis	97.9	1701	1202	1078	63.4	593	219
24	F	BL	none	n/a	203.5	547	63	541	98.9	48	7
53	F	L	none	n/a	124.2	162	0	162	100.0	21	0

						Expert-Annotated 1-hour Recordings (Held-Out Testing Data)					
						Expert Annotations				# HEAnet Detections without scalp visible spikes	
Age	Sex	Seizure onset	Surgery	Surgical Pathology	Record Length (h)	# Positive HEA	# Scalp visible spikes	# Positive HEA without scalp visible spike	% Positive HEA without scalp visible spike	PPV 0.7	PPV 0.9
63	F	N/A*	none	n/a	42.6	131	11	125	95.4	21	12
53	M	L	L ATL	focal cortical dysplasia	118.2	236	31	221	93.6	162	53
35	M	BL	declined surgery	n/a	281.6	803	48	763	95.0	578	318
49	F	R	declined surgery	n/a	357.2	111	19	111	100.0	45	25
24	M	N/A*	declined surgery	n/a	119.0	165	3	163	98.8	69	47
75	M	N/A*	none	n/a	92.4	46	2	46	100.0	18	6
48	F	L	none	n/a	139.0	353	195	264	74.8	371	191
58	M	N/A*	none	n/a	97.3	45	1	45	100.0	47	28
28	M	R	R ATL	hippocampal sclerosis	167.5	410	90	374	91.2	183	78
12	M	L	none	n/a	74.6	45	12	44	97.8	22	9
34	M	L	none	n/a	345.3	390	3	389	99.7	24	1
50	M	N/A*	none	n/a	502.5	14	0	14	100.0	15	4
54	M	BL	none	n/a	93.6	251	98	232	92.4	133	40
28	M	N/A*	none	n/a	113.0	8	0	8	100.0	32	6
62	F	N/A*	none	n/a	318.9	28	0	28	100.0	3	1
25	M	BL	L ATL	hippocampal sclerosis	191.9	642	347	554	86.3	435	180
38	F	BL	declined surgery	n/a	260.2	96	2	95	99.0	49	23

M=male; F=female; L=left; R=right; BL = bilateral; N/A = epilepsy lateralization unclear; ATL = anterior temporal lobectomy; RNS = responsive neural stimulation

* No seizures were captured for these patients.
** No definite HEA were annotated in 1-hour recording; this patient's data was only used in the training dataset.

eTable 2. Clinical Details and HEAnet Performance for Data Set 2

							Entire Recording		**1-hour Sleep Recording**					
Age	**Sex**	**Dx**	**IEDs**	**Sz Onset**	**MRI**	**FDG-PET[5]**	HEAnet Detection Rate (per hr)[6]	ASym HEA	HEAnet Detection Rate (per hr)[7]	Expert-Annotated Spike Rate (per hr)[8]	% HEAnet Detections coinciding with Expert Detections	TLE Diagnosis		
												HEA net	Ex 1	Ex 2
39	F	L TLE	LT	NC	L MTS	N/A	109.4	-0.82	67	34	17.3%	x	x	x
52	F	L TLE	LT	LT	L mTL DNET	N/A	0.34	-1.0	0	0	N/A			
72	M	L TLE	LT	NC	LT radiation necrosis	N/A	6.4	-0.45	3	0	0%			
47	M	L TLE	None	LT	Non-lesional	N/A	355.8	-0.07	209	0	0%	x		
31	F	L TLE	LT, LTP	LT	Non-lesional	N/A	598.7	-0.97	394	20	3.2%	x	x	x
50	M	L TLE	LT	LT	L mTL T2 FLAIR hyperintensity	None	554	-0.99	132	47	9.0%	x	x	x
62	F	L TLE	LT > RT	LT	Mild bilateral mTL and anterior temporal atrophy	None	56.1	0.61	14	0	0%			
56	M	L TLE	LT	LT	L MTS	None	401	-0.97	317	5	1.6%	x	x	x
65	M	L TLE	LT	LT	L MTS	LT	43.8	-0.93	23	0	0%			
27	F	L TLE	None	LT	Non-lesional	N/A	1.1	0.48	1	0	0%			
65	F	L TLE	LT	LT[3]	L mTL T2 FLAIR hyperintensity	LT	54.5	-0.77	3	0	0%			
57	F	L TLE	LT	LT[3]	L MTS. Diffuse atrophy	LT	31.5	-0.83	7	1	12.5%			x
41	F	L TLE	LT	LT	L MTS. Diffuse atrophy	LT	7.6	-1.00	1	0	0%			
19	F	L TLE	LT	LT	L MTS	LT	213.7	-0.87	33	315	5.7%	x	x	x
32	M	R TLE	RT	NC	R fornix < L fornix	None	175.6	0.93	91	4	0%	x	x	x
32	F	R TLE	RT > LT	RT	Small R inf temporal encephalocoele.	RT	916.6	-0.56	388	6	0.3%	x	x	x

							Entire Recording		1-hour Sleep Recording					
Age	**Sex**	**Dx**	**IEDs**	**Sz Onset**	**MRI**	**FDG-PET[5]**	HEAnet Detection Rate (per hr)[6]	ASym HEA	HEAnet Detection Rate (per hr)[7]	Expert-Annotated Spike Rate (per hr)[8]	% HEAnet Detections coinciding with Expert Detections	TLE Diagnosis		
												HEA net	Ex 1	Ex 2
30	M	R TLE	RT	RT	Hemorrhagic foci in R mTL and posterior limb of R internal capsule (trauma)	RT	925.8	0.99	301	2	0.7%	x		x
33	M	R TLE	RT	RT	R mTL T2 FLAIR hyperintensity	RT	399.8	0.70	323	67	11.8%	x	x	x
25	F	R TLE	RT > LT	RT	R inferior temporal, L frontal opercular, L corona radiata hamartomas (NF1)	RT	64.1	0.97	3	3	0%			
24	M	R TLE	RT	RT	Loss of internal architecture of R hippocampus. Fullness of R amygdala/uncus.	RT	189.8	0.91	177	5	2.2%	x	x	x
46	F	B TLE	LT	LT, RT[2]	Non-lesional	N/A	217.5	-0.72	172	104	26.2%	x	x	x
40	M	B TLE	RT, LT	NC[4]	Non-lesional	LT	97.5	0.51	17	5	0%		x	x
47	M	B TLE	RT[1]	RT[3]	Non-lesional	RT	366.5	0.40	95	1	1.0%	x		
19	M	B TLE / L TLE	RT > LT	LT	Non-lesional	RT	1134.7	0.61	870	7	0.5%	x	x	x
40	F	HC					1.8		0					
56	F	HC					40.4		24					
78	F	HC					0.8		1					
22	M	HC					0.5		3					
45	M	HC					16.6		12					
57	F	HC					6.1		2	2				x
49	F	HC					0		0					

							Entire Recording		1-hour Sleep Recording					
Age	**Sex**	**Dx**	**IEDs**	**Sz Onset**	**MRI**	**FDG-PET[5]**	HEAnet Detection Rate (per hr)[6]	ASym HEA	HEAnet Detection Rate (per hr)[7]	Expert-Annotated Spike Rate (per hr)[8]	% HEAnet Detections coinciding with Expert Detections	TLE Diagnosis		
												HEA net	Ex 1	Ex 2
73	F	HC					179.6		57					
32	M	HC					1		0					
23	F	HC					10.9		8	16				x
30	M	HC					0.8		2					
20	F	HC					1.6		1					
52	F	HC					1.6		0					
37	F	HC					2.3		1					
29	M	HC					8.2		8	3				x
67	F	HC					10.8		7	6				x
21	F	HC					3.0		8	7				x
41	F	HC					10.5		3					
29	F	HC					6.7		3					
50	F	HC					4.3		1	2				x

B – bilateral; Ex – expert; HC – healthy control; L – left; NC – none captured; NF1 – Neurofibromatosis type 1; R – right; T – temporal; Sz - seizure

[1] Subsequent Phase 2 study showed L and R hippocampal IEDs
[2] LT were clinical seizures; RT were subclinical
[3] Subsequent Phase 2 study showed L hippocampal onset
[4] Subsequent Phase 2 study showed independent L and R hippocampal onsets
[5] Hypometabolic regions
[6] For entire recording, excludes HEAnet detections that overlapped with SpikeNet detections. Detection rate includes all detections made on both right and left sides.
[7] For 1-hour expert-annotated recordings, excludes HEAnet detections that overlapped with expert annotations of visible epileptiform discharges. Detection rate includes all detections made on both right and left sides.
[8] Includes all epileptiform discharges annotated by at least 1 expert

eTable 3. Clinical Details and HEAnet Performance for Data Set 3

Age	Sex	Dx	IEDs	Sz Onset	MRI	FDG-PET[1]	Hours NREM	HEAnet Detection Rate[2] Left (per hr)	HEAnet Detection Rate[2] Right (per hr)	$ASym_{HEA}$	HEAnet Prediction (TLE vs HC)	HEAnet Prediction (laterality)
44	M	L TLE	LT	LT	L MTS	LT	4.98	17.49	2.61	-0.74	HC	
65	F	L TLE	LT	LT	Non-lesional	LT	8.13	8.49	0.62	-0.86	HC	
28	M	L TLE	LT	LT	L post T Cavernoma	N/A	6.68	35.66	2.85	-0.85	TLE	L
32	M	L TLE	LT	LT	L mTL DNET		8.00	98.38	1.75	-0.97	TLE	L
34	F	L TLE	LT	LT	L MTS	LT	4.39	122.50	2.73	-0.96	TLE	L
44	M	L TLE	LT	LT	Non-lesional	None	8.07	56.16	8.80	-0.73	TLE	I
75	M	L TLE	LT	LT	Non-lesional	LT	39.73	141.19	17.74	-0.78	TLE	L
63	F	L TLE	BT	LT	Non-lesional	LT	4.17	60.96	0.72	-0.98	TLE	L
64	M	L TLE	LT	LT	L mTL T2 FLAIR hyperintensity	N/A	18.05	32.47	0.44	-0.97	HC	
68	F	L TLE	LT	LT	L MTS	N/A	5.33	40.00	0.38	-0.98	TLE	L
52	F	L TLE	LT	LT	Non-lesional	None	25.81	202.49	1.36	-0.99	TLE	L
46	M	L TLE	LT	LT	Non-lesional	N/A	30.04	438.32	0.03	-1.00	TLE	L
63	M	R TLE	RT	RT	Non-lesional	None	0.37	190.91	120.00	-0.23	TLE	I
37	F	R TLE	RT	RT	Non-lesional	RT	3.13	2.24	10.88	0.66	HC	
33	M	R TLE	None	RT	R MTS	RT	4.74	1.05	8.01	0.77	HC	
53	M	R TLE	RT	RT	Non-lesional	None	8.63	34.43	46.72	0.15	TLE	I
59	M	R TLE	RT	RT	R T Encephalocele	RT	35.93	1.25	4.54	0.57	HC	
34	F	R TLE	RT	RT	R MTS, R superior T Cavernoma	RT	14.63	0.82	59.93	0.97	TLE	R
36	F	B TLE	BT (L>R)	BT	R MTS	RT	27.01	44.47	3.22	-0.86	TLE	L
49	F	B TLE	BT	BT	L MTS	None	35.64	37.57	41.66	0.05	TLE	I
26	M	B TLE	BT	BT	Non-lesional	N/A	17.54	162.58	72.68	-0.38	TLE	I
45	F	B TLE	LT	BT	Non-lesional	N/A	31.46	29.37	25.81	-0.06	HC	

Age	Sex	Dx	IEDs	Sz Onset	MRI	FDG-PET[1]	Hours NREM	HEAnet Detection Rate Left (per hr)	HEAnet Detection Rate Right (per hr)	$ASym_{HEA}$	HEAnet Prediction (TLE vs HC)	HEAnet Prediction (laterality)
30	F	HC					2.60	0.77	0.38		HC	
44	F	HC					1.96	1.02	2.55		HC	
32	M	HC					3.96	3.28	9.09		HC	
47	M	HC					1.11	3.61	0.90		HC	
60	M	HC					1.02	3.93	2.95		HC	
53	F	HC					0.84	1.19	0.00		HC	
58	M	HC					1.20	28.33	11.67		HC	
35	F	HC					1.68	0.60	0.00		HC	
27	M	HC					12.83	0.23	0.55		HC	
32	F	HC					0.77	3.91	1.30		HC	
26	F	HC					7.28	0.82	0.41		HC	
43	M	HC					1.10	0.00	1.82		HC	
19	F	HC					7.61	0.39	0.39		HC	
39	F	HC					0.93	0.00	3.24		HC	
62	F	HC					2.75	2.18	1.82		HC	
28	M	HC					1.69	2.36	0.59		HC	
20	F	HC					1.40	1.43	0.71		HC	
21	F	HC					1.66	3.62	0.60		HC	
52	F	HC					4.73	2.54	3.17		HC	
49	F	HC					19.16	16.86	206.33		TLE	R
35	F	HC					27.99	0.00	0.00		HC	
43	M	HC					26.37	0.00	0.34		HC	
22	F	HC					16.13	0.00	0.00		HC	
52	F	HC					18.29	0.05	0.11		HC	

B – bilateral; HC – healthy control; I – Indeterminate; L – left; R – right; T – temporal; Sz - seizure

[1] Hypometabolic regions

eTable 4. Performance and Hyperparameters of Individual CNNs Comprising HEAnet

Model #	ROC-AUC	PR-AUC	Structure	N_B	N_{conv}	f_0	lr	N_{cycles}
1	0.856	0.327	2D	2	3	32	0.0005	30
2	0.863	0.324	2D	2	3	32	0.0005	30
3	0.866	0.318	2D	1	4	64	0.0005	30
4	0.865	0.299	1D	1	4	16	0.0005	40
5	0.853	0.296	1D	1	4	32	0.0005	60
6	0.858	0.277	1D	1	4	32	0.0005	60

eTable 5. Size of the Temporal Dimension (D) of 1-D Filters ($K_{i,1D}$) for Each Convolutional Layer

Input of time dimension	**$K_{i,1D}$**
8	3
16	4
32	8
64	16
128	32
256	32

eTable 6. CNN Hyperparameters and Range of Values Evaluated

Hyperparameter	**Values tested**
Learning Rate (lr)	0.001 - 0.00001
Batch Size (bs)	256, 1024
Number of convolutional layers (N_{conv})	3, 4, 5
Number of blocks per convolutional layer (N_B)	1, 2
Number of filters in the first convolutional layer (f_0)	16, 32, 64
Filter size 2D ($K_{i,2D}$)	3, 4, 5, 7, 8, 9, 16, 32
Input size (W_0)	64, 128, 256 (0.25s, 0.5s, 1s)
Number of input channels (C_0)	14 (ipsilateral only), 19, 25 (all channels)
Threshold for FO spike detector (th_{FO})	0.870, 0.923, 0.966 (Corresponding PPVs 0.7, 0.8, 0.9)
Jitter frequency (J)	0, 5, 10
Early Stopping	Use, do not use.
Number of cycles (N_{cycles}) - when no early stopping was used	20, 40, 60, 80, 100

eTable 7. Analysis of False-Positive Detections by HEAnet (Data Set 1)

	% of False Positives	Additional Details
HEA on same side (True positive detections)	43.1%	Definite HEA on FO electrodes (based on morphology and amplitude) ~ 23.8% Probable HEA on FO electrodes (less robust morphology or lower amplitude, but still highly likely to be HEA) ~ 19.1%
No obvious background change	27.9%	Most common: sleep spindles; high frequency, low amplitude backgrounds
Artifact	10.3%	Most common: EKG artifact; electrode pop artifact; low-amplitude myogenic artifact
Spiky deflection	11.0%	Insufficient morphology or amplitude on scalp EEG to qualify as scalp-visible epileptiform discharge, and no HEA on FO electrodes. Most common: small sharp spikes; or suspected forme fruste epileptiform discharges.
HEA on contralateral side	7.5%	
Scalp-visible epileptiform discharge without HEA	0.7%	

eTable 8. Performance of HEAnet Across Awake and Asleep States (Data Set 1)

	% Total Sleep in Testing Dataset	AUC_{ROC}	AUC_{PR}	Sensitivity at PPV ~ 0.9 [4]	Specificity at PPV ~ 0.9 [4]	Sleep Stage-specific PPV
Awake	N/A[1]	0.85 ± 0.02	0.19 ± 0.07 [2]	0.06 ± 0.02	0.999 ± 0.000	0.81 ± 0.10
Asleep (all sleep stages)	100%	0.89 ± 0.01	0.39 ± 0.03 [3]	0.13 ± 0.05	0.999 ± 0.001	
N1	5.1%	0.91 ± 0.02	0.42 ± 0.10	0.14 ± 0.05	0.999 ± 0.001	0.87 ± 0.13
N2	59.2%	0.88 ± 0.01	0.36 ± 0.06	0.13 ± 0.05	0.999 ± 0.001	0.89 ± 0.04
N3	15.9%	0.90 ± 0.03	0.47 ± 0.14	0.18 ± 0.09	0.998 ± 0.002	0.90 ± 0.06
REM	18.2%	0.90 ± 0.02	0.46 ± 0.09	0.14 ± 0.07	0.999 ± 0.001	0.92 ± 0.04

[1] Awake data, consisting of 1-hour of awake recording from each patient, was not included as part of the expert-annotated testing data set. FOnet detections were used to determine ground truth occurrence of HEA for the awake data.

[2] AUC_{PR} for chance level prediction in the awake state is 0.008 (the frequency of HEA in the awake state).

[3] AUC_{PR} for chance level prediction in the asleep state is 0.016 (the frequency of HEA in the asleep state).

[4] Results were calculated using a threshold for HEAnet that yielded a PPV ~ 0.9 in asleep recordings (combined NREM and REM).