Journal of Neuroscience Methods 360 (2021) 109239

Contents lists available at ScienceDirect

Journal of Neuroscience Methods

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

Application of a convolutional neural network for fully-automated detection of spike ripples in the scalp electroencephalogram

Jessica K. Nadalin [a], Uri T. Eden [a,b], Xue Han [b,c], R. Mark Richardson [d], Catherine J. Chu [e,1], Mark A. Kramer [a,b,*,1]

[a] Department of Mathematics and Statistics, Boston University, Boston, MA 02215, United States
[b] Center for Systems Neuroscience, Boston University, Boston, MA 02215, United States
[c] Department of Biomedical Engineering, Boston University, Boston, MA 02215, United States
[d] Department of Neurosurgery, Massachusetts General Hospital, Boston, MA 02114, United States
[e] Department of Neurology, Massachusetts General Hospital, Boston, MA 02114, United States

ARTICLE INFO

Keywords:
EEG
high frequency oscillations
ripples
convolutional neural network

ABSTRACT

Background: A reliable biomarker to identify cortical tissue responsible for generating epileptic seizures is required to guide prognosis and treatment in epilepsy. Combined spike ripple events are a promising biomarker for epileptogenic tissue that currently require expert review for accurate identification. This expert review is time consuming and subjective, limiting reproducibility and high-throughput applications.
New method: To address this limitation, we develop a fully-automated method for spike ripple detection. The method consists of a convolutional neural network trained to compute the probability that a spectrogram image contains a spike ripple.
Results: We validate the proposed spike ripple detector on expert-labeled data and show that this detector accurately separates subjects with low and high seizure risks.
Comparison with Existing Method: The proposed method performs as well as existing methods that require manual validation of candidate spike ripple events. The introduction of a fully automated method reduces subjectivity and increases rigor and reproducibility of this epilepsy biomarker.
Conclusion: We introduce and validate a fully-automated spike ripple detector to support utilization of this epilepsy biomarker in clinical and translational work.

1. Introduction

A reliable biomarker to identify cortical tissue responsible for generating epileptic seizures is required to guide prognosis and treatment in epilepsy (Engel, 2011). Ripples - low amplitude (~0.1 μV), high frequency (80-250 Hz), short (~10 ms), focal (~1 cm^2) events - have emerged as a promising biomarker of epileptogenic brain tissue (Frauscher et al., 2017). Clinical investigations in patients with epilepsy have shown that ripples occur more frequently inside the seizure onset zone (Andrade-Valença et al., 2012; Cimbalnik et al., 2018; Dümpelmann et al., 2015; Guragain et al., 2018; Jacobs et al., 2008; Liu et al., 2018; Worrell et al., 2008), and that removing brain tissue generating ripples improves surgical outcome (Jacobs et al., 2008, 2010; Modur et al., 2011; Okanishi et al., 2014; Otárula et al., 2019; van Klink et al., 2014; Wang et al., 2013; Wu et al., 2010). Electrophysiological experiments and computational studies provide mechanistic context for these clinical results, identifying the cellular and network properties that produce – and potentially disrupt – ripples (Jiruska et al., 2017; Köhling & Staley, 2011; Stacey et al., 2009; Traub et al., 2001).

Despite the promise of this biomarker, at least two important challenges prevent widespread application of ripples in clinical practice. First, the current gold standard for detection of ripples is manual review of electroencephalogram (EEG) recordings (e.g., through direct visualization of the EEG (Cepeda et al., 2020; Klotz et al., 2021; Nariai et al., 2019) or as validation following an automated procedure (Boran et al., 2019)). However, this manual inspection is difficult, subjective, and prohibitively time consuming; manual review of 10 minutes of data can take an expert reviewer up to 15 hours per electrode channel

* Corresponding author at: Department of Mathematics and Statistics, Boston University, Boston, MA 02215, United States.
E-mail address: mak@bu.edu (M.A. Kramer).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.jneumeth.2021.109239
Received 11 January 2021; Received in revised form 17 May 2021; Accepted 30 May 2021
Available online 4 June 2021

(Urrestarazu et al., 2007). This existing gold-standard both introduces reviewer bias and prevents more widespread clinical investigation. To address these limitations, many automated and semi-automated ripple detection methods have been developed (Blanco et al., 2010; Charupanit & Lopour, 2017; Gardner et al., 2007; Gliske et al., 2016; Jacobs et al., 2018; Zelmann et al., 2012). However, technical challenges remain in the application of these methods, including limiting false positive detections (Amiri et al., 2016; Bénar et al., 2010; Gliske et al., 2020), and calibrating detectors for optimal performance (Sindhu et al., 2020).

Whether manual or automated ripple detection methods improve clinical treatment of patients with epilepsy also remains unclear. A recent multicenter prospective trial, applying both manual and automated ripple detections to identify tissue with high rates of ripple activity, did not find an improved outcome at the individual level (Jacobs et al., 2018). The failure of this prospective trial was at least partially attributed to a second challenge: accurately separating physiologic from pathologic ripples. Physiologic ripples appear in human cortical tissue and are thought to be necessary for normal memory processes (Buzsáki & Silva, 2012; Joo & Frank, 2018; Kucewicz et al., 2014; Matsumoto et al., 2013). These physiologic ripples are indistinguishable from pathologic ripples in spectral frequency, amplitude and duration (Alkawadri et al., 2014; Ellenrieder et al., 2016; Liu et al., 2018; Malinowska et al., 2015; Matsumoto et al., 2013). Thus, separating pathologic ripples from physiologic ripples remains a significant - and clinically important - challenge (Bénar et al., 2010; Cimbalnik et al., 2018; Jacobs & Schönberger, 2019; Matsumoto et al., 2013; Ren et al., 2019).

One strategy to address these challenges is to focus on ripples that co-occur with epileptiform spikes, a hallmark of epilepsy with high disease specificity (Ayala et al., 1973; Engel, 2012; Sabolek et al., 2012; Staley & Dudek, 2006; Worrell & Gotman, 2011; Chu et al., 2017, Kramer et al., 2019). Spikes are large amplitude (~100 μV), long lasting (~100 ms) deviations in brain voltage activity that, like ripples, occur between seizures. Although spikes and ripples are understood to represent separate neurophysiological events, in spiking channels the majority of ripples co-occur with spikes (Frauscher et al., 2017). Observations from both non-invasive (Kramer et al., 2019; van Klink, van't Klooster, et al., 2016) and invasive (Gliske et al., 2018; Jacobs et al., 2018; Wang et al., 2013, 2017) recordings suggest that ripples co-occurring with spikes provide a more focal and specific biomarker for seizures than spikes or ripples alone. A focus on synchronous spike and ripple events, "spike ripples", addresses both of the major issues limiting progress in ripple research. First, by selecting ripples that coincide with interictal spikes, pathologic ripples are disentangled from physiologic ripples (Thomschewski et al., 2019). Second, compared to detection of ripples alone, the combined spike ripple events provide additional features for automated detection, enabling better detector performance. Current approaches to identify spike ripple events require manual categorization by expert reviewers (Chu et al., 2017; Kobayashi et al., 2010; Kramer et al., 2019; van Klink, Frauscher, et al., 2016; van Klink, van't Klooster, et al., 2016). An accurate fully-automated method, which operates directly on observed brain voltage recordings, would avoid subjectivity and inefficiency, further enhance reproducibility and facilitate rapid, high-throughput analysis of this promising biomarker in large data sets.

In this manuscript, we introduce a fully-automated spike ripple detection method. To do so, we create a neural network model that leverages a pre-trained convolutional neural network for image data. We apply this method to scalp voltage recordings from patients with epilepsy and show that this method is fast, accurate, and flexible; and that this method accurately separates subjects with active epilepsy from those with a low risk of future seizures and healthy controls.

2. Materials and Methods

2.1. Human subject data

In this work, we re-use the human subject data described in (Kramer et al., 2019); the subject recruitment, EEG acquisition, and EEG preparation match the procedures described in (Kramer et al., 2019). Briefly, we utilize EEG data (2035 Hz sampling rate) from 34 subjects: 21 children diagnosed with a self-limited epilepsy syndrome, childhood epilepsy with centrotemporal spikes (CECTS, aged 4.9–16.8 years, 17 males), and 13 control subjects (aged 8.7–14.3 years, 5 males). We divide the subjects with epilepsy into two groups: those with high seizure risk (active CECTS, A-CECTS, defined as having had a seizure within the last 12 months); and (ii) low seizure risk (resolved CECTS, R-CECTS, defined as seizure-free for at least 12 months) (Ross et al., 2019). Two children with CECTS returned after a minimum of 12 months for repeat evaluations. For each subject, one EEG channel was selected from each hemisphere for analysis. If interictal spikes were present, then the channel in which the spike amplitude was maximal was selected. If no spikes were observed, the C3 and C4 electrodes were selected, as these electrodes are most commonly involved in this focal epilepsy syndrome (Koutroumanidis et al., 2017). See (Kramer et al., 2019) for a description of detailed clinical data for each subject.

2.2. Manual spike ripple detection

For manually marked spike ripples, we utilized a dataset of 10 subjects analyzed in (Chu et al., 2017). For each subject, an expert (CJC) manually inspected 10 minutes of data from one EEG channel, and identified 151 spike ripple events in total; see (Chu et al., 2017) for details.

2.3. Semi-automated spike ripple detection

In (Chu et al., 2017), a semi-automated method was proposed to detect ripples that co-occur with spikes. Given a brain voltage signal (e.g., the scalp EEG) the method first identifies candidate time intervals of increased high-frequency activity, then computes seven features for each candidate event. These features require that the candidate event contain high-frequency fluctuations approximately sinusoidal in shape, with at least three cycles, that co-occur with a large amplitude discharge, consistent with existing notions of a regular and persistent high frequency oscillation initiating on the ascent of an interictal spike (van Klink, van't Klooster, et al., 2016; von Ellenrieder et al., 2014; Zelmann et al., 2009; Zijlmans et al., 2009). The method successfully detects spike ripple events, with high positive predictive value, low false

Table 1
Low and high decision thresholds result in high classification performance. The positive predictive value (PPV), negative predictive value (NPV), sensitivity, and specificity for the automated spike ripple detector, and 95% confidence intervals. These values represent cumulative results from each subject i as test data, where the CNN is trained and validated on all patients excluding subject i. Both the optimal decision threshold from the ROC curve ($d_s = 0.298$) and higher decision threshold ($d_s = 0.95$) produce consistent results.

Decision threshold	Low [optimal ROC] ($d_s = 0.298$)	High ($d_s = 0.95$)
PPV (95% CI)	0.86 (0.841, 0.890)	0.96 (0.947, 0.980)
NPV (95% CI)	0.98 (0.977, 0.988)	0.92 (0.910, 0.932)
Sensitivity (95% CI)	0.95 (0.932, 0.965)	0.74 (0.704, 0.770)
Specificity (95% CI)	0.95 (0.943, 0.962)	0.99 (0.987, 0.995)

positive rate, and high intra-rater reliability (Chu et al., 2017). Application of this method to scalp EEG recordings from subjects with focal childhood epilepsy showed that spike ripples are more spatially restricted than spikes, and that reduced spike ripple rates correlate with epilepsy resolution (Kramer et al., 2019). While this semi-automated method is less time consuming (requiring approximately 2 minutes of manual analysis per 10 minute recording) than most existing ripple detection procedures, the method still requires validation by a human expert. In this manuscript, we utilize the data and expert classifications in (Kramer et al., 2019) to train and validate a fully-automated spike ripple detector. In contrast to the semi-automated detector, this new fully-automated detector does not require subsequent expert validation of results, as we show the results are on par with expert classifications.

2.4. Spectral analysis

To compute spectrograms, we implement the following procedure. For each 1 s of EEG data, we compute the spectrum in 0.2 s sliding subintervals (overlap 0.195 s). Within each subinterval, we subtract the signal mean, apply the Hanning taper, and then compute the spectrum. We smooth the resulting spectrogram in time using a Gaussian lowpass filter of size $\pm$25 ms. We display the two-dimensional spectrogram as an image of size 163 $\times$ 44, corresponding to times 0.1 s to 0.9 s (spacing 0.0049 s), and frequencies 30 Hz to 245 Hz (spacing 5 Hz) using the default *jet* colormap in MATLAB.

2.5. Convolutional neural network

To identify spectrograms consistent with spike ripples we develop a spike ripple detector from a convolutional neural network (CNN). We note that spike ripples are well-characterized by their spectrograms: typically, they consist of an island of power in the high frequency range for a duration of about 0.1 s, along with high power in the low frequency range over the entire course of the spectrogram (example in Fig. 1A, see also (Amiri et al., 2016; Kobayashi et al., 2010; Kramer et al., 2019; van Klink, van't Klooster, et al., 2016)). Classifying spike ripples via their spectrogram images motivates the use of a CNN, a natural choice for image data (LeCun et al., 2015). We note that while other methods, such as support vector machines, could be applied, these methods require explicit feature specification. In contrast, a CNN requires no feature specification, instead learning the features of interest from the data. We opt here for the CNN method, as the features that define a spike ripple remain incompletely understood. As these characteristic islands of power resemble natural images, we use the architecture ResNet (He et al., 2016), previously trained on ImageNet data (*ImageNet*, 2020).

We use transfer learning to train the CNN (i.e., we retrain layers of this existing architecture on the spectrogram image training data). This allows us to leverage the increased performance from a CNN with many layers without requiring the time and computing power to learn these weights from scratch. We maintain all convolutional layers from ResNet, and modify the final linear layers for binary classification. In the first round of training, we fix the parameters of the convolutional layers and tune only the final linear layers. For the remaining duration of training, we tune parameters of all layers (Kornblith et al., 2019; *Vision.Learner | Fastai*, 2020). To illustrate the computational advantage of transfer learning, we randomly initialize all weights in the network and train the entire network using the labeled spectrograms for 1000 epochs on an 8 GB GPU (30 GB RAM). Doing so required 2 hours and 13 minutes, and failed to achieve a validation loss below 0.1. In contrast, using transfer learning for the same set of labeled spectrograms on the same machine, we achieve a validation loss of 0.04 in 4 minutes, after 18 epochs. We conclude that transfer learning considerably decreases computation time and increases performance.

The CNN uses gradient descent to minimize the loss, which we take to be cross-entropy, a typical choice for binary classification problems. We train with a batch size of 8 images, and use differential learning rates for different groups of layers in the ResNet architecture: we use 1e-4 as the learning rate for the initial layers, 1e-3 for the middle layers, and 1e-2 for the final layers. In this way, we use a smaller learning rate for the earlier layers, which pick out simple shapes and edges, and a larger learning rate for the later layers, which pick out more complicated visual features. We additionally use stochastic gradient descent with restarts (SGDR) in training, wherein we gradually decrease the learning rate over the course of a cycle with cosine annealing, and increase the length of this cycle by a factor of 2 as training continues. This prevents solutions from becoming trapped in local minima. At the end of each training epoch, if the validation loss has not improved in the past 6 epochs, we stop training and use the weights from the epoch with the lowest validation loss in our final model.

In training, ¼ of the training data is set aside for validation, and training is stopped early when the validation loss fails to decrease, which prevents overfitting. The training data is augmented to artificially expand the size of the training set via random horizontal flips and random scaling by a factor of up to 1.2. The input images are 44 x 44 pixels. The resulting CNN receives as input a spectrogram from a 1 s segment of EEG data and outputs the probability of that segment containing a spike ripple. This probability is obtained by taking the softmax function of the final 2 output nodes of the network.

2.6. Training, validation, and testing datasets

The data consist of spectrograms computed from 1 s segments of EEG data from two electrodes for each subject. We use leave-one-out cross-validation to assess the ability of the proposed method to detect spike

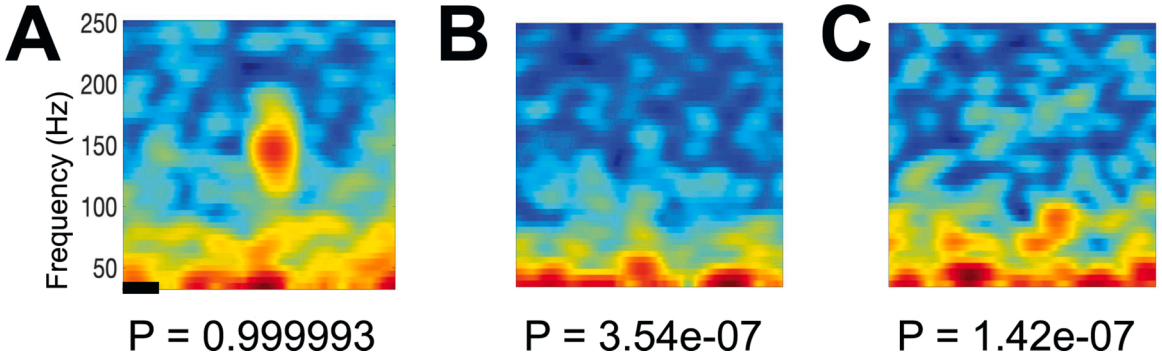

Fig. 1. The CNN outputs probabilities for candidate spectrograms in agreement with expert labels. Example spectrograms and corresponding CNN output probabilities (P) for candidate events **(A)** mutually labeled "yes" by both experts, with a spectral island evident at ~150 Hz, **(B)** mutually labeled "no" by both experts, and **(C)** undetected by an existing semi-automated spike ripple detector and expert validated with a "no" label. All three examples calculated for the same subject with active epilepsy. The spectrograms display power (in decibels, scaled 0 to 1 separately for each image) as a function of frequency and time; scale bar indicates 0.1 s.

ripples. That is, for each subject *i*, we create a training set composed of spectrograms from all subjects excluding *i*, and a test set composed of spectrograms exclusively from subject *i*.

For each patient, we divide the EEG signal(s) into 1 s segments with 0.5 s overlap, and compute the spectrogram of each segment. We choose the segment and overlap durations for two reasons. First, the segment is long enough to encompass the approximate duration of a spike (approximately 200 ms). Second, the segment is short enough to contain at most one spike ripple (maximum rate < 0.33 Hz) (Kramer et al., 2019). Half of the training set are these spectrograms from time segments identified as candidate spike ripples by the semi-automated detector and validated by two human experts; these spectrograms are assigned a "yes" label in the training set. One quarter of the training set are spectrograms from time segments identified as candidate spike ripples by the semi-automated detector and rejected by both experts; these spectrograms are assigned a "no" label in the training set. We proceed in this way, rather than using the original centered spectrograms from the semi-automated detector as our training data, so that the algorithm learns to detect spike ripples at all time points in the spectrogram, rather than only identifying spike ripples centered in the spectrogram. The remaining 25% of the training set are spectrograms from time segments undetected by the semi-automated detector, validated by an expert to not contain a spike ripple, and assigned a "no" label. In this way, we create a balanced training set of spectrograms: 50% "yes" labels and 50% "no" labels.

The test set for each subject *i* consists of the subset of time segments for subject *i* detected by the semi-automated detector and validated by both experts ("yes" label), rejected by both experts ("no" label), or undetected by the semi-automated detector and verified not to contain a spike ripple ("no" label).

On average, for each subject, the training data consists of 1664 images (minimum 953, maximum 1729). The test data consists of an average of 93 images (minimum 51, maximum 441). In total, the test data from all expert labeled spectrograms from all subjects consists of 2243 images.

For analysis of manually marked spike ripples, we first train the CNN spike ripple detector using the expert classified spectrograms detected and undetected by the semi-automated detector, as described above, excluding subject *i*. We then divide the entire EEG recording from patient *i* into 1 s segments (0.5 s overlap), compute the spectrogram for each segment, and input these spectrograms to the trained CNN. We compare the output probabilities to the manually marked spike ripple events, and repeat this process for each subject.

The trained CNN, and the procedure to train the CNN, is available for reuse and further development at: https://github.com/Eden-Kramer-Lab/CNN_Spectrogram_Algorithm

2.7. Statistical analysis

We compare the outputs of the test set data to the expert classifications and compute measures of detector performance: sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), false positive rate, the F1 score (i.e., the harmonic mean of the precision and recall), and the area under the curve (AUC) for the receiver operating characteristics (ROC) curve over all subjects. The false positive rate is defined as the number of false positive spike ripple detections per second. We compute confidence intervals for the sensitivity, specificity, PPV, and NPV using the R-package DTComPair (Stock and Hielscher, 2014).

2.8. Assessment of classifier performance

To assess classifier performance, we perform a randomized labels test (Ojala et al., 2009), in which we randomly permute the labels of the data 1000 times (i.e., we permute the "yes" and "no" labels assigned to each spectrogram). We then evaluate the detector performance using these randomized labels in comparison to the detector performance using the original, non-permuted labels.

3. Results

3.1. The fully automated spike ripple detector performs well on expert-labeled data

We propose an automated procedure to detect spike ripples in scalp EEG recordings. Our goal is to replace the time-consuming procedure of visual inspection with an automated approach that has comparable classification performance to expert validation. To do so, we train a convolutional neural network (CNN) to identify the unique spectral signatures of a spike ripple. In general, a spectrogram of a spike ripple consists of two components: (i) increased power across a broad range of low frequencies due to the spike, and (ii) increased power across a narrow interval of high frequencies due to the ripple - a "spectral island" (Fig. 1). To train and evaluate the detector, we analyze scalp EEG data recorded from $M = 34$ subjects (21 subjects with epilepsy and 13 control subjects, see *Materials and Methods: Human subject data*). Each subject has a set of labeled spectrograms, visually marked by two experts to either contain a spike ripple, or not (see *Materials and Methods: Training, validation, and testing datasets)*.

To test the performance of the CNN, we utilize these data to perform leave-one-out cross-validation: for each subject *i*, we set aside the corresponding N_i expert labeled spectrograms $\{s_i^1, \ldots, s_i^{N_i}\}$ as the test data, and use the expert labeled data from all other subjects $\{s_1^1, \ldots, s_1^{N_1}, \ldots, s_{i-1}^1, \ldots, s_{i-1}^{N_{i-1}}, s_{i+1}^1, \ldots, s_{i+1}^{N_{i+1}}, \ldots, s_M^1, \ldots, s_M^{N_M}\}$ as the training data. We train the spike ripple detector, and from this trained classifier obtain probabilities for the test data $\{p_i^1, \ldots, p_i^{N_i}\}$, where p_i^k is the probability of spectrogram s_i^k containing a spike ripple (Fig. 1).

Repeating this process for each subject, we obtain a collection of probabilities $\{p_1^1, \ldots, p_1^{N_1}, \ldots, p_M^1, \ldots, p_M^{N_M}\}$ for all spectrograms from each subject. To assess detector performance, we first compute the area under the ROC curve (AUC, Fig. 2); the AUC is large (0.99) consistent with excellent classification performance. Performing a randomized labels test (Ojala et al., 2009) (see *Materials and Methods: Assessment of classifier performance*), we find $p = 0.001$, suggesting high significance for this result. We then choose a decision threshold d_s such that if $p_i^k < d_s$, we conclude that spectrogram *k* from subject *i* does not contain a spike ripple, and if $p_i^k \geq d_s$, we conclude that spectrogram *k* from subject *i* does contain a spike ripple. The choice of d_s is based on the relative importance of limiting false positives or false negatives. A higher value of d_s

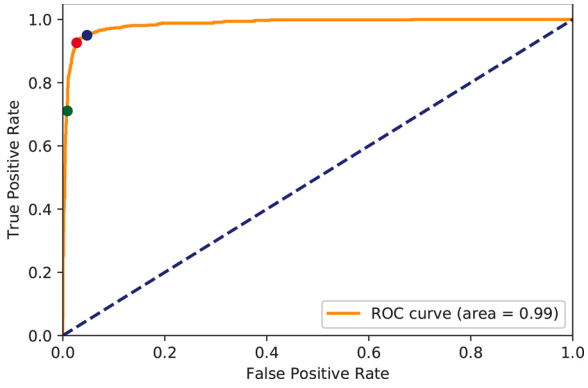

Fig. 2. The ROC curve shows strong diagnostic capability of the automated CNN spike ripple detector. The ROC curve for different decision thresholds d_s has an area under the curve (AUC) of 0.99. The optimal decision threshold ($d_s = 0.298$, blue circle) occurs at a false positive rate of 0.048 and true positive rate of 0.950. A decision threshold of 0.95 (green circle) reduces the true and false positive rates.

yields a more conservative estimate, limiting detections to those with high probabilities, and hence decreasing false positives. Alternatively, a lower value yields a more liberal estimate and decreases the number of false negatives.

To illustrate detector performance, we consider two decision thresholds d_s and compute four measures of detector performance: positive predictive value (PPV), negative predictive value (NPV), specificity, and sensitivity. Using the optimal value from the ROC curve, $d_s = 0.298$, we find excellent detector performance: high values (≥ 0.86) for all four measures (Table 1). Alternatively, we may desire detections that utilize a stricter interpretation of spike ripples, i.e., only time segments with very high probabilities are classified as spike ripples. Repeating this analysis with a higher decision threshold ($d_s = 0.95$, green dot in Fig. 2), we find consistent results (Table 1).

Though disagreements between expert and automated classifications comprise the minority of cases (Table 1), these cases provide insights into factors that contribute to the CNN classification. As expected, when the CNN and experts agree on the presence of a spike ripple (examples in Fig. 3A and B), we find a well-characterized spectral island in the spectrogram and large amplitude spike in the unfiltered EEG signal. Similarly, when the CNN and experts agree on the absence of a spike ripple (examples Fig. 3C and D), we find no evidence for a spike ripple in either the spectrogram or EEG signal. In examples where the CNN fails to detect a spike ripple identified by experts (Fig. 3E and F), we observe weak evidence of a spectral island (i.e., the power is low in the ripple frequency range). In an example where the CNN erroneously detects a spike ripple, the spectral island is clear but the large amplitude deviation in the EEG signal is too wide to be considered a spike (Fig. 3G). We also show a rare case consistent with an error in expert classification (Fig. 3H); here the CNN identifies a spike ripple missed in expert classification.

We conclude that the proposed automated CNN spike ripple detector accurately classifies spectrograms computed from EEG data. Given examples classified by expert reviewers as containing a spike ripple or not, the detector performs with high sensitivity and specificity. This high performance occurs for both a low and high decision threshold.

3.2. The automated spike ripple detector performs well against manual markings

In the previous section, we trained and analyzed the CNN spike ripple detector using examples identified by the semi-automated method in (Kramer et al., 2019), and then validated by human experts. As a second test of detector performance, we compare the fully-automated detections to the current gold standard approach - manual review. Manual, visual inspection, in which an expert reviews the entire recording and evaluates all spikes in 0.6 s increments in three different visualizations (raw time series, filtered time series, spectrogram) is highly time consuming, but more sensitive to detect spike ripple events than the semi-automated approach (Chu et al., 2017), and therefore provides a more accurate estimate of detector performance. We apply the trained CNN spike ripple detector to EEG data from 10 subjects with manually marked spike ripples (see *Materials and Methods: Manual spike ripple detection*), and compute the probability of each 1 s time segment containing a spike ripple (example in Fig. 4). We then use the precision-recall curve to choose a decision threshold where precision approximately equals recall ($d_s = 0.78$), and classify each 1 s time segment k as having a spike ripple if $p^k \geq d_s$ or not having a spike ripple if $p^k < d_s$. Comparing these classifications to the expert manual detections, we find a sensitivity of 0.627 (CI [0.56, 0.69]), positive

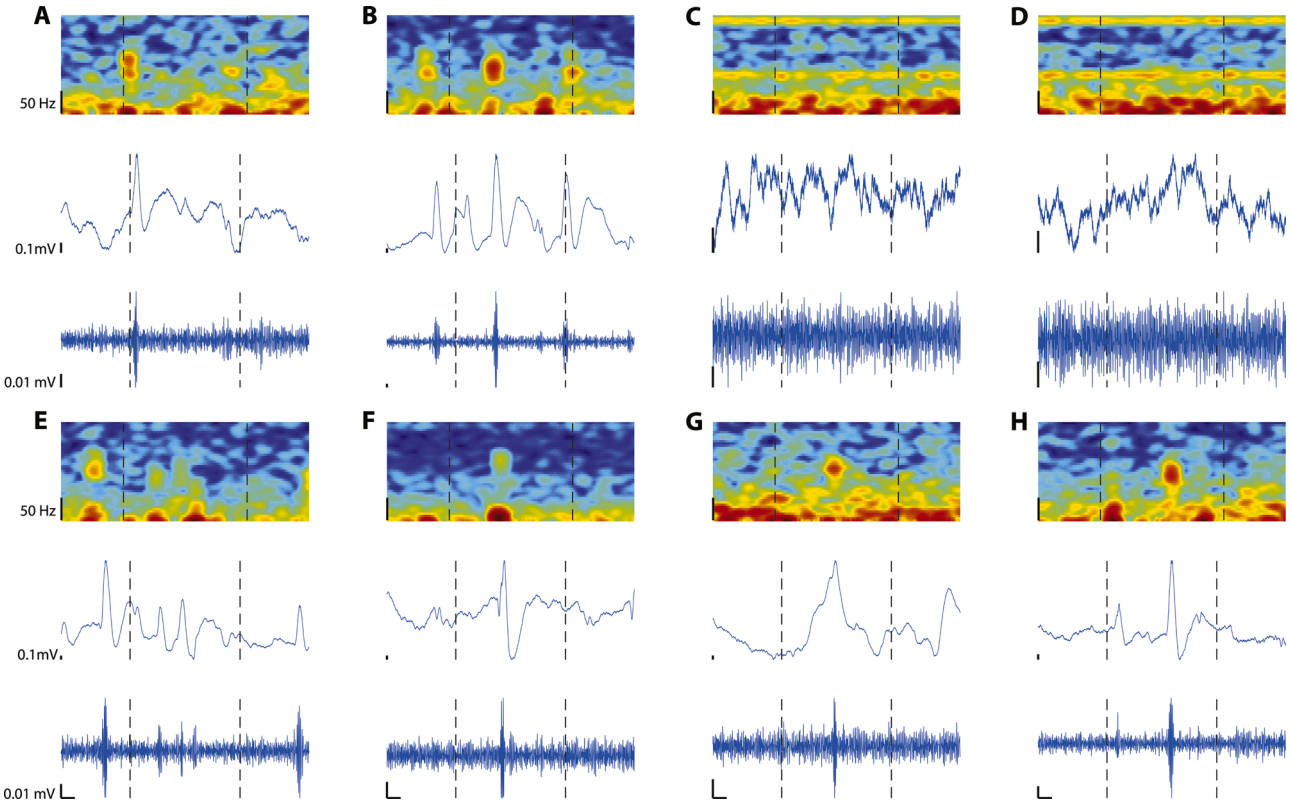

Fig. 3. Examples of agreement and disagreement between the CNN spike ripple detector and expert classifications. Example spectrograms (top), unfiltered EEG signals (middle), and filtered high frequency signals (bottom, highpass filtered to 100-300 Hz) from EEG excerpts classified by the CNN spike ripple detector. **(A, B)** The detector and expert classifications agree on the presence of a spike ripple (detector output probabilities $P > 0.99$). **(C,D)** The detector and expert classifications agree on the absence of a spike ripple ($P < 1e\text{-}12$). **(E-H)** Examples of detector and expert classification disagreement. **(E,F)** Example false negatives; the spike ripple is not detected ($P < 1e\text{-}2$). **(G,H)** Example false positives; a spike ripple is detected ($P > 0.99$). In all figures, the horizontal scale bars indicate 0.1 s, and the vertical dashed lines indicate the 1 s time intervals classified by the CNN spike ripple detector.

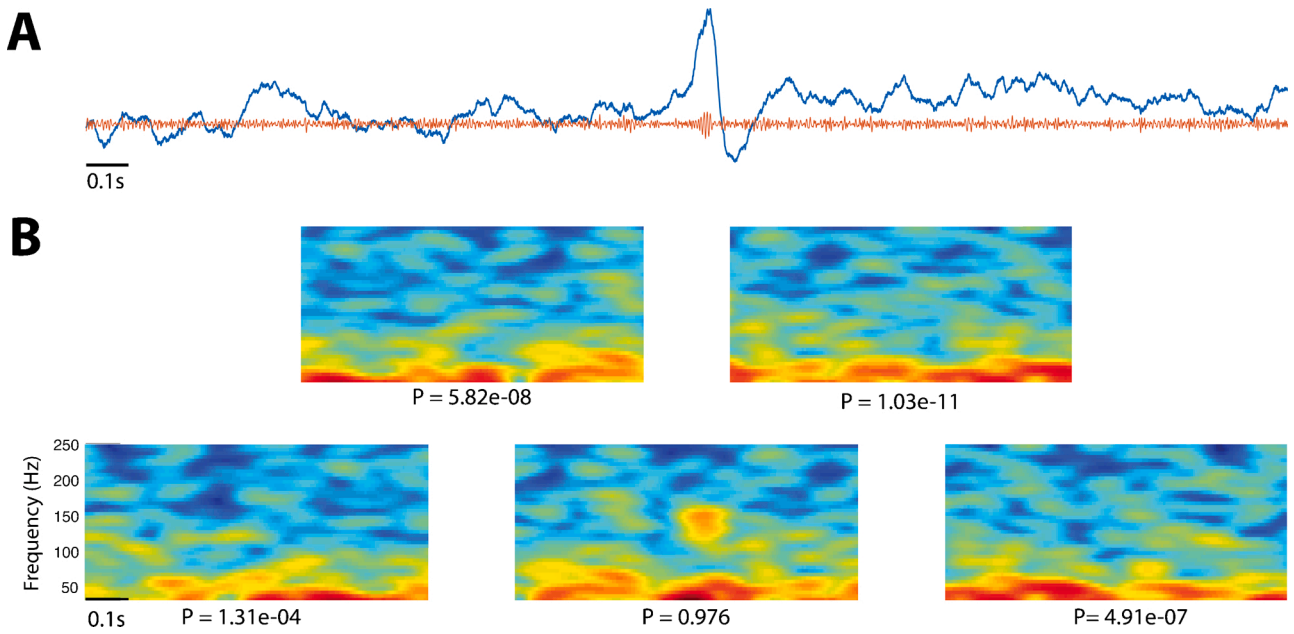

Fig. 4. Example spectrograms computed from the EEG of a subject, and the corresponding outputs from the spike ripple detector. (A) Example raw (blue) and high frequency (orange, 100-300 Hz) signals from a subject's EEG signal, and **(B)** the corresponding 1 s spectrograms with 0.5 s overlap, with probability outputs (P) from the spike ripple detector.

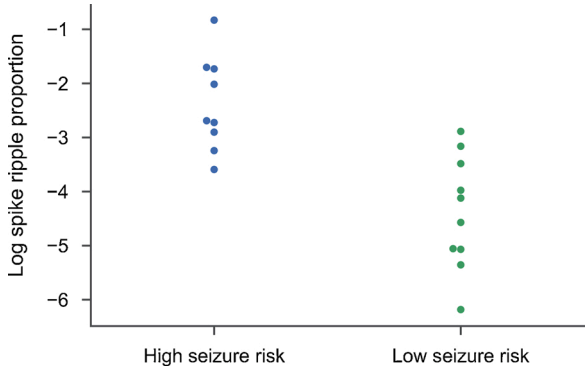

Fig. 5. The spike ripple proportion is higher in subjects with high seizure risk. The proportion of 1 s segments containing a spike ripple for high and low risk seizure groups with $d_s = 0.298$.

predictive value (PPV) of 0.630 (CI [0.57, 0.69]), and false positive rate of 0.007. We note that these values are consistent with the performance of the existing semi-automated method in (Chu et al., 2017). However, the automated CNN spike ripple detector requires no manual validation by expert reviewers. We conclude that the automated detector performs well - comparable to human experts - while eliminating the time consuming and subjective process of visual inspection and classification.

3.3. The automated spike ripple detector differentiates between subjects with high and low seizure risk

We now consider application of the automated CNN spike ripple detector to a specific clinical goal: classification of disease severity. Here, motivated by clinical utility, we restrict consideration to children with CECTS divided into two groups: low risk of a future seizure (e.g., resolving epilepsy, R-CECTS) or high risk of a future seizure (e.g., active epilepsy, A-CECTS). We now test whether the CNN spike ripple detector differentiates between the high and low risk seizure groups.

To do so, we apply the trained CNN spike ripple detector to the EEG data of each CECTS subject i, and use the optimal decision threshold $d_s = 0.298$ to classify each segment as having a spike ripple or not. From this classified dataset, for each subject i we evaluate the spike ripple proportion: the number of spike ripples detected divided by the total number of 1 s segments analyzed for subject i. Visual inspection suggests that the spike ripple proportion tends to be higher in subjects with high seizure risk (Fig. 5).

We then categorize each patient using the optimal ROC value as the threshold to find performance consistent with the existing naive automated spike ripple detector reported in (Kramer et al., 2019); excellent PPV and specificity, and high NPV and sensitivity. Repeating this analysis with an increased spike ripple decision threshold ($d_s = 0.95$), we find improved classification performance (Table 2); excellent specificity and PPV, and improved sensitivity and NPV compared to the CNN spike ripple detector with lower decision threshold ($d_s = 0.298$), and the naive automated spike ripple detector reported in (Kramer et al., 2019). To summarize detector performance, we compute the F1 score and find the highest value using the CNN spike ripple detector with strict decision threshold (F1 = 0.94) compared to the other methods (CNN with lower decision threshold, F1 = 0.80; naive automated detector, F1 = 0.89; semi-automated detector, F1 = 0.91). We conclude that a strict classification of spike ripples, wherein only high-probability time segments are classified as positive, provides excellent performance in classifying subjects into low versus high risk seizure groups.

3.4. Spike ripple proportion decreases with duration seizure-free

As a second assessment of the detector's clinical utility, we compare the results of the CNN spike ripple detector to an existing measure of disease severity: duration seizure-free. The longer a patient remains seizure-free, the less likely that patient will have a subsequent seizure, independent of medication use (Berg et al., 2001; Ross et al., 2020; Sillanpää et al., 2017). Using the optimal decision threshold ($d_s = 0.298$), we find a negative relationship between the log spike ripple proportion and time since last seizure (p = 2.7e-6, slope = -0.15, 95% CI [-0.20, -0.10], Fig. 6). Consistent with the results in (Kramer et al., 2019), we conclude that a negative relationship exists between spike ripple proportion and duration seizure-free; the longer duration seizure-free, the lower the spike ripple proportion.

Table 2
Subject-level diagnostic characteristics for different spike ripple decision thresholds and detectors. The CNN spike ripple detector with a strict decision threshold improves classification of subjects with high and low seizure risk. Results for the naive automated detector and semi-automated detector from (Kramer et al., 2019).

Spike ripple detector	CNN, Low ($d_s = 0.298$)	CNN, High ($d_s = 0.95$)	Naive automated detector	Semi-automated detector
PPV (95% CI)	1 (1.0, 1.0)	1 (1.0,1.0)	1 (1.0, 1.0)	0.83 (0.62, 1.0)
NPV (95% CI)	0.77 (0.54, 1.0)	0.91 (0.74, 1.0)	0.87 (0.70, 1.0)	1 (1.0, 1.0)
Sensitivity (95% CI)	0.67 (0.36,1.0)	0.89 (0.69,1.0)	0.8 (0.55, 1.0)	1 (1.0, 1.0)
Specificity (95% CI)	1 (1.0,1.0)	1 (1.0, 1.0)	1 (1.0, 1.0)	0.85 (0.65, 1.0)

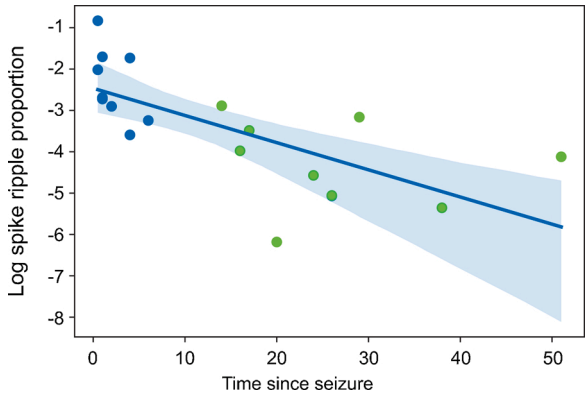

Fig. 6. The spike ripple proportion decreases with time since last seizure. Circles indicate patients with A-CECTS (blue) and R-CECTS (green). The blue line indicates the model fit with mean (solid) and 95% CIs (shaded).

3.5. The automated spike ripple detector is robust to added image noise

To evaluate the effect of noise on the CNN spike ripple detector, we construct new test sets for each patient by taking the original test set (Section 3.1) and adding Gaussian image noise. More specifically, for each image in the test set for all patients, we add Gaussian image noise with mean 0 and 20 different values of standard deviation $\sigma = \{0, 0.01, 0.02, \ldots, 0.1, 0.2, 0.3, \ldots, 1.0\}$, constructing 20 different test sets for each patient. Example noised images are shown in Fig. 7A. We then evaluate the resulting AUC for the ROC curve, as in Section 3.1, for the test sets at each level of σ (Fig. 7B) across all patients. We find that, as expected, detector performance decreases with increasing noise. However, the detector performs well when the noise is small (AUC > 0.9 for $\sigma \leq 0.1$), and maintains AUC > 0.686 even when the test image is considerably distorted by the image noise. We conclude that the automated CNN spike ripple detector is robust to noise in the spectrogram images.

3.6. Impact of sampling rate on detector performance

To evaluate how the sampling rate of the original EEG signal impacts the CNN spike ripple detector, we analyze the manually marked EEG data (Section 3.2) sampled at different frequencies (2035 Hz, the original sampling frequency; 1017.5 Hz, 407 Hz, and 254.375 Hz computed using the function *decimate* in MATLAB). As in Section 3.2, we create 1 s spectrograms from these signals, use these spectrograms as input to the CNN spike ripple detector, and compute the probability of each spectrogram containing a spike ripple. We calculate the optimized decision threshold d_s for each sampling frequency via the precision-recall curve, as in Section 3.2, and use d_s to predict the presence of spike ripples. Comparing these classifications to the expert manual detections, we compute the PPV, sensitivity, and FPR (Table 3) for each sampling frequency. While we find similar performance for data sampled at 2035 Hz and 1017.5 Hz, the performance worsens for the lower sampling frequencies 407 Hz and 254.375 Hz, as expected (Chu et al., 2017). We conclude that the CNN spike ripple detector performs best on data collected at sampling rates above 1000 Hz.

4. Discussion

In this manuscript, we introduced a new method to automatically detect an emerging electrographic biomarker for epilepsy, spike ripples. To do so, we trained a convolutional neural network to identify spectrogram images with the unique features of a spike ripple. We showed that the spike ripple detector performed well against expert-labeled data, and accurately predicted seizure risk. The method is fully-automated, removing the subjective and time-consuming visual inspection required by existing spike ripple detection methods. Expert validation is no longer required on the output from the detector, as we have shown the performance from the detector alone is comparable to this expert validation. The method requires no additional labeled data, as the model has been pre-trained on the data described here, and can be simply applied to future unlabeled data.

The spike ripple detector presented here differs from existing approaches in the following ways. First, we focus on automated classification of spike ripples, rather than ripples (or high-frequency oscillations) alone; for example, many machine learning algorithms exist - each analyzing a select feature set - to identify ripples (Amiri et al., 2016; Blanco et al., 2010; Migliorelli et al., 2020; Sciaraffa et al., 2020). While selecting specific features for ripple detection improves interpretability, doing so may also limit performance. To address this, neural networks have been recently developed to detect ripples without requiring explicit feature selection. In (Zuo et al., 2019), the authors developed a CNN to identify ripples, training the network on a small number of filtered EEG signals, stacked to create two-dimensional images. In (Lai et al., 2019), the authors use a CNN to classify two-dimensional time-frequency maps, and find improved detector performance. In (Hagen et al., 2020), the authors implement a recurrent neural network to detect sharp-wave ripples, training the network on the raw local field potential. In the method presented here, we use the spectrograms of EEG data as input to the CNN; rather than requiring explicit time and frequency characteristics to classify spike ripples, the model learns these characteristics from the spectrogram images.

The methodology of the spike ripple detector proposed here is most similar to the ripple detector developed in (Lai et al., 2019). In that work, a CNN classifies 2D time-frequency maps to identify ripples. While similar, the spike ripple detector developed here differs from the ripple detector in (Lai et al., 2019) in three ways. First, we focus on detection of the composite electrographic phenomena of spike ripples, not ripples alone. With this difference, we expect our method to overcome the challenges in separating pathologic from physiologic ripples (Thomschewski et al., 2019). Second, in (Lai et al., 2019) an initial detection, utilizing the data filtered into high-frequency bands, is performed to identify times of candidate ripples. Then, only the time-frequency maps at these candidate times are input to the CNN and classified. Here, we perform no initial step to identify candidate spike ripples; instead, the spectrograms for each 1 s interval (0.5 s overlap) are input to the CNN. Third, in (Lai et al., 2019) the CNN architecture is built and trained from scratch, and is limited to two convolutional and two pooling layers. In contrast, we use transfer learning, which allows us to utilize the performance of a complex 34-layer CNN without requiring the time and computing power to learn these weights from scratch. We note that

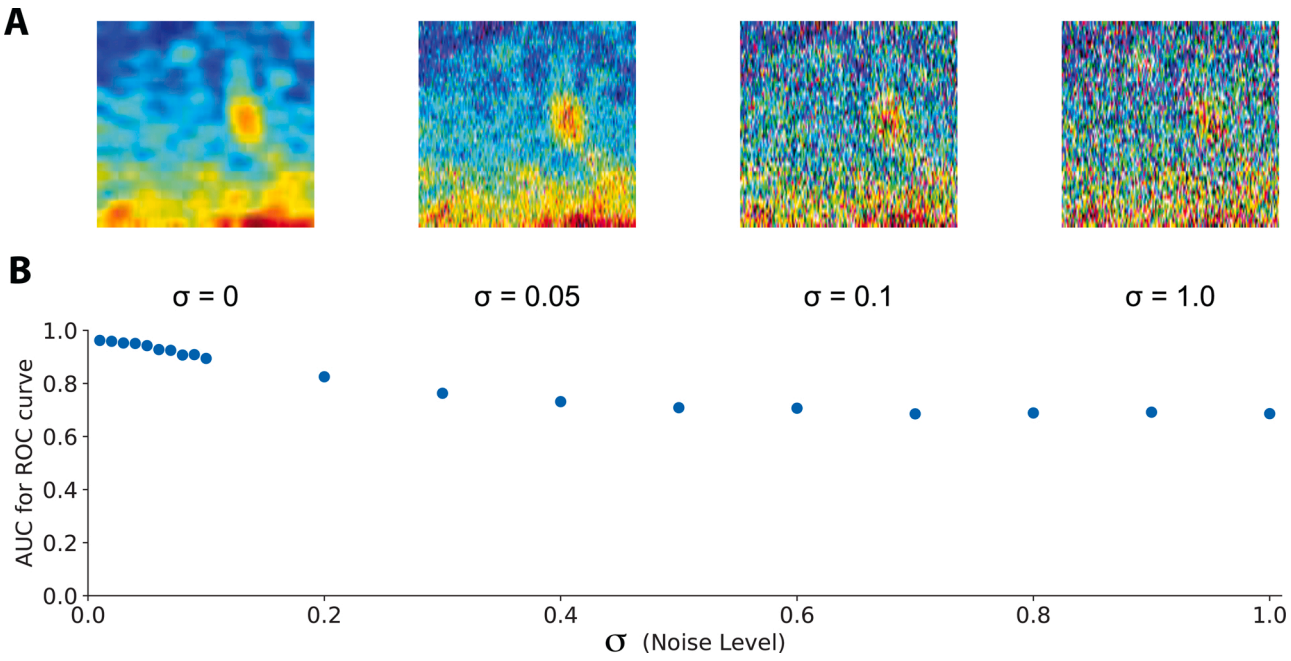

Fig. 7. Automated detections persist for noisy spectrogram images. (A) An example spectrogram distorted with different Gaussian image noise (standard deviation σ). **(B)** Detector performance (AUC for the ROC curve) versus noise level (σ).

Table 3
Performance of the CNN spike ripple detector is high at higher sampling frequencies, and low at lower sampling frequencies. Positive predictive value (PPV), sensitivity, and false positive rate (FPR) on test sets from identical signals sampled at different frequencies.

Sampling Frequency (Hz)	PPV	Sensitivity	FPR
2035	0.790 [0.736, 0.844]	0.786 [0.732, 0.840]	0.088
1017.5	0.781 [0.726, 0.836]	0.777 [0.722, 0.832]	0.092
407	0.656 [0.592, 0.719]	0.653 [0.589, 0.716]	0.142
254.375	0.341 [0.277, 0.406]	0.340 [0.275, 0.404]	0.258

CNNs have been applied in other contexts to assess spectrogram images of neural data, e.g., for rapid eye movement behavior disorder (Ruffini et al., 2019), and epileptiform spike detection (Johansen et al., 2016). Again, our methodology differs from these examples in that we utilize transfer learning to leverage the power of a CNN trained on millions of natural images for our model.

While the automated detection of spike ripples motivated the detector developed here, we note that the framework presented is highly flexible: theoretically, the detector could be repurposed to detect any signal pattern well-characterized by its spectrogram, with a reasonable amount of training data (on the order of 1000 labeled images). With the architecture completely intact, one would simply provide new training data to re-train the model using the same approach described here. We expect this approach would work best for spectrograms with distinct features resembling natural images (in our case, isolated regions of increased power formed simple, connected shapes). For example, with training data that identified ripples in different frequency bands, the spike ripple detector could be extended to distinguish between the spectral islands of ripples (80-250 Hz) and fast-ripples (250-500 Hz) (Blanco et al., 2010; Sciaraffa et al., 2020; Zuo et al., 2019). For more complex spectra, e.g., from audio data, it may be better to use an underlying architecture trained on non-natural images such as satellite imagery, rather than the ResNet architecture employed here.

While the spike ripple detector performed well in the application to EEG data considered here, we note the following limitations. First, we trained the detector using spike ripple (and non-spike ripple) events classified by two human experts. We treated these human expert classifications as true, despite the necessarily subjective nature of these classifications through visual inspection. We note that additional spike ripple classifications - performed by other human experts on different data sets - may be used to retrain the spike ripple detector. Doing so may further improve detector performance and limit the impact of subjective classifications by different human experts. More generally, the spike ripple detector supports a feedback approach to increase performance of the model: experts can analyze and validate or edit the output labels of test spectrograms, and include these newly labeled spectrograms in the training data for future iterations of the model.

Second, the spike ripple classification is not perfect. For example, we observed that sharp artifacts in the signal were sometimes incorrectly classified as spike ripples. A sharp event increases power at all frequencies, producing vertical streaks in the spectrogram. In some cases, these vertical streaks qualitatively resemble the islands of power in spectrograms of spike ripples (Amiri et al., 2016). To address this, future work could pre-process the data to remove common artifacts, and hence decrease the false positive rate. Alternatively, the model could be re-trained using the spectrograms of these artifacts, assigned a “no” label.

Third, in training and applying the CNN spike ripple detector, we made numerous choices; we utilized data sampled at a specific rate (2035 Hz), with a specific reference montage (average reference), and during a specific sleep stage (non-REM sleep) (Kramer et al., 2019). In previous work, the semi-automated spike ripple detector was shown to perform consistently across different sampling rates, reference montages, and states of consciousness (Chu et al., 2017). We showed here that automated CNN spike ripple detections persist when the sampling rate is high enough (above 1000 Hz, Table 3), and despite noise in the spectrogram images (Fig. 7). However, how these choices and others - alone and in combination - impact performance of the automated CNN spike ripple detector remains a topic for further investigation. We recommend analyzing data at the highest sampling rate available (i.e., without downsampling) during non-REM sleep, when muscle artifacts are less common.

Fourth, we note that any classification method utilizing neural networks is subject to the black box problem: it is particularly difficult to interpret the millions of trained parameters in the network. While we

assume that the model is identifying spectral islands in the spectrograms to classify spike ripples, we cannot verify this claim. An understanding of precisely what the detector identifies in spectrograms could provide insight into mechanisms supporting the pathology. Much research exists on the interpretation of neural networks (Benitez et al., 1997; Ghorbani et al., 2018; Montavon et al., 2018), and future work could focus on interpreting the trained CNN spike ripple detector. For example, feature visualization (Yosinski et al., 2015), which seeks to visualize the activations in different layers, may perhaps identify the activations of some layers that specifically target islands of power in the spectrogram. Additionally, saliency maps could be used to identify which regions of the spectrogram, and correspondingly which patterns in time-frequency space, have the largest impact on classification (Mundhenk et al., 2020). We expect saliency maps to identify the brief, spectral island at high frequencies, and broad increase in power at low frequencies, as regions that strongly impact classification. However, saliency maps may also reveal other features in the spectrogram important for classification. Understanding these features may improve detector performance, and provide additional insight into the spectral features of spike ripples.

Here we present a reliable, fully automated method to detect a promising biomarker of seizure risk, spike ripples. The method has a single, intuitive parameter to adjust - the decision threshold - resulting in higher positive predictive value or sensitivity. We showed that the spike ripple proportion produced by the detector separates subjects with low and high seizure risk. Such a diagnostic tool has the potential to aid clinicians in their determination of disease severity and treatment, and support high-throughput analysis of large data sets.

Declaration of Competing Interest

CJC consults for SleepMed Inc and Biogen Inc on EEG interpretation and analysis. MAK consults for Biogen Inc on EEG interpretation and analysis.

CRediT authorship contribution statement

Jessica K. Nadalin: Conceptualization, Methodology, Software, Validation, Formal analysis, Visualization, Writing - original draft. **Uri T. Eden:** Conceptualization, Methodology, Funding acquisition. **Xue Han:** Writing - review & editing, Funding acquisition. **R. Mark Richardson:** Writing - review & editing, Funding acquisition. **Catherine J. Chu:** Conceptualization, Investigation, Supervision, Resources, Data curation, Writing - original draft, Funding acquisition. **Mark A. Kramer:** Conceptualization, Methodology, Supervision, Data curation, Visualization, Writing - original draft, Funding acquisition.

Acknowledgements

M.A.K acknowledges support by the National Science Foundation DMS Award #1451384, National Institutes of Health NINDS R01NS110669 and R01NS119483. J.K.N. acknowledges support by the National Science Foundation GRFP. C.J.C acknowledges support by the National Institutes of Health NINDS R01NS119483.

References

Alkawadri, R., Gaspard, N., Goncharova, I.I., Spencer, D.D., Gerrard, J.L., Zaveri, H., Duckrow, R.B., Blumenfeld, H., Hirsch, L.J., 2014. The spatial and signal characteristics of physiologic high frequency oscillations. Epilepsia 55 (12), 1986–1995. https://doi.org/10.1111/epi.12851.

Amiri, M., Lina, J.-M., Pizzo, F., Gotman, J., 2016. High Frequency Oscillations and spikes: Separating real HFOs from false oscillations. Clinical Neurophysiology 127 (1), 187–196. https://doi.org/10.1016/j.clinph.2015.04.290.

Andrade-Valença, L., Mari, F., Jacobs, J., Zijlmans, M., Olivier, A., Gotman, J., Dubeau, F., 2012. Interictal high frequency oscillations (HFOs) in patients with focal epilepsy and normal MRI. Clinical Neurophysiology 123 (1), 100–105. https://doi.org/10.1016/j.clinph.2011.06.004.

Ayala, G.F., Dichter, M., Gumnit, R.J., Matsumoto, H., Spencer, W.A., 1973. Genesis of epileptic interictal spikes. New knowledge of cortical feedback systems suggests a neurophysiological explanation of brief paroxysms. Brain Research 52, 1–17. https://doi.org/10.1016/0006-8993(73)90647-1.

Bénar, C.G., Chauviere, L., Bartolomei, F., Wendling, F., 2010. Pitfalls of high-pass filtering for detecting epileptic oscillations: A technical note on "false" ripples. Clinical Neurophysiology 121 (3), 301–310. https://doi.org/10.1016/j.clinph.2009.10.019.

Benitez, J.M., Castro, J.L., Requena, I., 1997. Are artificial neural networks black boxes? IEEE Transactions on Neural Networks 8 (5), 1156–1164. https://doi.org/10.1109/72.623216.

Berg, A.T., Shinnar, S., Levy, S.R., Testa, F.M., Smith-Rapaport, S., Beckerman, B., Ebrahimi, N., 2001. Two-Year Remission and Subsequent Relapse in Children with Newly Diagnosed Epilepsy. Epilepsia 42 (12), 1553–1562. https://doi.org/10.1046/j.1528-1157.2001.21101.x.

Blanco, J.A., Stead, M., Krieger, A., Viventi, J., Marsh, W.R., Lee, K.H., Worrell, G.A., Litt, B., 2010. Unsupervised Classification of High-Frequency Oscillations in Human Neocortical Epilepsy and Control Patients. Journal of Neurophysiology 104 (5), 2900–2912. https://doi.org/10.1152/jn.01082.2009.

Boran, Ece, Sarnthein, Johannes, Krayenbühl, Niklaus, Ramantani, Georgia, Fedele, Tommaso, 2019. High-Frequency Oscillations in Scalp EEG Mirror Seizure Frequency in Pediatric Focal Epilepsy. Scientific Reports 9 (no. 1), 1–10.

Buzsáki, G., da Silva, F.L., 2012. High frequency oscillations in the intact brain. Progress in Neurobiology 98 (3), 241–249. https://doi.org/10.1016/j.pneurobio.2012.02.004.

Cepeda, Carlos, Levinson, Simon, Nariai, Hiroki, Yazon, Vannah-Wila, Tran, Conny, Barry, Joshua, Oikonomou, Katerina D., et al., 2020. Pathological High Frequency Oscillations Associate with Increased GABA Synaptic Activity in Pediatric Epilepsy Surgery Patients. Neurobiology of Disease 134, 104618.

Charupanit, K., Lopour, B.A., 2017. A Simple Statistical Method for the Automatic Detection of Ripples in Human Intracranial EEG. Brain Topography 127, 1–15. https://doi.org/10.1007/s10548-017-0579-6.

Chu, C.J., Chan, A., Song, D., Staley, K.J., Stufflebeam, S.M., Kramer, M.A., 2017. A semi-automated method for rapid detection of ripple events on interictal voltage discharges in the scalp electroencephalogram. Journal of Neuroscience Methods 277, 46–55. https://doi.org/10.1016/j.jneumeth.2016.12.009.

Cimbalnik, J., Brinkmann, B., Kremen, V., Jurak, P., Berry, B., Gompel, J.V., Stead, M., Worrell, G., 2018. Physiological and pathological high frequency oscillations in focal epilepsy. Annals of Clinical and Translational Neurology 5 (9), 1062–1076. https://doi.org/10.1002/acn3.618.

Dümpelmann, M., Jacobs, J., Schulze-Bonhage, A., 2015. Temporal and spatial characteristics of high frequency oscillations as a new biomarker in epilepsy. Epilepsia 56 (2), 197–206. https://doi.org/10.1111/epi.12844.

Ellenrieder, N. von, Frauscher, B., Dubeau, F., Gotman, J., 2016. Interaction with slow waves during sleep improves discrimination of physiologic and pathologic high-frequency oscillations (80–500 Hz). Epilepsia 57 (6), 869–878. https://doi.org/10.1111/epi.13380.

Engel, J., 2011. Biomarkers in epilepsy: Introduction. Biomarkers in Medicine 5 (5), 537–544. https://doi.org/10.2217/bmm.11.62.

Engel, J., 2012. Seizures and Epilepsy, Second Edition. Oxford University Press.

Frauscher, B., Bartolomei, F., Kobayashi, K., Cimbalnik, J., Klooster, M.A., Rampp, S., Otsubo, H., Höller, Y., Wu, J.Y., Asano, E., Engel, J., Kahane, P., Jacobs, J., Gotman, J., 2017. High-frequency oscillations: The state of clinical research. Epilepsia 98 (1). https://doi.org/10.1111/epi.13829 n/a-n/a.

Gardner, A.B., Worrell, G.A., Marsh, E., Dlugos, D., Litt, B., 2007. Human and automated detection of high-frequency oscillations in clinical intracranial EEG recordings. Clinical Neurophysiology: Official Journal of the International Federation of Clinical Neurophysiology 118 (5), 1134–1143. https://doi.org/10.1016/j.clinph.2006.12.019.

Ghorbani, A., Abid, A., Zou, J., 2018. Interpretation of Neural Networks is Fragile. ArXiv: 1710.10547 [Cs, Stat]. http://arxiv.org/abs/1710.10547.

Gliske, S.V., Irwin, Z.T., Chestek, C., Hegeman, G.L., Brinkmann, B., Sagher, O., Garton, H.J.L., Worrell, G.A., Stacey, W.C., 2018. Variability in the location of high frequency oscillations during prolonged intracranial EEG recordings. Nature Communications 9 (1), 1–14. https://doi.org/10.1038/s41467-018-04549-2.

Gliske, S.V., Irwin, Z.T., Davis, K.A., Sahaya, K., Chestek, C., Stacey, W.C., 2016. Universal automated high frequency oscillation detector for real-time, long term EEG. Clinical Neurophysiology 127 (2), 1057–1066. https://doi.org/10.1016/j.clinph.2015.07.016.

Gliske, S.V., Qin, Z., Lau, K., Alvarado-Rojas, C., Salami, P., Zelmann, R., Stacey, W.C., 2020. Distinguishing false and true positive detections of high frequency oscillations. Journal of Neural Engineering 17 (5), 056005. https://doi.org/10.1088/1741-2552/abb89b.

Guragain, H., Cimbalnik, J., Stead, M., Groppe, D.M., Berry, B.M., Kremen, V., Kenney-Jung, D., Britton, J., Worrell, G.A., Brinkmann, B.H., 2018. Spatial variation in high-frequency oscillation rates and amplitudes in intracranial EEG. Neurology 90 (8), e639–e646. https://doi.org/10.1212/WNL.0000000000004998.

Hagen, E., Chambers, A.R., Einevoll, G.T., Pettersen, K.H., Enger, R., Stasik, A.J., 2020. RippleNet: A Recurrent Neural Network for Sharp Wave Ripple (SPW-R) Detection. BioRxiv. https://doi.org/10.1101/2020.05.11.087874, 2020.05.11.087874.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep Residual Learning for Image Recognition. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 770–778. https://doi.org/10.1109/CVPR.2016.90.

ImageNet, 2020. ImageNet. http://www.image-net.org/.

Jacobs, J., LeVan, P., Chander, R., Hall, J., Dubeau, F., Gotman, J., 2008. Interictal high-frequency oscillations (80-500 Hz) are an indicator of seizure onset areas independent of spikes in the human epileptic brain. Epilepsia 49 (11), 1893–1907. https://doi.org/10.1111/j.1528-1167.2008.01656.x.

Jacobs, J., Schönberger, J., 2019. In search of epileptic scalp high-frequency oscillations. Clinical Neurophysiology 130 (7), 1172–1174. https://doi.org/10.1016/j.clinph.2019.04.006.

Jacobs, J., Wu, J.Y., Perucca, P., Zelmann, R., Mader, M., Dubeau, F., Mathern, G.W., Schulze-Bonhage, A., Gotman, J., 2018. Removing high-frequency oscillations: A prospective multicenter study on seizure outcome. Neurology 91 (11), e1040–e1052. https://doi.org/10.1212/WNL.0000000000006158.

Jacobs, J., Zijlmans, M., Zelmann, R., Chatillon, C.-É., Hall, J., Olivier, A., Dubeau, F., Gotman, J., 2010. High-frequency electroencephalographic oscillations correlate with outcome of epilepsy surgery. Annals of Neurology 67 (2), 209–220. https://doi.org/10.1002/ana.21847.

Jiruska, P., Alvarado-Rojas, C., Schevon, C.A., Staba, R., Stacey, W., Wendling, F., Avoli, M., 2017. Update on the mechanisms and roles of high-frequency oscillations in seizures and epileptic disorders. Epilepsia 98 (Suppl. 6). https://doi.org/10.1111/epi.13830, 250-n/a.

Johansen, A.R., Jin, J., Maszczyk, T., Dauwels, J., Cash, S.S., Westover, M.B., 2016. Epileptiform spike detection via convolutional neural networks. 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) 754–758. https://doi.org/10.1109/ICASSP.2016.7471776.

Joo, H.R., Frank, L.M., 2018. The hippocampal sharp wave–ripple in memory retrieval for immediate use and consolidation. Nature Reviews Neuroscience 19 (12), 744–757. https://doi.org/10.1038/s41583-018-0077-1.

Kobayashi, K., Watanabe, Y., Inoue, T., Oka, M., Yoshinaga, H., Ohtsuka, Y., 2010. Scalp-recorded high-frequency oscillations in childhood sleep-induced electrical status epilepticus. Epilepsia 51 (10), 2190–2194. https://doi.org/10.1111/j.1528-1167.2010.02565.x.

Köhling, R., Staley, K., 2011. Network mechanisms for fast ripple activity in epileptic tissue. Epilepsy Research 97 (3), 318–323. https://doi.org/10.1016/j.eplepsyres.2011.03.006.

Kornblith, S., Shlens, J., Le, Q.V., 2019. Do Better ImageNet Models Transfer Better? ArXiv:1805.08974 [Cs, Stat]. http://arxiv.org/abs/1805.08974.

Koutroumanidis, M., Arzimanoglou, A., Caraballo, R., Goyal, S., Kaminska, A., Laoprasert, P., Oguni, H., Rubboli, G., Tatum, W., Thomas, P., Trinka, E., Vignatelli, L., Moshé, S.L., 2017. The role of EEG in the diagnosis and classification of the epilepsy syndromes: A tool for clinical practice by the ILAE Neurophysiology Task Force (Part 1). Epileptic Disorders 19 (3), 233–298. https://doi.org/10.1684/epd.2017.0935.

Klotz, Kerstin A., Sag, Yusuf, Schönberger, Jan, Jacobs, Julia, 2021. Scalp Ripples Can Predict Development of Epilepsy After First Unprovoked Seizure in Childhood. Annals of Neurology 89 (1), 134–142.

Kramer, M.A., Ostrowski, L.M., Song, D.Y., Thorn, E.L., Stoyell, S.M., Parnes, M., Chinappen, D., Xiao, G., Eden, U.T., Staley, K.J., Stufflebeam, S.M., Chu, C.J., 2019. Scalp recorded spike ripples predict seizure risk in childhood epilepsy better than spikes. Brain 142 (5), 1296–1309. https://doi.org/10.1093/brain/awz059.

Kucewicz, M.T., Cimbalnik, J., Matsumoto, J.Y., Brinkmann, B.H., Bower, M.R., Vasoli, V., Sulc, V., Meyer, F., Marsh, W.R., Stead, S.M., Worrell, G.A., 2014. High frequency oscillations are associated with cognitive processing in human recognition memory. Brain 137 (8), 2231–2244. https://doi.org/10.1093/brain/awu149.

Lai, D., Zhang, X., Ma, K., Chen, Z., Chen, W., Zhang, H., Yuan, H., Ding, L., 2019. Automated Detection of High Frequency Oscillations in Intracranial EEG Using the Combination of Short-Time Energy and Convolutional Neural Networks. IEEE Access 7, 82501–82511. https://doi.org/10.1109/ACCESS.2019.2923281.

LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep Learning. Nature 521 (7553), 436–444.

Liu, S., Gurses, C., Sha, Z., Quach, M.M., Sencer, A., Bebek, N., Curry, D.J., Prabhu, S., Tummala, S., Henry, T.R., Ince, N.F., 2018. Stereotyped high-frequency oscillations discriminate seizure onset zones and critical functional cortex in focal epilepsy. Brain 141 (3), 713–730. https://doi.org/10.1093/brain/awx374.

Malinowska, U., Bergey, G.K., Harezlak, J., Jouny, C.C., 2015. Identification of seizure onset zone and preictal state based on characteristics of high frequency oscillations. Clinical Neurophysiology: Official Journal of the International Federation of Clinical Neurophysiology 126 (8), 1505–1513. https://doi.org/10.1016/j.clinph.2014.11.007.

Matsumoto, A., Brinkmann, B.H., Matthew Stead, S., Matsumoto, J., Kucewicz, M.T., Marsh, W.R., Meyer, F., Worrell, G., 2013. Pathological and physiological high-frequency oscillations in focal human epilepsy. Journal of Neurophysiology 110 (8), 1958–1964. https://doi.org/10.1152/jn.00341.2013.

Migliorelli, C., Bachiller, A., Alonso, J.F., Romero, S., Aparicio, J., Jacobs, J., Mananas, M.A., San Antonio-Arce, V., 2020. SGM: A novel time-frequency algorithm based on unsupervised learning improves high-frequency oscillation detection in epilepsy. Journal of Neural Engineering. https://doi.org/10.1088/1741-2552/ab8345.

Modur, P.N., Zhang, S., Vitaz, T.W., 2011. Ictal high-frequency oscillations in neocortical epilepsy: Implications for seizure localization and surgical resection. Epilepsia 52 (10), 1792–1801. https://doi.org/10.1111/j.1528-1167.2011.03165.x.

Montavon, G., Samek, W., Müller, K.-R., 2018. Methods for interpreting and understanding deep neural networks. Digital Signal Processing 73, 1–15. https://doi.org/10.1016/j.dsp.2017.10.011.

Mundhenk, Nathan, T., Chen, Barry Y., Friedland, Gerald, 2020. Efficient Saliency Maps for Explainable AI. ArXiv:1911.11293 [Cs]. http://arxiv.org/abs/1911.11293.

Nariai, Hiroki, Hussain, Shaun A., Bernardo, Danilo, Fallah, Aria, Murata, Kristina K., Nguyen, Jimmy C., Rajaraman, Rajsekar R., et al., 2019. Prospective Observational Study: Fast Ripple Localization Delineates the Epileptogenic Zone. Clinical Neurophysiology 130 (November (11)), 2144–2152.

Ojala, Markus, Garriga, Gemma C., 2009. Permutation Tests for Studying Classifier Performance. In: 2009 Ninth IEEE International Conference on Data Mining. Miami Beach, FL, USA: IEEE, pp. 908–913. https://doi.org/10.1109/ICDM.2009.108.

Okanishi, T., Akiyama, T., Tanaka, S.-I., Mayo, E., Mitsutake, A., Boelman, C., Go, C., Snead, O.C., Drake, J., Rutka, J., Ochi, A., Otsubo, H., 2014. Interictal high frequency oscillations correlating with seizure outcome in patients with widespread epileptic networks in tuberous sclerosis complex. Epilepsia 55 (10), 1602–1610. https://doi.org/10.1111/epi.12761.

Otárula, K.A.G., von Ellenrieder, N., Cuello-Oderiz, C., Dubeau, F., Gotman, J., 2019. High-Frequency Oscillation Networks and Surgical Outcome in Adult Focal Epilepsy. Annals of Neurology 85 (4), 485–494. https://doi.org/10.1002/ana.25442.

Ren, S., Gliske, S.V., Brang, D., Stacey, W.C., 2019. Redaction of false high frequency oscillations due to muscle artifact improves specificity to epileptic tissue. Clinical Neurophysiology 130 (6), 976–985. https://doi.org/10.1016/j.clinph.2019.03.028.

Ross, E.E., Stoyell, S.M., Kramer, M.A., Berg, A.T., Chu, C.J., 2019. The natural history of seizures and neuropsychiatric symptoms in BECTS: results from a large prospective, longitudinal US cohort.

Ross, E.E., Stoyell, S.M., Kramer, M.A., Berg, A.T., Chu, C.J., 2020. The natural history of seizures and neuropsychiatric symptoms in childhood epilepsy with centrotemporal spikes (CECTS). Epilepsy & Behavior 103. https://doi.org/10.1016/j.yebeh.2019.07.038.

Ruffini, G., Ibañez, D., Castellano, M., Dubreuil-Vall, L., Soria-Frisch, A., Postuma, R., Gagnon, J.-F., Montplaisir, J., 2019. Deep Learning With EEG Spectrograms in Rapid Eye Movement Behavior Disorder. Frontiers in Neurology 10. https://doi.org/10.3389/fneur.2019.00806.

Sabolek, H.R., Swiercz, W.B., Lillis, K.P., Cash, S.S., Huberfeld, G., Zhao, G., Ste Marie, L., Clemenceau, S., Barsh, G., Miles, R., Staley, K.J., 2012. A candidate mechanism underlying the variance of interictal spike propagation. J Neurosci 32 (9), 3009–3021. https://doi.org/10.1523/JNEUROSCI.5853-11.2012.

Sciaraffa, N., Klados, M.A., Borghini, G., Di Flumeri, G., Babiloni, F., Aricò, P., 2020. Double-Step Machine Learning Based Procedure for HFOs Detection and Classification. Brain Sciences 10 (4), 220. https://doi.org/10.3390/brainsci10040220.

Sillanpää, M., Schmidt, D., Saarinen, M.M., Shinnar, S., 2017. Remission in epilepsy: How long is enough? Epilepsia 58 (5), 901–906. https://doi.org/10.1111/epi.13732.

Sindhu, R., Staba, R., Lopour, B.A., 2020. Trends in the use of automated algorithms for the detection of high-frequency oscillations associated with human epilepsy. Epilepsia 61 (8), 1553–1569. https://doi.org/10.1111/epi.16622.

Stacey, W.C., Lazarewicz, M.T., Litt, B., 2009. Synaptic Noise and Physiological Coupling Generate High-Frequency Oscillations in a Hippocampal Computational Model. Journal of Neurophysiology 102 (4), 2342–2357. https://doi.org/10.1152/jn.00397.2009.

Staley, K.J., Dudek, F.E., 2006. Interictal spikes and epileptogenesis. Epilepsy Currents / American Epilepsy Society 6 (6), 199–202. https://doi.org/10.1111/j.1535-7511.2006.00145.x.

Stock, C., Hielscher, T., 2014. DTComPair: comparison of binary diagnostic tests in a paired study design. R package version 1.0.3. http://CRAN.R-project.org/package=DTComPair.

Thomschewski, A., Hincapié, A.-S., Frauscher, B., 2019. Localization of the Epileptogenic Zone Using High Frequency Oscillations. Frontiers in Neurology 10. https://doi.org/10.3389/fneur.2019.00094.

Traub, R.D., Whittington, M.A., Buhl, E.H., LeBeau, F.E.N., Bibbig, A., Boyd, S., Cross, H., Baldeweg, T., 2001. A Possible Role for Gap Junctions in Generation of Very Fast EEG Oscillations Preceding the Onset of, and Perhaps Initiating, Seizures. Epilepsia 42 (2), 153–170. https://doi.org/10.1046/j.1528-1157.2001.26900.x.

Urrestarazu, E., Chander, R., Dubeau, F., Gotman, J., 2007. Interictal high-frequency oscillations (100–500 Hz) in the intracerebral EEG of epileptic patients. Brain 130 (9), 2354–2366. https://doi.org/10.1093/brain/awm149.

van Klink, N., Frauscher, B., Zijlmans, M., Gotman, J., 2016. Relationships between interictal epileptic spikes and ripples in surface EEG. Clinical Neurophysiology : Official Journal of the International Federation of Clinical Neurophysiology 127 (1), 143–149. https://doi.org/10.1016/j.clinph.2015.04.059.

van Klink, N., van't Klooster, M.A., Leijten, F.S.S., Jacobs, J., Braun, K.P.J., Zijlmans, M., 2016. Ripples on rolandic spikes: A marker of epilepsy severity. Epilepsia. https://doi.org/10.1111/epi.13423 n/a-n/a.

van Klink, N., van't Klooster, M.A., Zelmann, R., Leijten, F.S.S., Ferrier, C.H., Braun, K.P.J., van Rijen, P.C., van Putten, M.J.A.M., Huiskamp, G.J.M., Zijlmans, M., 2014. High frequency oscillations in intra-operative electrocorticography before and after epilepsy surgery. Clinical Neurophysiology 125 (11), 2212–2219. https://doi.org/10.1016/j.clinph.2014.03.004.

Vision.learner | fastai, 2020. https://docs.fast.ai/vision.learner.html.

von Ellenrieder, N., Beltrachini, L., Perucca, P., Gotman, J., 2014. Size of cortical generators of epileptic interictal events and visibility on scalp EEG. NEUROIMAGE 94, 47–54. https://doi.org/10.1016/j.neuroimage.2014.02.032.

Wang, S., So, N.K., Jin, B., Wang, I.Z., Bulacio, J.C., Enatsu, R., Dai, S., Chen, Z., Gonzalez-Martinez, J., Najm, I.M., 2017. Interictal ripples nested in epileptiform discharge help to identify the epileptogenic zone in neocortical epilepsy. Clinical Neurophysiology 128 (6), 945–951. https://doi.org/10.1016/j.clinph.2017.03.033.

Wang, S., Wang, I.Z., Bulacio, J.C., Mosher, J.C., Gonzalez-Martinez, J., Alexopoulos, A. V., Najm, I.M., So, N.K., 2013. Ripple classification helps to localize the seizure-onset zone in neocortical epilepsy. Epilepsia 54 (2), 370–376. https://doi.org/10.1111/j.1528-1167.2012.03721.x.

Worrell, G.A., Gardner, A.B., Stead, S.M., Hu, S., Goerss, S., Cascino, G.J., Meyer, F.B., Marsh, R., Litt, B., 2008. High-frequency oscillations in human temporal lobe: Simultaneous microwire and clinical macroelectrode recordings. Brain 131 (4), 928–937. https://doi.org/10.1093/brain/awn006.

Worrell, G., Gotman, J., 2011. High-frequency oscillations and other electrophysiological biomarkers of epilepsy: Clinical studies. Biomarkers in Medicine 5 (5), 557–566. https://doi.org/10.2217/bmm.11.74.

Wu, J.Y., Sankar, R., Lerner, J.T., Matsumoto, J.H., Vinters, H.V., Mathern, G.W., 2010. Removing interictal fast ripples on electrocorticography linked with seizure freedom in children (e–Pub ahead of print). Neurology 75 (19), 1686–1694. https://doi.org/10.1212/WNL.0b013e3181fc27d0.

Yosinski, J., Clune, J., Nguyen, A., Fuchs, T., Lipson, H., 2015. Understanding Neural Networks Through Deep Visualization. ArXiv:1506.06579 [Cs]. http://arxiv.org/abs/1506.06579.

Zelmann, R., Mari, F., Jacobs, J., Zijlmans, M., Dubeau, F., Gotman, J., 2012. A comparison between detectors of high frequency oscillations. Clinical Neurophysiology 123 (1), 106–116. https://doi.org/10.1016/j.clinph.2011.06.006.

Zelmann, R., Zijlmans, M., Jacobs, J., Châtillon, C.-E., Gotman, J., 2009. Improving the identification of High Frequency Oscillations. Clinical Neurophysiology 120 (8), 1457–1464. https://doi.org/10.1016/j.clinph.2009.05.029.

Zijlmans, M., Jacobs, J., Zelmann, R., Dubeau, F., Gotman, J., 2009. High-frequency oscillations mirror disease activity in patients with epilepsy. Neurology 72 (11), 979–986. https://doi.org/10.1212/01.wnl.0000344402.20334.81.

Zuo, R., Wei, J., Li, X., Li, C., Zhao, C., Ren, Z., Liang, Y., Geng, X., Jiang, C., Yang, X., Zhang, X., 2019. Automated Detection of High-Frequency Oscillations in Epilepsy Based on a Convolutional Neural Network. Frontiers in Computational Neuroscience 13. https://doi.org/10.3389/fncom.2019.00006.