





ORIGINAL ARTICLE

An approach for reliably identifying high-frequency oscillations and reducing false-positive detections

Yufeng Zhou¹ | Jing You¹ | Udaya Kumar²  | Shennan A Weiss^{3,4}  | Anatol Bragin^{2,5}  | Jerome Engel Jr^{2,5,6,7} | Christos Papadelis^{8,9,10} | Lin Li^{1,2} 

¹Department of Biomedical Engineering, University of North Texas, Texas, USA

²Department of Neurology, University of California Los Angeles, Los Angeles, California, USA

³Departments of Neurology, Department of Physiology and Pharmacology, State University of New York Downstate, Brooklyn, New York, USA

⁴Department of Neurology, New York City Health + Hospitals/Kings County, Brooklyn, New York, USA

⁵Brain Research Institute, University of California, Los Angeles, California, USA

⁶Department of Neurobiology, David Geffen School of Medicine at UCLA, Los Angeles, California, USA

⁷Department of Psychiatry and Biobehavioral Sciences, David Geffen School of Medicine at UCLA, California, USA

⁸Jane and John Justin Neurosciences Center, Cook Children's Health Care System, Fort Worth, Texas, USA

⁹School of Medicine, Texas Christian University, Fort Worth, Texas, USA

¹⁰Department of Bioengineering, University of Texas at Arlington, Arlington, Texas, USA

Correspondence

Lin Li, Department of Biomedical Engineering, University of North Texas, Texas, USA.

Email: lin.li@unt.edu

Funding information

National Institutes of Health, Grant/Award Number: 1R16-NS131108-01, 7R01-NS104116-02, R01-NS033310, R01-NS065877 and U54-NS100064

Abstract

Objective: Aiming to improve the feasibility and reliability of using high-frequency oscillations (HFOs) for translational studies of epilepsy, we present a pipeline with features specifically designed to *reject false positives for HFOs* to improve the automatic HFO detector.

Methods: We presented an integrated, multi-layered procedure capable of automatically rejecting HFOs from a variety of common false positives, such as motion, background signals, and sharp transients. This method utilizes a time-frequency contour approach that embeds three different layers including peak constraints, power thresholds, and morphological identification to discard false positives. Four experts were involved in rating detected HFO events that were randomly selected from different posttraumatic epilepsy (PTE) animals for a comprehensive evaluation.

Results: The algorithm was run on 768-h recordings of intracranial electrodes in 48 PTE animals. A total of 453 917 HFOs were identified by initial HFO detection, of which 450 917 were implemented for HFO refinement and 203 531 events were retained. Random sampling was used to evaluate the performance of the detector. The HFO detection yielded an overall accuracy of 0.95 ± 0.03 , with precision, recall, and F1 scores of 0.92 ± 0.05 , 0.99 ± 0.01 , and 0.94 ± 0.03 , respectively. For the HFO classification, our algorithm obtained an accuracy of

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2022 The Authors. *Epilepsia Open* published by Wiley Periodicals LLC on behalf of International League Against Epilepsy.

0.97 ± 0.02 . For the inter-rater reliability of algorithm evaluation, the agreement among four experts was 0.94 ± 0.03 for HFO detection and 0.85 ± 0.04 for HFO classification.

Significance: Our approach shows that a segregated pipeline design with a focus on false-positive rejection can improve the detection efficiency and provide reliable results. This pipeline does not require customization and uses fixed parameters, making it highly feasible and translatable for basic and clinical applications of epilepsy.

KEYWORDS

complex wavelet, epilepsy, high-frequency oscillations, topographical analysis

1 | INTRODUCTION

High-frequency oscillation (HFO), classified as ripples (80-240 Hz) and fast ripples (240-500 Hz), is regarded as a promising biomarker of epilepsy.¹⁻⁴ Such biomarkers can help localize the epileptogenic zone, the brain region that is indispensable for generating seizures in patients with drug-resistant epilepsy (DRE).⁵⁻¹⁰ Yet, the clinical utility of HFOs in the presurgical evaluation of patients with DRE is limited. Currently, clinicians perform the detection of HFO through visual inspection, a process that is challenging and time-consuming since artifacts and environmental noise often mimic HFOs. A reliable, automated computational method is preferred to discard these artifacts and extract the true HFOs.

Conventional HFO detection methods implement bandpass filters for local field potentials (LFPs) followed by multiple thresholding in terms of amplitude and power to identify HFOs.¹¹⁻¹⁶ The short-time-energy (STE) method proposed by our group¹⁷ is one of these methods, whose main advantages are the low computational cost and the short processing time. This method does not require a prior information of HFO labels and has been shown to be effective for detecting HFOs in animals and patients with epilepsy.¹⁷⁻¹⁹ However, bandpass filtering and the use of amplitude thresholds do not separate true HFOs from sharp transients,²⁰ sharp-contoured epileptiform spikes,²¹ and other physiological events. Therefore, STE methods still require expert review and manual flagging of true positives from a database, rendering the detection pipeline labor-intensive and not fully automated.

Similar methodological approaches based on filtered signals have been proposed aiming to improve the HFO detection.²¹⁻²³ Several research groups have suggested that an isolated “island” in the time-frequency plot is the biosignature that characterizes a true HFO.^{20,23,24} Yet, nontrue HFO events may also present a similar distinct biosignature,²⁵ making it more challenging to exclude false HFOs

Key Points

- A novel algorithm was introduced to reject false positives after initial HFO detection.
- A high level of accuracy and reliability was achieved in the comprehensive evaluation of multiple datasets.
- Translatable and adaptable to other HFO detection algorithms.

by using this approach. Moreover, fast brain activities, such as the physiological sharp spikes, and transient motion artifacts also show the “blob-like” feature in the time-frequency analysis plot.²⁶ Thus, there is an urgent need to develop new methodological approaches that discard false HFO positives (HFO FPs) in order to improve the reliability of the HFO detectors.

To address the above challenges, we proposed here a novel HFO fully automated method that combines the extraction of HFOs spectral feature with the analysis of events morphology.²⁶⁻²⁹ This method categorized three common HFO FPs, including sharp transients, sharply contoured epileptiform spikes, and background signals. Our main hypothesis is that the rejection of these three types of HFO FPs can significantly improve the robustness of the intracranial HFO detection. Here, we analyzed a dataset of posttraumatic epileptic animals with long-term sustained intracranial electrophysiological recordings to test the above hypothesis. This algorithm was developed as a separate module to reversely reject the events after the HFO detection. The module-based design provides more step-by-step performance information than the integrated pipeline. A comprehensive evaluation was designed to evaluate each step of HFO rejection. The assessment of the validity and reliability of this new approach was reported.

2 | METHOD

2.1 | Experimental setups

The data pool consisted of 48 male Sprague-Dawley rats (300-350 g) that were treated with lateral fluid percussion injury (FPI) in a posttraumatic epilepsy model from our early studies.^{19,30-32} The data pool consisted of LFPs recorded at depth electrodes with sampling frequencies of 3000-10 000 Hz. Specifically, 16 electrodes were implanted in the following areas: prefrontal cortex, thalamus, traumatic brain injury (TBI) areas, and hippocampal. The recordings were selected from the interictal period with a length of 3 h. A total of 768 h of LFPs during the epileptogenesis were analyzed in this study. Details of the experimental setup and data acquisition have been presented in our previous publications.^{18,32}

2.2 | Data preprocessing

Prior to the HFO analysis, raw data were first down-sampled to 3KHz and then manually selected for the interictal period. During the data selection, recordings contaminated by long-lasting (>5 s) noise, poor connections, or large shocks were removed. Instead of using all 16 channels of data, we manually identified high-quality 8 channels for data analysis because our electrodes were designed as dual channels at each location. The main remaining disturbances of the signal were background noise and sharp transients that were difficult to be removed by visual inspection. Therefore, the core algorithm was built aiming to distinguish these two types of events from true HFOs. Specifically, eight brain regions were selected for our study, including bilateral prefrontal cortex (LFC, RFC), striatum (LST, RST), perilesional areas (LTBI, RTBI), and hippocampus (LHP, RHP).^{19,30-32}

2.3 | HFO initial detection

Before we run the algorithm, a HFO initial detection was performed based on the two purposes. The first is to provide the small portion of dataset for feature selection, and the second is to include as many potential HFOs as possible for refinement. In this step, STE method¹⁷ was implemented to find all HFO candidates and retrieve a 2-s HFO epoch (with HFO peak at the center) for further analysis.

2.4 | Algorithm Framework

The complete pipeline consists of three major modules, including: (1) feature selection module, (2) HFO

detection module, and (3) HFO classification module (Figure 1).

2.4.1 | Module 1: Feature selection

To explore key features that lead to different types of events, two experts were invited to review the database generated by the HFO initial detection. Experts manually selected 1000 true HFO events, 1000 all except sharp transients and true HFOs (AESH), and 1000 sharp transients (ST). Only events that were agreed by both reviewers were used for feature selection. The labeled datasets were excluded from the further HFO refinement process. The HFO refinement process was performed based on the remaining unlabeled datasets ($n = 450\,917$) (Figure 1).

The reviewers scored the events by examining the local field potential (LFP). Specifically, the reviewers screened the LFP raw signal, filtered signal, and the time-frequency plot and created the descriptive characteristics of HFOs from them. Guided by reviewers' criteria, we analyzed the labeled data generated two feature sets and they were defined as follows: (1) set A included the normalized average power and power ratio, and (2) set B included peak ratio and the start/end time of the peak. The interpretation of constructing two sets was that the reviewers found set A to be effective in distinguishing sharp transients from the data pool (Figure 2A), and meanwhile set B worked in differentiating HFOs from AESH (Figure 2B). Based on the analysis of these two numerical distributions, the thresholds used to establish the HFO refinement were set, and they are as follows: normalized average power: 0.2; power ratio: 0.7; peak ratio: 0.6; start time: less than 0; end time: greater than 0.

2.4.2 | Module 2: HFO refinement

To discard HFO FPs from the original database, a multi-threshold layer was created based on the two feature sets generated by feature selection module. There are three layers in this section, and the detailed design is described below.

(1) Peak constraint layer

Based on the LFP of labeled dataset, we found that the HFOs always represent rapid oscillations around the center, which contrasts with the AESH. Consequently, it is plausible to assume that the difference in shape between the true HFOs and AESH will be reflected and enhanced after implementing the 80-520 Hz bandpass filter on the LFPs. Then, it is also feasible to separate

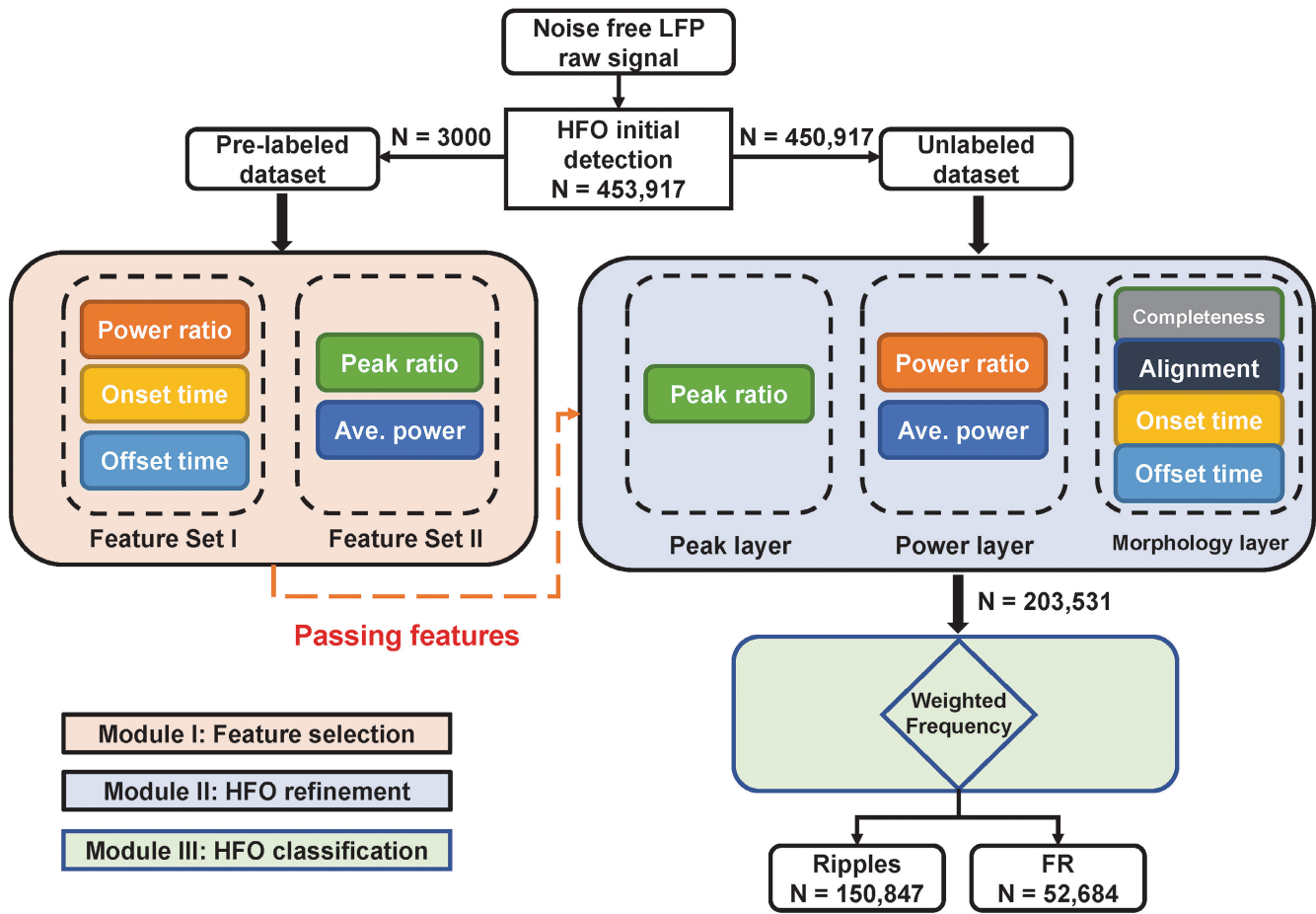


FIGURE 1 Schematic diagram of the HFO detection pipeline. In the HFO refinement module, the completeness and alignment subsystems are introduced from the previous study. All remaining are the re-combination of features from the feature selection module.

the HFOs from the background noise by quantifying the burst of peaks on filtered data. Thus, we calculated the ratio between the peaks in the center 0.1 s range and the total number of peaks for that event. The ratio is denoted as P_d and peaks are defined as follows: if points are above or below the mean ± 2 times of standard deviation of the baseline, they will be considered as peaks. We set the ratio to 0.6 by looking at the 3D distribution on labeled dataset (Figure 2B).

(2) Power thresholding layer

It is worth noting that sharp transients also reflect a burst of peaks, inevitably making it hard to differentiating them from true HFOs. Therefore, adding a layer that can distinguish sharp transients is necessary. During the feature selection process, the reviewers found that the time-frequency maps obtained by Gabor wavelets clearly showed differences. Hotspots of ST reflected a thin and long candle-shaped contour, while HFOs typically represent more of a squeezed or orb shape around the center, with AESH not indicating any obvious hotspot. Here, we introduced a similar approach to quantify the predominant

characteristic of sharp transients.³³ The first property introduced was the average power \bar{Z} , and the power of the k^{th} time window is denoting as: $\bar{Z}_k = \frac{1}{TF} \sum_l \sum_{m=t_k}^{t_{k+1}-1} Z_{lm}$, it was set to $\frac{1}{TF} \sum_l \sum_m Z_{lm}$, which was identical to our previous study.³³ The other was the power ratio HPR , and the ratio of the k^{th} window is as follows: $HPR_k = \frac{1}{TF} 1^T \left(\sum_{m=t_k}^{t_{k+1}-1} Z_{lm} \geq \bar{Z}_k \right)$, which was set to 0.7 following the value of previous study.³³ Specifically, the terms T , F , and t_k refer to the length of the time vector, frequency vector, and start point of the k^{th} time window, respectively. Z represents the time-frequency matrix of an epoch, and 1 is the vector whose all entries are 1. Moreover, the length of the time window was set to 100 ms, and the stride was set to ensure the 90% overlap of two consecutive windows.

(3) Contour morphology layer

The feature P_d demonstrates its effectiveness in separating HFOs from background signals. The HFO usually reflects the blob shape around the center, while the AESH show no sign of energy concentration. The location of hotspot can be considered as another key factor to distinguish these two. The hotspot was quantified by the contour method,²⁶⁻²⁹ and

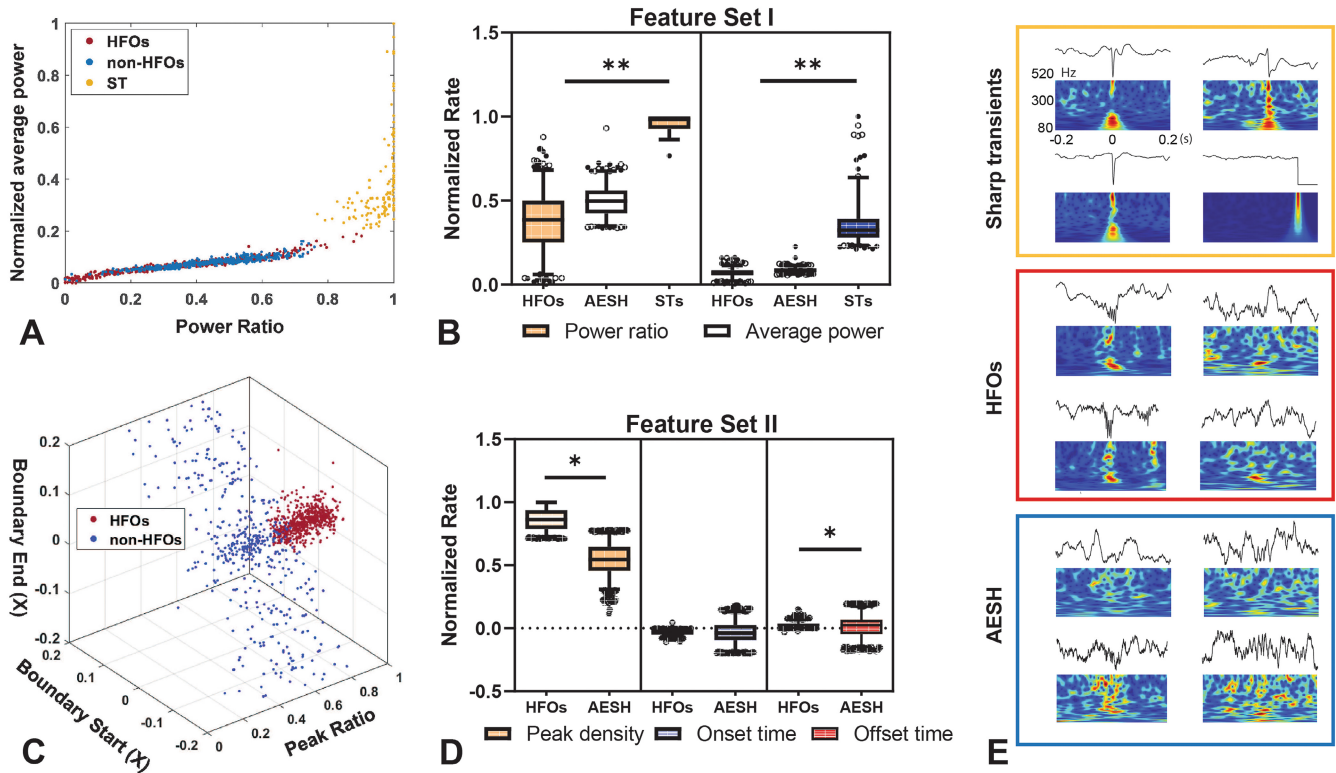


FIGURE 2 HFO selection process. A, 2D distribution of three types of events based on power ratio HPR_k and normalized power \bar{Z}_k , with red, blue, and yellow dots representing HFOs, AESH, and sharp transients (ST), respectively. B, The statistical analysis of feature set one with respect to HFOs, AESH, and STs. C, 3D distribution of HFO and AESH regards to peak ratio, start/end time of POE. The red dots refer to HFOs, while blue dots represent AESH. D, The statistical analysis of feature set two with respect to HFOs and AESH. E, The demonstrative events of the ST (top), HFO (middle), and AESH (bottom).

the contour analysis can be generalized into three aspects. The first aspect is the completeness of the contour. In this study, only closed-loop contours (CLCs) are included for further analysis. Specifically, contours with identical starting and ending points are defined as CLCs, while open-loop contours (OLCs) refer to the ones that have different starting and ending points. The second aspect is the alignment. The purpose is to discard epochs that do not follow certain rules, and they are defined as follows. (i) Discard the low power-level CLCs by setting the threshold to $0.6 \times (\max Z - \min Z) + \min Z$ to make a trade-off between too many false positives and too few true HFOs. (ii) Discard those groups with the highest energy CLC located in the inner circle. (iii) Discard the groups containing less than 3 CLCs. After the three-step threshold processing, the outermost CLC in each group is defined as the boundary of the group (BOG) and the CLC with the highest power level is defined as the peak of the group (POG). The third aspect is the on/offset of the hotspot, and the purpose is to ensure the events reflect a blob shape on the time-frequency map. The details are described as follows: (i) Find the highest power-level POG among all candidates and its BOG. These two parameters are denoted as the peak of epoch (POE) and boundary of epoch (BOE), respectively. (ii) Denote the left

and right edge of the POE as the starting and ending points of the events. Set the starting points boundary < 0 and set the ending points boundary > 0 (Figure 2). All other parameters in this section follow the settings in the previous study.³³

Module 3: HFO classification

All unlabeled data were processed according to the pipeline to form a refined dataset. The POE of each individual HFO was implemented to compute the corresponding weighted frequency F , which is defined as: $F_{ij} = f_D^T Z_D$, where D denotes all complete grid cells covered by the POE. The terms f_D and Z_D represent the frequency value and power level of a specific grid cell retrieved from the time-frequency matrix. If F is above 240 Hz, then this epoch is classified as a fast ripple. Otherwise, it is classified as ripple.

2.5 | Comprehensive evaluation of the HFO detection algorithm

At present, there is little consensus on how HFO detection algorithms should be comprehensively evaluated.³⁴

One of the problems is the difficulty of assembling a team of experts in HFO identification. Previous studies have found both disagreement³⁵ and relatively good agreement³⁶ among raters on HFO detection. HFO events are most often observed in the hippocampal CA1 region and the entorhinal cortex.³⁷ However, it is more difficult to confidently label HFO events in other brain regions and when the recording microelectrodes are away from the soma.^{38–41} In addition, scoring a large sample of events is very laborious. To assess the accuracy and reliability of the algorithm, we implemented a rating protocol involving four experts, either clinical epileptologists or neurology researchers. Given the very long recordings (nearly thousands of hours in total), visual inspection of all HFOs was impractical, so a random sampling was introduced for evaluation. Specifically, epochs were randomly selected among several subjects to create test sets to comprehensively evaluate the HFO detection accuracy and ripple/fast ripple distinguishability. We constructed four different event subgroups in the evaluation dataset, including 1000 detected HFOs, 1000 rejected HFOs, 500 ripples, and 500 fast ripples. Specifically, the 1000 detected HFOs used to evaluate HFO detection were separated from the 500 ripples/fast ripples used to evaluate HFO classification. All individual events were independently labeled by four experts. No platform/software was given to the reviewer, instead we provided the unfiltered LFPs and their corresponding time-frequency plots (see Supplementary Figure S1) for evaluation.

2.6 | Statistics

2.6.1 | Assessing the effectiveness of HFO detection and classification

To evaluate the performance of HFO detection, the task of the raters was to label: (1) true HFOs from 1000 detector-dropped HFO events, and (2) false HFOs from 1000 algorithm-identified HFO events. In the first experiment, the expert labeled HFO events were referred to as false negatives (FNs) and the remaining unlabeled events were referred to as true negatives (TNs). In the second experiment, expert labeled AEH (all events except HFOs) were considered as false positives (FPs) and the remaining unlabeled events were denoted as true positives (TPs). The results of evaluation from raters formed a 2-by-2 confusion matrix that further quantified accuracy (Acc), precision (P), recall (R), and F1 score ($F1$) according to the following equations. (i) $Acc = \frac{TP+TN}{N}$; (ii) $P = \frac{TP}{TP+FP}$; (iii) $R = \frac{TP}{TP+FN}$; (iv) $F1 = \frac{2*P*R}{P+R}$. The term N represents the total number of samples used for review, which in this study corresponds to 2000.

To verify the distinguishability of the HFO classification, the rater needs to indicate the source of the error from the dataset of 500 detector-identified ripples, 500 detector-identified fast ripple, and 500 detector-dropped events. Specifically, the rater should empirically indicate that an incorrectly detected ripple should be classified as a fast ripple or as other (either a AESH or a sharp transient). Therefore, a 3-by-3 confusion matrix was created to retrieve the overall accuracy (Acc), precision, recall, and F1 score for ripple ($P_r, R_r, F1_r$) and fast ripple ($P_{fr}, R_{fr}, F1_{fr}$), respectively. Specifically, these parameters were computed as follows (i) $Acc = \frac{\sum_{i=j}^3 N_{ij}}{\sum_{i=1}^3 \sum_{j=1}^3 N_{ij}}$; (ii) $P_r = \frac{N_{11}}{\sum_{i=1}^3 N_{i1}}$, $P_{fr} = \frac{N_{22}}{\sum_{i=1}^3 N_{i2}}$; (iii) $R_r = \frac{N_{11}}{\sum_{j=1}^3 N_{1j}}$, $R_{fr} = \frac{N_{22}}{\sum_{j=1}^3 N_{2j}}$; (iv) $F1_r = \frac{2*P_r*R_r}{P_r+R_r}$, $F1_{fr} = \frac{2*P_{fr}*R_{fr}}{P_{fr}+R_{fr}}$.

The term N_{ij} is denoted as the ij^{th} entry of the confusion matrix.

2.6.2 | Assessment of the inter-rater reliability

To evaluate the validity of the test and the stability of the results, we conducted an inter-rater reliability analysis. We chose Cohen's kappa to measure the inter-rater reliability of the scores from the previous step (section 2.5.1). For more details on computing Cohen's Kappa, see the supplementary materials. In this paper, the conventional description of κ is implemented, which follows: less than 0, poor reliability; 0 to 0.2, slight agreement; 0.21 to 0.4, fair; 0.41 to 0.6, moderate; 0.61 to 0.8, substantial; 0.81 to 1, almost perfect agreement.⁴²

3 | RESULTS

3.1 | Results of feature selection module

As described in 2.3.1, five features were included in the feature library, namely P_d, \bar{Z}_k, HPR_k , and onset/offset time of POE. \bar{Z}_k and HPR_k comprised the feature set 1 and are mainly responsible for separating sharp transients from the database (Figure 2A). The histogram of \bar{Z}_k and HPR_k shows statistical differences in distribution among all three categories, revealing that the power and shape of the hotspot do help to distinguish sharp transients (Figure S2A). On the contrary, the P_d and the onset/offset time of POE constitute the feature set 2, which aims to distinguish HFOs from the background signal. The true HFOs are clustered in a certain region of this 3D space, while the AESH are sparsely distributed (Figure 2C). The bar plot of the three attributes of feature set 2 shows the probability distribution of selected features in these two categories,

reflecting the more concentrated energy of HFOs compared with the background signals (Figure S2B). Thus, a favorable HFO detection pipeline was generated by these two feature sets being integrated into the thresholding process. The detailed distributions of these five selected features across different data types are represented in Figure 2B and Figure 2D, and the significant differences are listed in Table S1.

3.2 | Results of HFO detection and classification module

To investigate the details of HFO refinement, each processing layer of HFO rejection was carefully screened (Figure 3). Sharp transients revealed a long and narrow hotspot that spans across the entire frequency range, which was substantially different from HFOs/AESH, suggesting the effectiveness of power and shape of the contour in differentiating sharp transients and others (Figure 3A&B). Contours were split into OLCs and CLCs, and only CLCs with higher power level than the hyperparameters were kept for the following steps (Figure 3C&D). At last, the onset and offset time of POE was implemented to check the position of the hotspot with the left dashed line representing onset time of the event and the right dashed line referring to the offset time (Figure 3E). The AESH revealed the off-centered hotspot, which was captured by this layer and substantiating the effectiveness of feature selection layer.

The HFO classification was performed after HFO detection, and weighted frequency was the threshold implemented in this process. In this study, a total number of 453917 HFOs candidates in 228 LFP epochs were identified after HFO rough detection. In total, 250368 events were rejected during the HFO refinement process, including 84320 events deleted in the peak constraint layer, 15668 events deleted in the power thresholding layer, 4492 events deleted in the contour morphology layer, and 145906 events deleted at off-center checking layer. After the HFO rejection, 203531 events (44% of total) were retained in the HFO events library.

The data analysis was computed on a 64-core CPU, 264GB ram Dell Precision 7920 workstation and took 89.5h to complete the entire process. The computation load was also tested on a conventional computer with 8-core i7 CPU and 16GB memory capacity. With this build, the HFO initial detection ran for 1 min for a 30-min file containing 8 channels. For the HFO refinement process, the time varied and the average processing time per file was 5-10 min.

3.3 | Comprehensive evaluation for the HFO processing

3.3.1 | Evaluation of the HFO rejection

The evaluation of HFO detection included data from two perspectives: (1) labeling true HFO events from the 1000 algorithm-discarded HFOs and (2) labeling false HFOs from 1000 algorithm-detected HFOs. The confusion matrices based on the evaluation of each rater are presented along with the *Acc*, *P*, *R*, and *F1* scores (Table 1). Specifically, the highest and lowest *F1* scores among all raters were 0.99 and 0.92, indicating a solid performance in the HFO detection and reliability of features we selected for HFO refinement process. We obtained a *P* of 0.92 ± 0.05 , *R* of 0.99 ± 0.01 , with the *Acc* = 0.95 ± 0.03 and the *F1* score = 0.94 ± 0.03 .

3.3.2 | Evaluation of the ripple fast ripple classification

Table 2 not only reflects that our model is reliable in terms of distinguishing ripples and fast ripples, but more importantly summarizes the effectiveness of weighted frequencies in determining the category of the events. The overall accuracy was 0.97 ± 0.02 , revealing the high quality of the classifier. Specifically, the overall *P_r*, *R_r*, and *F1_r* were 0.98 ± 0.02 , 0.98 ± 0.01 , and 0.98 ± 0.01 . On the contrary, the overall *P_{fr}*, *R_{fr}*, and *F1_{fr}* were 0.94 ± 0.05 , 0.98 ± 0.01 , and 0.96 ± 0.03 .

3.3.3 | Inter-rater reliability of HFO identification

As described in section 2.4.3, Cohen's kappa was introduced to analyze the inter-rater reliability. Figure 4A provides details of raters' decision for each randomly selected event in HFO detection. Experts tended to have more disagreements on the incorrectly detected HFOs, indicating that raters were more confident in the events that were rejected by multi-layers. Strong inter-rater agreement was found in the HFO detection section, with an overall κ of 0.94 ± 0.03 (Figure 4C). Figure 4B documents rater disagreements with the detector in HFO classification. The raster plot shows that raters disagreed the most on the cases in the fast ripple labeling. The overall inter-rater reliability was slightly weaker than the HFO classification (Figure 4D), with a corresponding κ of 0.85 ± 0.04 , but still in the range of excellent agreement.

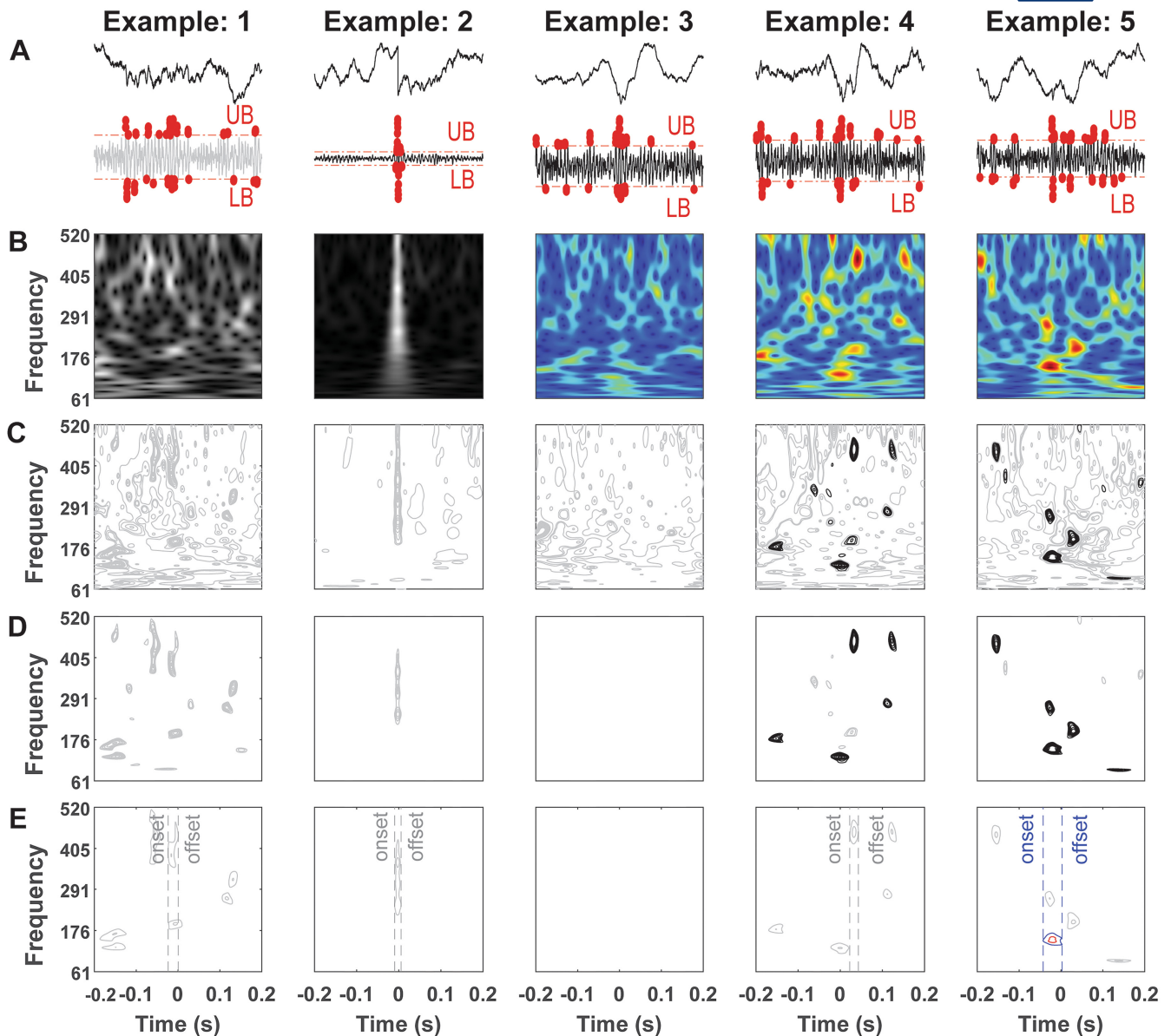


FIGURE 3 Details of HFO rejection. A, Raw plot of selected epochs and the details of peak thresholding layer. UB, upper bound. LB, lower bound. Red dots, peaks outside of the boundary. Gray line, epoch declined by thresholding. B, Time-frequency plot. Gray picture, epoch failed in sharp transients rejection. C, Decomposed contour plot. Gray contours, contours removed due to low-energy CLCs. D, The process of group checking layer. Gray contours, contours less than 4 times CLCs. Black contours, qualified contour. E, Off-center checking layer. Red circle, POE. Blue circle, BOE. Gray contour, unqualified POG and BOG.

3.3.4 | Performance comparison against other methods

To further test that our detector accurately selects true HFOs and discards false detections, we compared our approach with other detectors. As shown in Figure S4, our approach reached the accuracy rate of 0.91 ± 0.049 , while the short line detector method^{43,13} showed a rate of 0.54 ± 0.082 , and the Hilbert approach⁴⁴ showed an accuracy of 0.54 ± 0.08 . The details of reviewers' agreements have been well documented with each bar representing

a false positive (Figure S5). It was worth noting that the co-occurrence of high rater agreements on wrong detections and the low agreements on correct identifications are due to the inconsistent performance of the detectors. In our method, the agreement for detectors to incorrectly identify HFO events was 13.5%, while the agreement to correctly identify HFOs improved to 85%. In contrast, in the Hilbert method, the agreement for incorrect detection raised to 35.9%, while the agreement for accurate identification dropped to 42.3%. Again, the short line length method performed similarly to the Hilbert method on this

	Predicted HFO	Predicted AEH	ACC	P	R	F1-Score
R1						
Labeled HFO	869	13	92.8%	86.9%	98.5%	0.92
Labeled AEH	131	987				
R2						
Labeled HFO	937	8	96.5%	93.7%	99.2%	0.96
Labeled AEH	63	992				
R3						
Labeled HFO	983	2	99.1%	98.3%	99.8%	0.99
Labeled AEH	17	998				
R4						
Labeled HFO	865	11	92.7%	86.5%	98.7%	0.92
Labeled AEH	135	989				

TABLE 1 Confusion matrix of HFO detection

TABLE 2 Confusion matrix of HFO classification

	Predicted Ripple	Predicted Fast Ripple	Predicted AEH	Acc	Pr	Rr	F1r	Pfr	Rfr	F1fr
R1										
Labeled Ripple	495	0	1	99.1%	99%	99.8%	0.994	99.6%	98.6%	0.99
Labeled Fast Ripple	1	498	6							
Labeled AEH	4	2	493							
R2										
Labeled Ripple	499	5	3	97.2%	99.8%	98.4%	0.991	92.8%	99.4%	0.96
Labeled Fast Ripple	1	464	2							
Labeled AEH	0	31	495							
R3										
Labeled Ripple	496	3	0	98.9%	99.2%	99.4%	0.993	97.8%	99.0%	0.98
Labeled Fast Ripple	3	489	2							
Labeled AEH	1	8	498							
R4										
Labeled Ripple	481	6	5	94.5%	96.2%	97.8%	0.97	89.4%	97.6%	0.93
Labeled Fast Ripple	5	447	6							
Labeled AEH	14	47	489							

metric. Agreement on incorrectly detected events rose to 29.4%, while agreement on correctly detected HFOs fell to 36.9% (Figure S6).

We implemented our approach to a new dataset with the Kainic Acid animal model of epilepsy¹⁸ to evaluate the effectiveness of our algorithm. Notably, we did not manually inspect and remove artifacts in the KA dataset as we did in the TBI dataset, and thus, the accuracy on the KA dataset should decrease given the variation in data quality. Specifically, we randomly selected 500 algorithm-detected HFOs on the KA dataset and asked reviewers to mark AEH

(Figure S7). Our method obtained an accuracy of 0.92 ± 0.05 and 0.77 ± 0.08 for HFO detection on the TBI dataset and KA dataset, respectively. The precision decreased on KA dataset, but it still remains at a high level even though the recordings were affected by motion artifacts. In terms of the recall rate for HFO detection, our method has a solid performance, reaching 0.99 ± 0.01 and 1.00 ± 0.00 on TBI and KA dataset, respectively. Finally, the *F1* score reached 0.94 ± 0.03 on the TBI dataset and 0.87 ± 0.05 on the KA dataset, indicating the reliability and feasibility of the selected feature sets for refinement process (Figure S8).

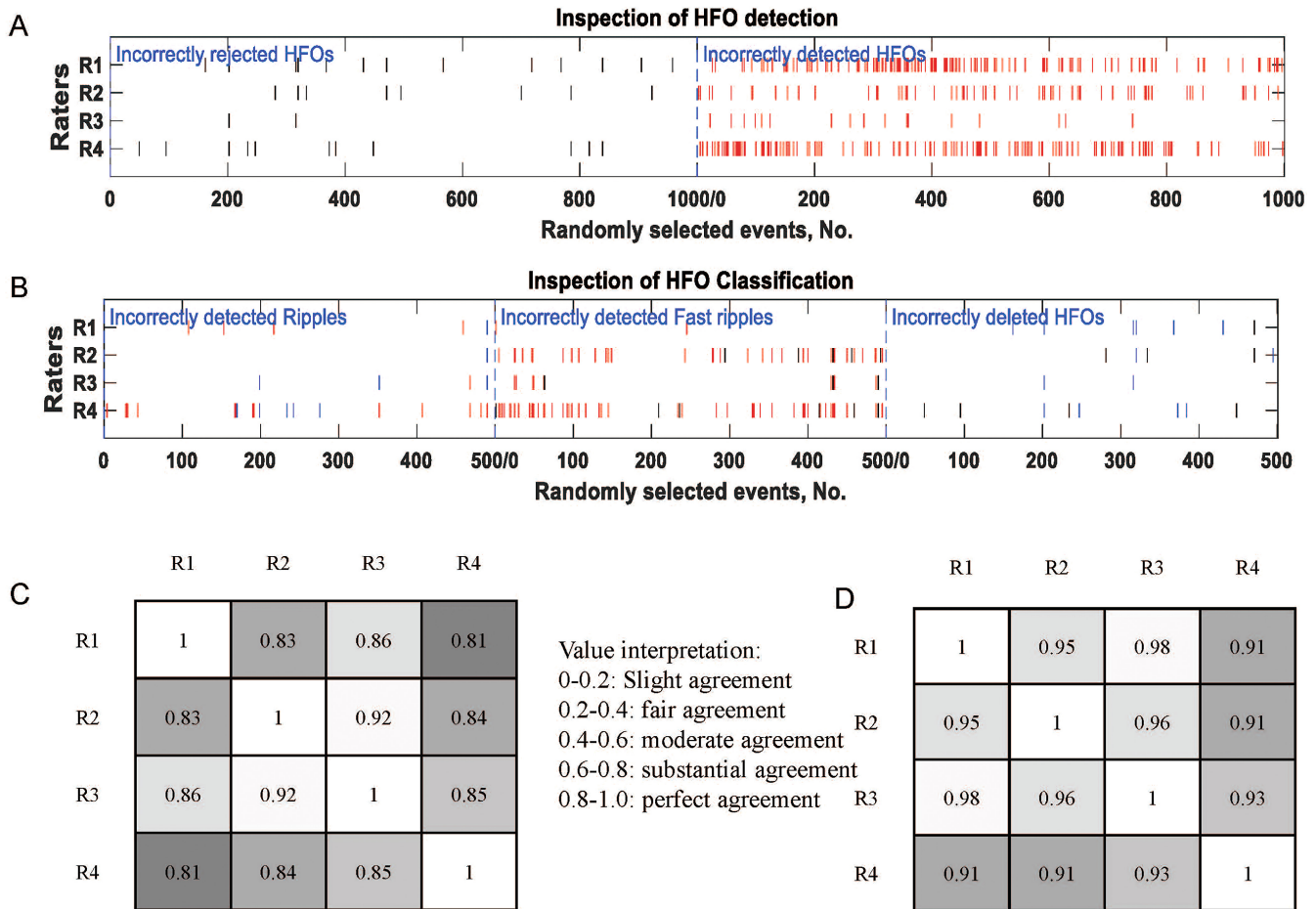


FIGURE 4 Evaluation of our approach. A, Raster plot of HFO algorithm evaluation. The black bars referred to the disputes over HFO events rejected by detector and red bars referred to disagreements over the detected HFOs. B, Raster plot of HFO classification evaluation. Incorrectly detected ripples: Blue bar from fast ripple, red bar from AESH. Incorrectly detected fast ripples: Black bar from ripple, red bar from non-HFO events. Incorrectly deleted HFOs: black bar from ripple, blue bar from fast ripple. C, Correlation matrix of HFO detection between four raters, the lighter color the stronger agreement, with the overall $0.94 \pm 0.03 \kappa$. D, Correlation matrix of HFO detection evaluation between four raters, the lighter color the stronger agreement, with the overall $0.85 \pm 0.04 \kappa$.

4 | DISCUSSION

This study presented a novel approach to reliably and effectively reject HFO FPs, one of the most common problems that hinders the clinical use of HFO. Our approach focused on the analysis of LFPs without machine noise, using AESH and sharp transients as the main rejection targets. We designed a method based on a library of relevant features, followed by a multi-layer threshold processing method for event rejection. Four experts thoroughly evaluated the method and achieved high F1 scores in both detection and classification tests, thus confirming the feasibility of our method. The stability of this study was further validated by high inter-rater reliability, indicating the advanced nature of our method in improving the quality of the HFO database. The highlight of this study is the introduction of the multi-layer threshold module. Five critical features were selected to distinguish between

HFOs, background signals, and sharp transients. The experts marked small portion of events for each category, and they empirically evaluated event type from the shape, power, and locations of the LFPs and time-frequency plots. We used these features and designed the HFO rejection pipeline that simulates the expert reviewing process. Specifically, the peak constraint and morphological identification layers are responsible for identifying HFOs and AESH, while the power thresholding layer aims to distinguish sharp transients.

In the HFO initial detection, the main drawback was the large number of HFO FPs; specifically, 84% of the detected events were later found to be either background noise or sharp transients, which indicated a need for HFO FP rejection. In our multi-layer thresholding process, 33.73% of the HFO FPs were removed by the peak constraint layer and 6.28% of sharp transients by the power thresholding layer and the remaining 59.98% were rejected

in the morphological identification layer. According to the evaluation of HFO detection, our algorithm reached the accuracy rate of 92% and a precision rate of 99%. This result is close to the machine learning approaches with powerful recognition capabilities brought by neural networks.^{45–50} However, the machine learning models hold an end-to-end design, which lack the interpretation and mathematical description of HFOs important to clinical utilization.^{45–48} The HFO classification module of our method distinguished ripples and fast ripples with high accuracy presumably because high-quality HFO candidates had already passed through the feature selection module. Specifically, our design had the precision rate of 98% and 94% in ripple and fast ripple classification, and 98% and 98% in recall rate of ripple and fast ripple, respectively.

It is important to clarify that intracranial EEG implementations of HFO detection using the topographical method consider “candles” as exclusively physiological sharp spikes.^{13,29,34,35} In contrast, in the current LFP implementation, we considered candles as nonphysiological sharp transients and sharp spikes. The major difference is that in LFP recordings from rodents, sharp transients occur frequently and need to be automatically redacted, along with sharp spikes (eg, [Figure 2E](#)). In the current version of the topographic analysis, we do not distinguish HFOs on oscillations from HFOs on spikes, but this can be implemented in later versions. Other potential issue comes from the use of the same parameters for topographical detection of ripple and fast ripples. In our approach, the ripples and fast ripples are processed simultaneously, which inevitably introduces bias to contour morphology layer and then leading to fewer fast ripples been detected since power level of CLC in high-frequency range (>240Hz) is typically lower than the one in the lower frequency region (<240Hz). Additionally, our algorithm identifies only one class of event for each epoch, either ripple or fast ripple. It would omit the case when ripples and fast ripples occur simultaneously. Under this circumstance, our algorithm would be forced to choose the one with higher energy even if the power levels of these two POGs are similar. The case of “fast ripple on ripple” provides valuable information in epilepsy assessment¹⁴ and may not be well addressed in our approach.

In this study, we are still manually selecting thresholds in the feature selection process. In future work, as more data are visually inspected and more features are added to the feature selection process, we hope to establish a tree model that selects the most critical features based on their importance ranking and then automatically detects HFOs without the need to manually set criteria for HFO rejection.

Inter-rater reliability test has been implemented in many studies and has become a golden standard to justify

the performance of the automatic detection method. The inter-rater agreement was applied to identify interictal epileptiform discharges (IEDs) and justify the robustness of the detection method.⁵¹ It was also introduced in seizure detection to make a collective clinical decision.⁵² Former studies reported the inter-rater reliability in the HFO analysis.^{45,49} However, one of the issues was the difficulties to obtain the “ground truth,” especially when the dataset was large. In our approach, the deleted epochs were assembled to form the negative dataset. Instead of labeling all the data, we reported the true negatives by sampling and rating the events from the negative dataset. Our results indicated a high degree of consistency across raters, suggesting that features selected are representative and able to generate the detector without instability issues. Consequently, the implementation of important features and mathematical description of HFOs would be beneficial to HFO study in academic and clinical scenarios.

One issue to note is that we did not test our algorithm in a clinical dataset. Considering that common SEEG data has hundreds of channels, it is expected to take longer to compute compared with the animal data we tested in this study. Additional adjustments may be needed because of the different sampling rates. In addition, we note that our method is not computationally efficient, especially slow in computing the morphological layers.

Instead of one-dimensional data, the morphological analysis in our algorithm expands the data into a two-dimensional graph, which increases the computational load to a large extent. For the 768-h dataset we tested, it took about 90 h to complete the entire process on a 64-core Intel workstation. More optimization efforts should be implemented in future studies to improve the processing efficiency.

5 | CONCLUSION

In this study, we presented a refined pipeline with a segregated design focused on rejecting false positives for HFOs. Through comprehensive expert evaluation, our algorithm showed a promising reliability and accuracy in rejecting HFO FPs. Future work can focus on refining the algorithm by testing larger populations and different recording environments.

ACKNOWLEDGMENTS

This study was supported by the National Institute of Neurological Disorders & Stroke R01-NS065877 (A.B), R01-NS033310 and U54-NS100064 (J.E), 7R01-NS104116-02 (C.P) and 1R16-NS131108-01 (L.L).

We are indebted to Richard Staba and Zachary Waldman for their pioneer work on algorithm development, and

Aristea Galanopoulou, Terence O'Brien, Matt Hudson, and Patricia G. Saletti for their assistance and advice on the data evaluations.

CONFLICTS OF INTEREST

The authors report no conflicts of interest. We confirm that we have read the Journal's position on issues involved in ethical publication and affirm that this report is consistent with those guidelines.

ORCID

Udaya Kumar  <https://orcid.org/0000-0003-0725-0195>

Shennan A Weiss  <https://orcid.org/0000-0003-0825-8439>

Anatol Bragin  <https://orcid.org/0000-0003-1207-8481>

Lin Li  <https://orcid.org/0000-0002-9996-9377>

REFERENCES

- Bragin A, Engel J, Wilson CL, Fried I, Buzsáki G. High-Frequency Oscillations in Human Brain. *Hippocampus*. 1999;9:137–42.
- Bragin A, Wilson CL, Staba RJ, Reddick M, Fried I, Engel J. Interictal high-frequency oscillations (80–500Hz) in the human epileptic brain: Entorhinal cortex. *Ann Neurol*. 2002;52:407–15.
- Chen Z, Maturana MI, Burkitt AN, Cook MJ, Grayden DB. High-Frequency Oscillations in Epilepsy. *Neurology*. 2021;96:439–48.
- Bragin A, Engel J, Wilson CL, Fried I, Mathern GW. Hippocampal and Entorhinal Cortex High-Frequency Oscillations (100–500 Hz) in Human Epileptic Brain and in Kainic Acid-Treated Rats with Chronic Seizures. *Epilepsia*. 1999;40(2):127–37.
- Engel J, Bragin A, Staba R, Mody I. High-frequency oscillations: What is normal and what is not? *Epilepsia*. 2009;50:598–604. <https://doi.org/10.1111/j.1528-1167.2008.01917.x>
- Gotman J. High frequency oscillations: The new EEG frontier? *Epilepsia*. 2010;51:63–5.
- Jacobs J, Staba R, Asano E, Otsubo H, Wu JY, Zijlmans M, et al. High-frequency oscillations (HFOs) in clinical epilepsy. *Prog Neurobiol*. 2012;98:302–15. <https://doi.org/10.1016/j.pneurobio.2012.03.001>
- Tamília E, Matarrese MAG, Ntolkeras G, Grant PE, Madsen JR, Stufflebeam SM, et al. Noninvasive Mapping of Ripple Onset Predicts Outcome in Epilepsy Surgery. *Ann Neurol*. 2021;89:911–25.
- Tamília E, Dirodi M, Alhilani M, Grant PE, Madsen JR, Stufflebeam SM, et al. Scalp ripples as prognostic biomarkers of epileptogenicity in pediatric surgery. *Ann. Clin. Transl. Neurol*. 2020;7:329–42.
- Tamília E, Park EH, Percivati S, Bolton J, Taffoni F, Peters JM, et al. Surgical resection of ripple onset predicts outcome in pediatric epilepsy. *Ann Neurol*. 2018;84:331–46.
- Weiss SA, Alvarado-Rojas C, Bragin A, Behnke E, Fields T, Fried I, et al. Ictal onset patterns of local field potentials, high frequency oscillations, and unit activity in human mesial temporal lobe epilepsy. *Epilepsia*. 2016;57:111–21.
- Weiss SA, Orosz I, Salamon N, Moy S, Wei L, van't Klooster MA, et al. Ripples on spikes show increased phase-amplitude coupling in mesial temporal lobe epilepsy seizure-onset zones. *Epilepsia*. 2016;57:1916–30.
- Worrell GA, Gardner AB, Stead SM, Hu S, Goerss S, Cascino GJ, et al. High-frequency oscillations in human temporal lobe: Simultaneous microwire and clinical macroelectrode recordings. *Brain*. 2008;131:928–37.
- Worrell GA, Parish L, Cranstoun SD, Jonas R, Baltuch G, Litt B. High-frequency oscillations and seizure generation in neocortical epilepsy. *Brain*. 2004;127:1496–506.
- Worrell GA, Jerbi K, Kobayashi K, Lina JM, Zelmann R, Le Van Quyen M. Recording and analysis techniques for high-frequency oscillations. *Prog. Neurobiol*. 2012;98:265–78. <https://doi.org/10.1016/j.pneurobio.2012.02.006>
- von Ellenrieder N, Andrade-Valença LP, Dubeau F, Gotman J. Automatic detection of fast oscillations (40–200Hz) in scalp EEG recordings. *Clin Neurophysiol*. 2012;123:670–80.
- Staba RJ, Wilson CL, Bragin A, Fried I, Engel J. Quantitative analysis of high-frequency oscillations (80–500 Hz) recorded in human epileptic hippocampus and entorhinal cortex. *J Neurophysiol*. 2002;88:1743–52. <https://doi.org/10.1152/jn.2002.88.4.1743>
- Li L, Patel M, Almajano J, Engel J, Bragin A. Extrahippocampal high-frequency oscillations during epileptogenesis. *Epilepsia*. 2018;59:e51–5.
- Kumar U, Li L, Bragin A, Engel J. Spike and wave discharges and fast ripples during posttraumatic epileptogenesis. *Epilepsia*. 2021;62:1842–51.
- Bénar CG, Chauvière L, Bartolomei F, Wendling F. Pitfalls of high-pass filtering for detecting epileptic oscillations: A technical note on “false” ripples. *Clin Neurophysiol*. 2010;121:301–10.
- Liu S, Sha Z, Sencer A, Aydoseli A, Bebek N, Abosch A, et al. Exploring the time-frequency content of high frequency oscillations for automated identification of seizure onset zone in epilepsy. *J Neural Eng*. 2016;13:026026.
- Blanco JA, Stead M, Krieger A, Stacey W, Maus D, Marsh E, et al. Data mining neocortical high-frequency oscillations in epilepsy and controls. *Brain*. 2011;134:2948–59.
- Biro G, Kachenoura A, Albera L, Bénar C, Wendling F. Automatic detection of fast ripples. *J Neurosci Methods*. 2013;213:236–49.
- Papadelis C, Tamília E, Stufflebeam S, Grant PE, Madsen JR, Pearl PL, et al. Interictal high frequency oscillations detected with simultaneous magnetoencephalography and electroencephalography as biomarker of pediatric epilepsy. *J Vis Exp*. 2016;118:54883.
- Amiri M, Lina JM, Pizzo F, Gotman J. High Frequency Oscillations and spikes: Separating real HFOs from false oscillations. *Clin Neurophysiol*. 2016;127:187–96.
- Waldman ZJ, Shimamoto S, Song I, Orosz I, Bragin A, Fried I, et al. A method for the topographical identification and quantification of high frequency oscillations in intracranial electroencephalography recordings. *Clin Neurophysiol*. 2018;129:308–18.
- Waldman ZJ, Camarillo-Rodriguez L, Chervenova I, Berry B, Shimamoto S, Elahian B, et al. Ripple oscillations in the left temporal neocortex are associated with impaired verbal episodic memory encoding. *Epilepsy Behav*. 2018;88:33–40.
- Weiss SA, Berry B, Chervoneva I, Waldman Z, Guba J, Bower M, et al. Visually validated semi-automatic high-frequency oscillation detection aids the delineation of epileptogenic regions

- during intra-operative electrocorticography. *Clin Neurophysiol*. 2018;129:2089–98.
29. Weiss SA, Waldman Z, Raimondo F, Slezak D, Donmez M, Worrell G, et al. Localizing epileptogenic regions using high-frequency oscillations and machine learning. *Biomark Med*. 2019;13:409–18. <https://doi.org/10.2217/bmm-2018-0335>
 30. Bragin A, Li L, Almajano J, Alvarado-Rojas C, Reid AY, Staba RJ, et al. Pathologic electrographic changes after experimental traumatic brain injury. *Epilepsia*. 2016;57:735–45.
 31. Reid AY, Bragin A, Giza CC, Staba RJ, Engel J. The progression of electrophysiologic abnormalities during epileptogenesis after experimental traumatic brain injury. *Epilepsia*. 2016;57:1558–67.
 32. Li L, Kumar U, You J, Zhou Y, Weiss SA, Engel J, et al. Spatial and temporal profile of high-frequency oscillations in posttraumatic epileptogenesis. *Neurobiol Dis*. 2021;161:105544.
 33. Zhou Y, You J, Zhu F, Bragin A, Engel J, Li L. Automatic Electrophysiological Noise Reduction and Epileptic Seizure Detection for Stereoelectroencephalography. *Conf Proc IEEE Eng Med Biol Soc*. 2021;2021:107–12. <https://doi.org/10.1109/EMBC46164.2021.9630651>
 34. Remakanthakurup Sindhu K, Staba R, Lopour BA. Trends in the use of automated algorithms for the detection of high-frequency oscillations associated with human epilepsy. *Epilepsia*. 2020;61:1553–69.
 35. Spring AM, Pittman DJ, Aghakhani Y, Jirsch J, Pillay N, Bello-Espinosa LE, et al. Interrater reliability of visually evaluated high frequency oscillations. *Clin Neurophysiol*. 2017;128:433–41.
 36. Nariai H, Wu JY, Bernardo D, Fallah A, Sankar R, Hussain SA. Interrater reliability in visual identification of interictal high-frequency oscillations on electrocorticography and scalp EEG. *Epilepsia Open*. 2018;3:127–32.
 37. Williams, L. *et al.* Hippocampal and Entorhinal Cortex High-Frequency Oscillations (100–500 Hz) in Human Epileptic Brain and in Kainic Acid-Treated Rats with Chronic Seizures. (1999).
 38. Bragin A, Wilson CL, Almajano J, Mody I, Jerome Engel J. High-frequency Oscillations after Status Epilepticus: Epileptogenesis and Seizure Genesis. *Epilepsia*. 2004;45:1017–23.
 39. Bragin A, Wilson CL, Engel J. Voltage depth profiles of high-frequency oscillations after kainic acid-induced status epilepticus. *Epilepsia*. 2007;48:35–40.
 40. Frauscher B, Bartolomei F, Kobayashi K, Cimbalnik J, Maryse A v 't K, Rampf S, et al. High-frequency oscillations: The state of clinical research. *Epilepsia*. 2017;58:1316–29. <https://doi.org/10.1111/epi.13829>
 41. Jiruska P, Alvarado-Rojas C, Schevon CA, Staba R, Stacey W, Wendling F, et al. Update on the mechanisms and roles of high-frequency oscillations in seizures and epileptic disorders. *Epilepsia*. 2017;58:1330–9. <https://doi.org/10.1111/epi.13830>
 42. Hallgren, K. A. Computing Inter-Rater Reliability for Observational Data: An Overview and Tutorial A Primer on IRR. (2012).
 43. Gardner AB, Worrell GA, Marsh E, Dlugos D, Litt B. Human and automated detection of high-frequency oscillations in clinical intracranial EEG recordings. *Clin Neurophysiol*. 2007;118:1134–43.
 44. Crépon B, Navarro V, Hasboun D, Clemenceau S, Martinerie J, Baulac M, et al. Mapping interictal oscillations greater than 200 Hz recorded with intracranial macroelectrodes in human epilepsy. *Brain*. 2010;133:33–45.
 45. Guo J, Li H, Sun X, Qi L, Qiao H, Pan Y, et al. Detecting high frequency oscillations for stereoelectroencephalography in epilepsy via hypergraph learning. *IEEE Trans Neural Syst Rehabil Eng*. 2021;29:587–96.
 46. Wu M, Qin H, Wan X, Du Y. HFO Detection in Epilepsy: A Stacked Denoising Autoencoder and Sample Weight Adjusting Factors-Based Method. *IEEE Trans Neural Syst Rehabil Eng*. 2021;29:1965–76.
 47. Sciaraffa N, Klados MA, Borghini G, di Flumeri G, Babiloni F, Aricò P. Double-step machine learning based procedure for HFOs detection and classification. *Brain Sci*. 2020;10(4):220.
 48. Zuo R, Wei J, Li X, Li C, Zhao C, Ren Z, et al. Automated detection of high-frequency oscillations in epilepsy based on a convolutional neural network. *Front Comput Neurosci*. 2019;13:6.
 49. Migliorelli C, Bachiller A, Alonso JF, Romero S, Aparicio J, Jacobs-le van J, et al. SGM: A novel time-frequency algorithm based on unsupervised learning improves high-frequency oscillation detection in epilepsy. *J Neural Eng*. 2020;17:026032.
 50. Guo J, Li H, Pan Y, Gao Y, Sun J, Wu T, et al. Automatic and Accurate Epilepsy Ripple and Fast Ripple Detection via Virtual Sample Generation and Attention Neural Networks. *IEEE Trans Neural Syst Rehabil Eng*. 2020;28:1710–9.
 51. Jing J, Herlopian A, Karakis I, Ng M, Halford JJ, Lam A, et al. Interrater Reliability of Experts in Identifying Interictal Epileptiform Discharges in Electroencephalograms. *JAMA Neurol*. 2020;77:49–57.
 52. Rakshashbhavankar AA, Wagh D, Athikarisamy SE, Davis J, Nathan EA, Palumbo L, et al. Inter-rater reliability of amplitude-integrated EEG for the detection of neonatal seizures. *Early Hum Dev*. 2020;143:105011.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Zhou Y, You J, Kumar U, Weiss SA, Bragin A, Engel J, et al. An approach for reliably identifying high-frequency oscillations and reducing false-positive detections. *Epilepsia Open*. 2022;7:674–686. <https://doi.org/10.1002/epi4.12647>