BRAIN
A JOURNAL OF NEUROLOGY

Crowdsourcing seizure detection: algorithm development and validation on human implanted device recordings

Steven N. Baldassano,[1,2] Benjamin H. Brinkmann,[3,4] Hoameng Ung,[1,2] Tyler Blevins,[1,2] Erin C. Conrad,[5] Kent Leyde,[6] Mark J. Cook,[7,8] Ankit N. Khambhati,[1,2] Joost B. Wagenaar,[2,5] Gregory A. Worrell[3,4] and Brian Litt[1,2,5]

There exist significant clinical and basic research needs for accurate, automated seizure detection algorithms. These algorithms have translational potential in responsive neurostimulation devices and in automatic parsing of continuous intracranial electroencephalography data. An important barrier to developing accurate, validated algorithms for seizure detection is limited access to high-quality, expertly annotated seizure data from prolonged recordings. To overcome this, we hosted a kaggle.com competition to crowdsource the development of seizure detection algorithms using intracranial electroencephalography from canines and humans with epilepsy. The top three performing algorithms from the contest were then validated on out-of-sample patient data including standard clinical data and continuous ambulatory human data obtained over several years using the implantable NeuroVista seizure advisory system. Two hundred teams of data scientists from all over the world participated in the kaggle.com competition. The top performing teams submitted highly accurate algorithms with consistent performance in the out-of-sample validation study. The performance of these seizure detection algorithms, achieved using freely available code and data, sets a new reproducible benchmark for personalized seizure detection. We have also shared a 'plug and play' pipeline to allow other researchers to easily use these algorithms on their own datasets. The success of this competition demonstrates how sharing code and high quality data results in the creation of powerful translational tools with significant potential to impact patient care.

1 Department of Bioengineering, University of Pennsylvania, Philadelphia, PA 19104, USA
2 Center for Neuroengineering and Therapeutics, University of Pennsylvania, Philadelphia, PA 19104, USA
3 Mayo Systems Electrophysiology Laboratory, Departments of Neurology and Biomedical Engineering, Mayo Clinic, Rochester, MN 55905, USA
4 Department of Neurology, Mayo Clinic and Mayo Foundation, Rochester, MN 55905, USA
5 Department of Neurology, University of Pennsylvania, PA, USA
6 NeuroVista, Seattle, WA, USA
7 St. Vincent's Hospital, Melbourne, VIC, Australia
8 Department of Medicine, University of Melbourne, Melbourne, VIC, Australia

Correspondence to: Steven N. Baldassano
301 Hayden Hall, 3320 Smith Walk, University of Pennsylvania, Philadelphia, PA 19104, USA
E-mail: stevennb@mail.med.upenn.edu

Introduction

Epilepsy is a common chronic neurological condition affecting 1–2% of the population (Sun *et al.*, 2008). Despite current treatment options, including anti-epileptic drugs and surgical resection, up to 30% of patients with epilepsy continue to have seizures (Kwan and Brodie, 2000). In addition, many patients with seizures successfully controlled by anti-epileptic drugs suffer from medication side-effects (Morrell, 2002). Many medically-refractory patients are not candidates for surgical resection due to poorly localized epileptic networks or involvement of eloquent cortex, and even ideal candidates remain seizure-free for a year after surgery in only about 65% of cases (Wiebe *et al.*, 2001).

Closed-loop stimulation devices are receiving significant attention as an alternate method of therapy for medication-resistant seizures. These devices are implanted to record intracranial electroencephalography (iEEG) data from depth or subdural electrodes. Seizure detection algorithms are used to identify potential seizure epochs, and electrical stimulation is delivered to arrest seizure propagation. Despite the invasive nature of these devices, this approach has the benefit of providing therapy only when needed to targeted brain structures with few side effects (Morrell, 2011). The FDA approved the first of these devices, the NeuroPace RNS system, in 2013. This device administers electrical stimulation during a detected seizure, decreasing seizure frequency by about 50% in most patients (Chabolla *et al.*, 2006). While this device demonstrates clinical benefit to patients, its impact may be limited by the accuracy of seizure detection. As it is necessary to detect seizure events early and with high sensitivity, this device suffers from high rates of false positive detections. These false positives cause unnecessary brain stimulations that can reduce battery life, increasing cost and patient discomfort due to more frequent surgical procedures to replace spent batteries. Addressing the shortcomings of existing online seizure detection strategies is an essential step toward design of effective, long-lasting implantable devices for treatment of epilepsy. Given the existing framework for incorporating new algorithms into implantable devices, generating novel, effective methods for detecting seizures in real-time would immediately contribute to a rapidly expanding anti-seizure device market.

In addition to their direct therapeutic utility, more accurate automated seizure detection algorithms would address a significant clinical and research burden. Physicians working in epilepsy often review large quantities of continuous EEG data from inpatient monitoring studies to identify seizures (Gutierrez-Colina *et al.*, 2012), which in some patients may be quite subtle (Abend *et al.*, 2010, 2011). In patients with ambulatory recording devices, physicians must review hundreds of hours of continuous recordings in order to generate accurate seizure diaries. For research applications, the need to manually identify seizures places a significant bottleneck on producing high quality datasets for analysis.

In contrast to the use case of implantable device recordings, in which early detection of seizures is critical, these applications require seizure detection algorithms that are highly sensitive and specific without undue regard for detection latency.

Developing patient-specific seizure onset detection algorithms remains challenging. Intractable epilepsy can result from an enormous variety of pathologies, including trauma, tumours, stroke, infection, cortical malformations, genetic causes, and medications, resulting in significant diversity in seizure onset patterns. Existing devices rely on manually tuning preset detection parameters, and often fail to capture individual seizure dynamics, even after repeated follow-up appointments for algorithm tuning. Researchers have pursued automated algorithms for detecting epileptic events from patient EEG recordings since the early 1970s (Tzallas *et al.*, 2012). These algorithms traditionally rely on selecting discriminative features that are extracted from the data, coupled with a classification strategy. While a variety of signal features have been used in prior studies (Ramgopal *et al.*, 2014), including morphology-based features [e.g. line length (Esteller *et al.*, 2001), halfwave (Gotman, 1982), area, principal component analysis], biologically-inspired features [e.g. cross-channel correlation (Liu *et al.*, 2002), synchronization (Altenburg *et al.*, 2003)], and frequency-domain features [e.g. FFT analysis (Temko *et al.*, 2011), wavelet transform (Pradhan *et al.*, 1996; Casson *et al.*, 2007)], there is no consensus regarding the optimal feature or set for detection. Similarly, classification approaches have ranged from simplistic feature thresholds to more complex machine learning classifiers such as support vector machines (Liu *et al.*, 2002; Acharya *et al.*, 2011; Kharbouch *et al.*, 2011; Temko *et al.*, 2011), random forests, and artificial neural networks (Alkan *et al.*, 2005; D'Alessandro *et al.*, 2005), but no clearly optimal classification method has yet been established.

An important handicap to developing and validating algorithms for seizure detection is that only a limited number of data scientists have access to high-quality, expertly annotated seizure data. Thus far, EEG data have been largely acquired from intensive care unit (ICU) monitoring (Osorio *et al.*, 2002; Wilson *et al.*, 2004; Gardner *et al.*, 2006), presurgical inpatient studies (So, 2000; Chung *et al.*, 2015), animal models (White *et al.*, 2006; Raghunathan *et al.*, 2009), and some implantable devices (Davis *et al.*, 2011). The utility of these data is limited by shortcomings such as short recording duration, few recorded seizures, degraded data quality from faulty electrodes, and concerns regarding the validity of animal models. Furthermore, datasets largely remain at their acquiring institutions, limiting seizure detection research and our ability to validate published algorithms (Wagenaar *et al.*, 2015). The lack of a standardized database of seizure data (Wagenaar *et al.*, 2015), as well as a common 'gold standard' for annotating seizure entities (Cui *et al.*, 2012), hampers our ability to compare detection accuracy across algorithms. This

challenge is not unique to epilepsy, as other fields have tackled the 'data barrier' with open data platforms such as OSF.io and PhysioNet. In the field of epilepsy, approaches to address this problem are gradually emerging, such as the data archive, Epilepsiae, and the collaborative, cloud-based platform http://ieeg.org, which our group created to encourage sharing of data, computational tools, and expertise among researchers (Wagenaar et al., 2013).

The seizure detection challenge

We hosted a kaggle.com competition, sponsored by the National Institutes of Health's National Institute of Neurological Disorders and Stroke (NINDS) and the American Epilepsy Society (AES), to develop the best personalized seizure detection algorithms. This platform allowed for a crowdsourced approach by providing easily accessible, annotated recordings to data scientists. Contestants were provided with labelled interictal and ictal training data and unlabelled testing data derived from four canine and eight human subjects (described in detail in the 'Materials and methods' section). The contestants applied machine learning techniques of their choosing in an attempt to accurately classify the testing data as interictal or ictal. The contest was carried out using kaggle.com's infrastructure for tracking algorithm submissions, maintaining a live-updating leaderboard, and evaluating performance on a limited hidden test set. This endeavour was designed to engage algorithm and machine learning experts across a wide array of fields by providing high-quality data in a format easily accessible to data scientists without detailed knowledge of epilepsy. At the conclusion of the seizure detection competition the same organizations hosted a parallel competition in seizure prediction (Brinkmann et al., 2016).

The competition was held from May to August 2014. At the end of the competition, the top three teams were awarded prizes of $5000, $2000, and $1000. These teams' algorithms were further tested on a more extensive archive of prolonged, open-source, multi-institutional datasets hosted on the International Electrophysiology Portal (http://www.ieeg.org). In this validation study we evaluated the generalizability and robustness of these algorithms. This study included continuous human recordings from the implanted NeuroVista seizure advisory system (Cook et al., 2013), offering an unprecedented opportunity to prospectively evaluate seizure detection algorithm performance in this use-case. We also developed a pipeline infrastructure allowing rapid application of these algorithms to custom, user-supplied datasets.

Here, we present descriptions of the datasets used in the kaggle.com seizure detection contest, the methodology of each of the winning algorithms, and the performance of these algorithms in both the original competition and a larger validation trial. These algorithms have immediate utility for state-of-the-art seizure detection. We hope that this work, in addition to our published pipeline and archive, will encourage open data sharing to improve research reproducibility and facilitate algorithm comparison.

Materials and methods

Experimental design

The goals of this study were (i) to develop and objectively compare seizure detection algorithms through crowdsourcing; (ii) to assess the robustness of the top algorithms on out-of-sample data; and (iii) to produce a pipeline to facilitate application of these algorithms to new datasets. Crowdsourcing was carried out using a kaggle.com competition (described in detail below). This competition included data from eight human patients with epilepsy and four canines with naturally-occurring epilepsy. The top three algorithms from the competition were then evaluated using data from 18 out-of-sample human patients with epilepsy. No data were excluded from the study, and no subjects or data were removed as outliers.

Subjects and data

Twelve intracranial EEG datasets provided by the Mayo Clinic and University of Pennsylvania were selected for use in the kaggle.com competition. Four of the datasets were generated from chronic recordings of canines with naturally-occurring epilepsy using the ambulatory NeuroVista Seizure Advisory System implanted device described previously (Davis et al., 2011). The dogs were housed at the veterinary hospitals at the University of Minnesota and University of Pennsylvania and continuously monitored with video and iEEG. All dogs had normal neurologic examinations and MRIs. IEEG was acquired from the implanted device with a sampling rate of 400 Hz from 16 subdural electrodes arranged on two standard, human-sized, 4-contact strips implanted on each hemisphere in an antero-posterior position. An analogue anti-aliasing low-pass filter was applied with poles at 100 Hz and 150 Hz. The remaining eight datasets were produced from patients with drug-resistant epilepsy undergoing intracranial EEG monitoring at Mayo Clinic Rochester. These datasets were sampled continuously at 500 Hz or 5000 Hz (with anti-aliasing low-pass filters at 130 Hz and 1 kHz, respectively) with a varying number of subdural electrode grids as determined by individual clinical considerations. A sample recording is shown in Fig. 1. Subject recording information is shown in Table 1 and patient demographics and electrode locations are shown in Supplementary Table 1.

All intracranial EEG records were reviewed and seizures annotated by two board certified epileptologists (G.W. and B.L.). Malfunctioning or grossly non-physiological iEEG channels were removed by visual inspection. To prepare the data for competition use, each dataset was chronologically split into a training set and a testing set (Table 1). Testing and training data were organized into 1-s clips. Training data clips were labelled 'ictal' for the seizure data segments or 'interictal' for non-seizure data segments, and testing clips remained unlabelled. Training data clips were arranged sequentially, while the order of testing data clips was randomized. Both the training and testing datasets were intentionally unbalanced with a larger number of interictal clips to mimic the sparse
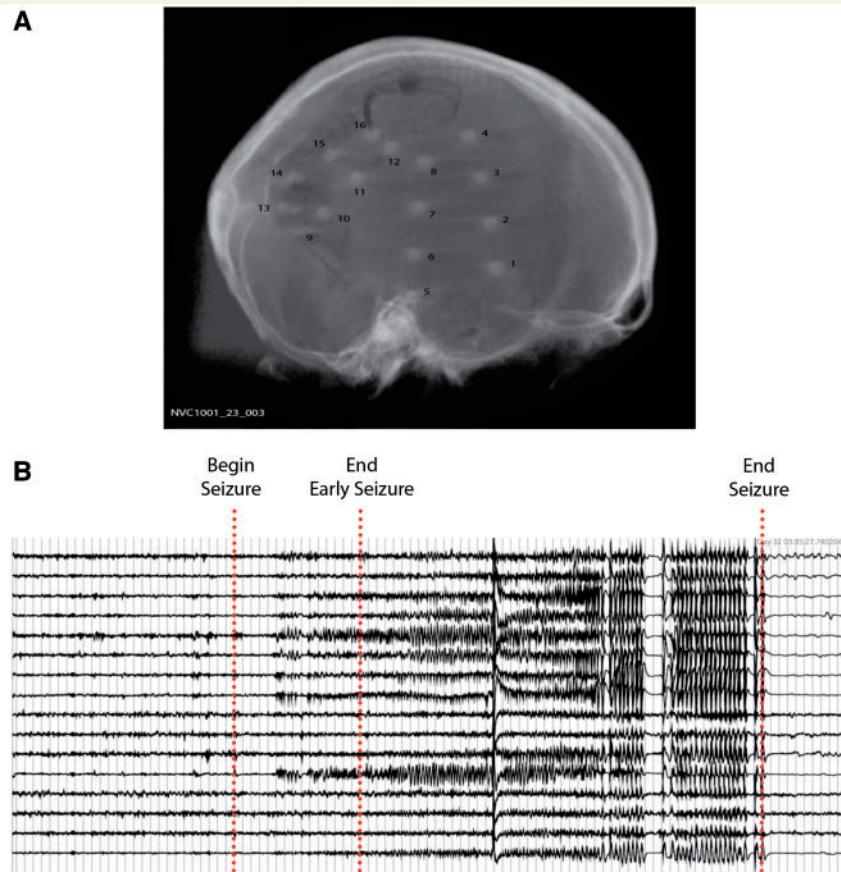
Figure 1 Representative EEG data. (**A**) MRI imaging of patient with implanted NeuroVista SAS device. This device was used for collection of canine data in the competition and human data in the validation study. (**B**) Sample EEG recording of a seizure. Vertical lines represent 1-s intervals. Boundaries of seizure and early seizure periods are marked.

Table 1 Recording characteristics for kaggle.com competition datasets

Subject	Channels, n	Sampling rate	Seizures, n (training)	Recording length used (h)	Training clips, n (% ictal, early)	Testing clips, n (% ictal, early)
Dog 1	16	400	9 (5)	8208	596 (0.30, 0.13)	3181 (0.05, 0.02)
Dog 2	16	400	5 (3)	7152	1320 (0.13, 0.03)	2997 (0.05, 0.01)
Dog 3	16	400	22 (12)	1920	5240 (0.09, 0.03)	4450 (0.09, 0.04)
Dog 4	16	400	6 (2)	1099	3047 (0.08, 0.01)	3013 (0.04, 0.02)
Patient 1	68	500	7 (2)	144	174 (0.40, 0.17)	2050 (0.08, 0.04)
Patient2	16	5000	7 (3)	75	3141 (0.05, 0.01)	3894 (0.06, 0.02)
Patient 3	55	5000	9 (7)	82	1041 (0.31, 0.10)	1281 (0.10, 0.02)
Patient 4	72	5000	5 (2)	96	210 (0.10, 0.10)	543 (0.09, 0.09)
Patient 5	64	5000	7 (3)	141	2745 (0.05, 0.02)	2986 (0.06, 0.02)
Patient 6	30	5000	8 (4)	159	2997 (0.08, 0.02)	2997 (0.07, 0.02)
Patient 7	36	5000	6 (3)	70	3521 (0.08, 0.01)	3601 (0.10, 0.01)
Patient 8	16	5000	4 (2)	71	1890 (0.10, 0.02)	1922 (0.09, 0.02)

nature of seizure events. Only seizures with a 4 h lead of seizure-free activity were included in the study, as previous evidence suggests that seizures less than 4 h apart may not be independent events (Litt *et al.*, 2001). Ictal segments were selected to cover the entirety of seizure from the earliest

electrographic change (EEC) to seizure termination. Interictal segments, each approximately equal to the mean seizure duration of the patient, were selected randomly under the provision that they were not within 1 h before or after a seizure. No further preprocessing or selection criteria were applied to these

clips. Each extracted data segment was individually mean centred. Data segments were stored as ordered structures including sample data, data segment length, iEEG sampling frequency, and channel names in uncompressed MATLAB format data files. Dataset characteristics, including the sequential nature of training segments and selection criteria for inter-ictal segments, and general descriptions of hardware used for recording were made available to competition participants. All data remain available for download at ieeg.org and from kaggle.com (www.kaggle.com/c/seizure-detection/data).

Competition details

Participants were required to use an algorithmic approach to classify data segments. Changes in methodology across subjects were permitted if done in an automated manner that could generalize to outside datasets. Classification of test segments had to be performed by an algorithm and not determined by visual inspection. Each test segment was to be assigned probabilities that it was (i) a seizure clip; and (ii) an early seizure clip. Early seizure clips were defined as occurring within the first 15 s of seizure start as marked by the EEC. These classifications were structured as parallel two-class problems (seizure versus non-seizure, early seizure versus non-early seizure). Participants submitted answers in a .csv file formatted by 'clip name', 'seizure probability', and 'early seizure probability'. Area under the receiver operating characteristic (ROC) curve scores for seizure detection and early seizure detection were then computed by applying varying thresholds to the probability values.

The competition included both public and private leader boards. The public leader board reflected ranked scores on 15% of the full test set and was visible to all participants. This leader board was updated during the algorithm development phase of the competition. Teams were permitted five submissions to the public leader board per day. The private leader board was hidden from view and ranked submissions on their performance on the remaining 85% of the test set. Final placement was awarded based on the private leader board results. Contestants were informed prior to the competition that winning solutions must be made publicly available under an Open Source Initiative (OSI) license to be eligible for recognition and prize money.

Validation study

Many existing seizure detection algorithms suffer from poor generalizability due to overfitting of the dataset used for development. To address this concern, we evaluated the robustness of the top three algorithms from the competition in a larger validation study using recordings from 18 human patients not included in the competition. This study included seven human iEEG recordings from the Hospital of the University of Pennsylvania (HUP) and Mayo Clinic, as well as 11 prolonged, continuous recordings from humans implanted with the NeuroVista seizure prediction device. The HUP and Mayo Clinic recordings were generated during patient monitoring studies in the epilepsy monitoring unit (EMU), mimicking the human datasets used in the competition. The human NeuroVista recordings were generated using the same implantable devices used to produce the canine recordings used in the competition (Davis *et al.*, 2011). These patients were recorded for up to 2.5 years with continuous intracranial EEG via an implanted telemeter device coupled to a belt-worn unit (Cook *et al.*, 2013). All seizure start times were annotated by board certified epileptologists (G.W., B.L., M.C.). Subject demographics and electrode locations are listed in Supplementary Table 1. Each recording was divided into training and testing datasets (Table 2) and segmented into 1-s clips as in the kaggle.com competition. Dataset size was scaled proportionally with the number of seizure clips

Table 2 Recording characteristics for validation study datasets

Subject	Channels, *n*	Sampling rate	Seizures, *n* (training)	Recording length used (h)	Training clips, *n* (% ictal, early)	Testing clips, *n* (% ictal, early)
Patient H1	79	512	5 (3)	310	2306 (0.13, 0.02)	1538 (0.13, 0.02)
Patient H2	61	512	8 (5)	221	408 (0.14, 0.14)	236 (0.11, 0.11)
Patient H3	46	512	5 (3)	146	1853 (0.13, 0.02)	1227 (0.12, 0.02)
Patient M1	56	500	7 (5)	120	5226 (0.06, 0.01)	2617 (0.25, 0.02)
Patient M2	89	500	10 (5)	75	2099 (0.12, 0.04)	2145 (0.14, 0.04)
Patient M3	112	500	5 (3)	163	1051 (0.12, 0.04)	712 (0.13, 0.04)
Patient M4	78	500	8 (5)	215	8721 (0.11, 0.01)	5423 (0.14, 0.01)
Patient NV1	16	400	40 (20)	7959	5876 (0.12, 0.05)	5804 (0.11, 0.04)
Patient NV2	16	400	31 (20)	1503	12466 (0.13, 0.02)	6825 (0.11, 0.02)
Patient NV3	16	400	40 (20)	2034	3752 (0.14, 0.07)	3606 (0.11, 0.07)
Patient NV4	14	400	17 (12)	4995	9339 (0.10, 0.02)	3947 (0.11, 0.02)
Patient NV5	16	400	6 (4)	4955	2004 (0.12, 0.03)	961 (0.08, 0.02)
Patient NV6	16	400	40 (20)	1899	4523 (0.10, 0.06)	4754 (0.15, 0.06)
Patient NV7	16	400	40 (20)	1279	10861 (0.14, 0.03)	10587 (0.14, 0.03)
Patient NV8	15	400	40 (20)	2862	9999 (0.13, 0.03)	9812 (0.12, 0.03)
Patient NV9	16	400	40 (20)	2254	7185 (0.12, 0.04)	7220 (0.13, 0.04)
Patient NV10	16	400	11 (7)	1252	3534 (0.13, 0.03)	1985 (0.11, 0.03)
Patient NV11	16	400	40 (20)	6014	8973 (0.13, 0.03)	8688 (0.10, 0.03)

on a per subject basis. Algorithm metrics were determined using performance on unlabelled test data.

Performance metrics

ROC curves were generated for (i) classification of seizure (ictal) clips versus non-seizure (interictal) clips; and (ii) classification of early seizure clips versus non-early seizure clips. Algorithm performance was assessed using the area under the ROC curve (AUC). The overall performance metric used for algorithm ranking was the average of the AUCs for seizure and early seizure classification (Equation 1). Early seizure clips are emphasized in this metric due to the importance of early seizure detection for successful intervention with a responsive therapeutic device.

$$Performance = \frac{1}{2}(AUC_{seizure} + AUC_{early}) \qquad (1)$$

Statistical analysis

Student's *t*-test was used for comparison of means, and *P*-values < 0.05 were considered significant.

Results

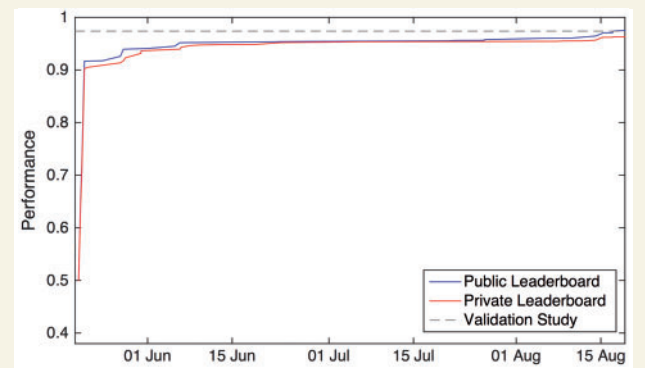### Kaggle competition results

Two hundred teams comprising 241 individuals took part in the competition. These teams submitted a combined 4503 classifications of the test data. The top performance on the public and private leader boards over the length of the competition is shown in Fig. 2. The final standings of the top performing teams are shown in Table 3. The top three teams achieved public leader board scores of 0.975, 0.968, and 0.962, and private leader board scores of 0.963, 0.957, and 0.956. Additional information regarding the distribution of scores across all teams is shown in Supplementary Fig. 2.

Algorithms

The algorithms of the top three performers on the private leader board are summarized below. More detailed descriptions of these models and the code used for implementation are available at www.kaggle.com/c/seizure-detection/details/winners.

Algorithm 1

The first place algorithm was developed by Michael Hills (Dessert Labs). This algorithm relies on three sets of features for classification. The first set of features consists of the pairwise cross-correlation between channel signals as well as the sorted eigenvalues of the cross-correlation matrix. The Fast Fourier Transform is then applied to each 1-s clip for preprocessing. The second set of features consists of the frequency magnitudes of each channel in the

Figure 2 Top algorithm performance over time. Leading score over the course of the kaggle.com competition on public (blue) and private (red) leader boards. The top score in the validation study is represented by the dashed grey line.

Table 3 Final private and public leader board standings for kaggle.com competition

Team name	Private	Public
Michael Hills	0.96288	0.97490
Olson and Mingle	0.95655	0.96803
cdipsters	0.95643	0.96199
alap	0.95600	0.96699
Fusion	0.95437	0.96861
Maineiac	0.95239	0.96017
ACG Mojtaba	0.95183	0.96214
Matthew Roos	0.95072	0.94963
Fitzgerald	0.94956	0.95845
Diba	0.94865	0.95845

range of 1–47 Hz. These power spectra are then normalized within each frequency bin. The third set of features consists of the pairwise cross-correlation between normalized channel power spectra in the range of 1–47 Hz as well as the sorted eigenvalues of the cross-correlation matrix. A random forest classifier of 3000 trees is trained on the complete feature set.

Algorithm 2

The second place algorithm was developed by Eben Olson (Yale University) and Damian Mingle (WPC Healthcare). Data are preprocessed in an automated filter selection step. Combinations of filters (each containing up to four filters) are chosen from a bank of 10 partially overlapping, approximately log-spaced bandpass filters covering the range 5–200 Hz and evaluated on each subject. The three combinations that perform best on cross-validation are retained for each subject. After filtering, covariance matrices are calculated and normalized for each clip to generate the feature set. Classification is carried out using an ensemble of 100 multi-layered neural networks each consisting of

two hidden layers of 200 and 100 units, respectively. These networks are trained with the AdaDelta method for 100 epochs. Each network is trained on a 12-channel subset of the full covariance matrix using a dropout of 0.5 in the hidden layers.

Algorithm 3

The third place algorithm was developed by Ishan Talukdar, Nathan Moore, and Alexander Sood (UC Berkeley). For preprocessing, samples are downsampled to 100 Hz to decrease noise. The algorithm relies on channel-specific features as well as global signal features. The channel-specific features include signal characteristics (maximum amplitude, mean amplitude, absolute deviation, and variance), as well as characteristics of the Fast Fourier Transform of the signal (maximum power, mean power, variance, and frequency at which the maximum power occurs). The global features also include signal features in the time domain (maximum amplitude, mean amplitude, maximum absolute deviation across channels, maximum, mean, and variance of the variance across channels, and covariance between channel signals) and characteristics of the Fast Fourier Transform of the signal (maximum power, mean power, maximum variance across channels, and the maximum, mean, and variance of the frequency at which maximum power occurs across channels). Channel-specific and global features are also extracted from the first and second derivatives of the time series data. Classification is carried out by averaging the outputs of an ensemble of 1000 decision trees using the Extremely Randomized Trees algorithm (Geurts *et al.*, 2006), implemented in python using scikit-learn's ExtraTreesClassifier.

Validation study results

The performance of the top three teams was further evaluated in the validation study (Fig. 3). In addition to standard clinical data from patients recorded at Mayo Clinic and the Hospital of the University of Pennsylvania (HUP), the validation study included prolonged, continuous recordings from humans implanted with the NeuroVista seizure prediction device (see 'Materials and methods' section). These datasets provide the first-ever opportunity to directly evaluate seizure detection algorithms on long-running, human implanted device data. As in the kaggle.com competition, performance metrics were based on clip-by-clip classifications. The teams achieved overall performance scores of 0.972, 0.946, and 0.974, respectively, representing mean changes in performance of +0.009, −0.011, and +0.017 relative to the private leader board scores (Table 3). The performance of the algorithms varied across patient cohorts, with better performance of all algorithms in the NeuroVista cohort than in the Mayo Clinic cohort (for each algorithm, respectively: effect sizes of 0.035, 0.0855, 0.055; P-values of 0.007, 0.005, and 0.004 by one-tailed *t*-test with 13 degrees of freedom). The performance in the HUP cohort was not significantly different from that in
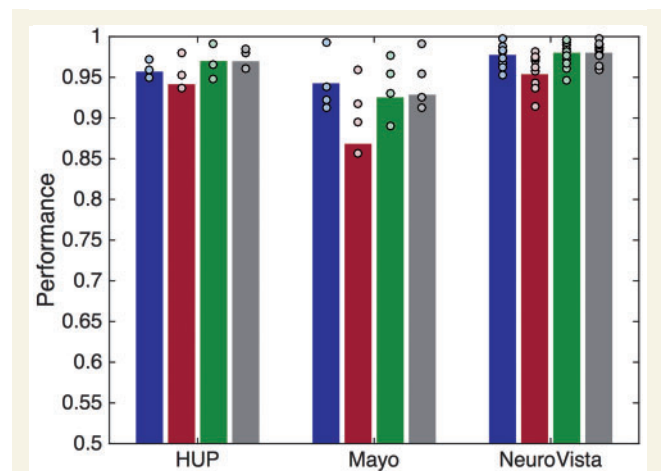

Figure 3 Validation study performance. Performance of (*left to right bars*) Algorithm 1 (blue), Algorithm 2 (red), Algorithm 3 (green), and the ensemble Algorithm (grey) on each cohort in the validation study. Each point represents performance on an individual subject.

either the NeuroVista or Mayo Clinic cohorts. To control for potential skewing of ROC curves due to differences in sample size among subjects, we also generated ROC curves for each algorithm for each individual subject. AUC values from these curves were used to compute performance when each subject is weighted equally [mean ± standard error of the mean (SEM); Algorithm 1: 0.966 ± 0.006; Algorithm 2: 0.946 ± 0.008; Algorithm 3: 0.968 ± 0.006]. Full performance metrics of each algorithm on the validation dataset are shown in Supplementary Table 2.

All algorithms performed better on seizure classification (AUCs of 0.981, 0.976, 0.984) than on early seizure classification (AUCs of 0.964, 0.916, 0.964) as shown in Fig. 4. While the algorithms achieved similar accuracy in seizure epoch classification, Algorithm 2 performed more poorly in early seizure classification than the other algorithms, particularly on the Mayo Clinic datasets (Supplementary Table 2).

To directly assess their utility for functional seizure detection, the algorithms were further evaluated on the NeuroVista cohort using two specific tuning strategies: (i) high specificity; and (ii) high early seizure sensitivity.

We first tuned the algorithms to have few false positives, and measured their detection sensitivity at high specificity. This analysis directly simulates the ideal application of these algorithms for automated generation of seizure diaries or clinical data parsing from high volume, continuous recordings. Seizure and early seizure detection sensitivities computed at a specificity threshold of one false positive per hour of interictal data (specificity of 0.9997) are shown in Supplementary Table 3, and seizure detection at this false positive rate is shown in Fig. 5. This false positive rate would represent a significant improvement relative to currently approved devices, which typically deliver 600 to
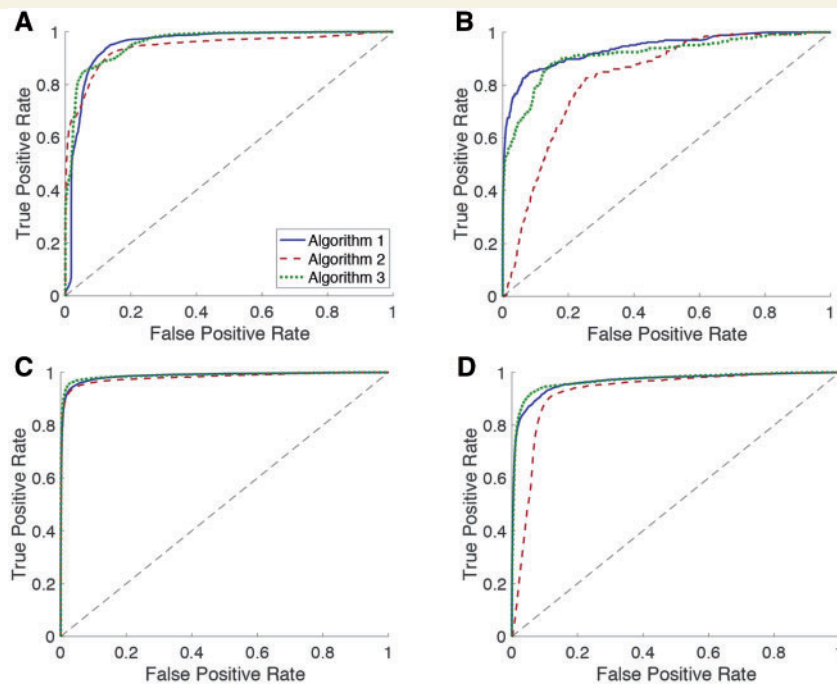
Figure 4 Validation study ROC curves. ROC curves for (**A**) seizure classification in the HUP and Mayo cohorts, (**B**) early seizure classification in the HUP and Mayo cohorts, (**C**) seizure classification in the NeuroVista cohort, (**D**) early seizure classification in the NeuroVista cohort.

2000 stimulations per day (25 to 83 false positives/h) (Sun and Morrell, 2014). These results indicate that these algorithms provide sufficient detection sensitivity with high positive predictive value (PPV) (Algorithms 1 and 3 PPV = 0.99, Algorithm 2 PPV = 0.97) for identification of seizure and early seizure periods with few false positives. Every seizure in the test set ($n = 162$) was detected by Algorithms 1 and 3 with at least one correctly identified seizure clip (seizure detection sensitivity = 100%; Algorithm 2 seizure detection sensitivity = 94%).

We next tuned the algorithms for highly sensitive early detection of seizure activity, which is necessary for effective neuroresponsive therapy. Early seizure detection poses a quite difficult problem as seizure onset patterns are often very diverse and may closely resemble interictal epileptiform bursts that can occur frequently between seizures (Davis *et al.*, 2016). Despite the negative impact of false positive detections on device battery life and the lack of understanding of the effects of frequent stimulation (Hodaie *et al.*, 2002; Chkhenkeli *et al.*, 2004), high false positive rates are often tolerated as the stimulation is below patient perception in order to reliably capture seizure onset. We examined the high sensitivity use case by computing the specificity when setting early seizure detection sensitivity to 75% (Supplementary Table 3). In this limit, Algorithms 1 and 3 achieved specificities of nearly 99%. It is important to note that the reported specificities are based on classification of individual 1-s clips and could be improved by postprocessing, such as integration of classification results
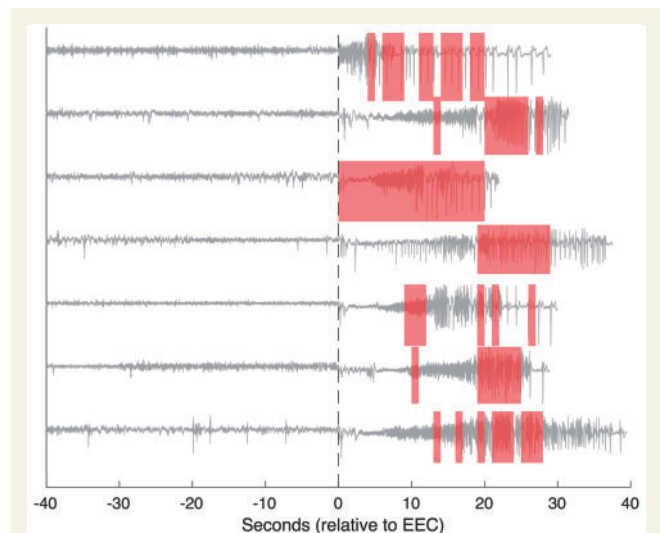
Figure 5 Representative seizure detection in the low-false-positive limit. Each EEG trace shows a single representative channel signal from a different seizure with 40 s of preictal recording. The seizure EEC is denoted by the dashed line. Seizures shown were all derived from Patient NV1. Areas highlighted in red were classified as seizure by Algorithm 1.

over several seconds. In addition, while incorrectly labelling late seizure activity as early seizure decreases algorithm specificity by our metrics it may not have any functional negative impact in patient care.

Algorithm performances varied on a per subject basis. This discrepancy makes it difficult to predict which algorithm will perform best on novel patient data. To address this concern, we evaluated the performance of an ensemble algorithm that averages the prediction scores of the top three algorithms (Fig. 3). Combining multiple, heterogeneous learning algorithms tends to decrease overfitting and provide a more generalizable solution (Bühlmann, 2012). This algorithm performed better than Algorithms 1 and 2 on a patient-by-patient basis, with mean score increases of 0.005 ($P = 0.012$) and 0.03 ($P < 0.001$), respectively (one-tailed paired t-test). The ensemble algorithm provided a mean score increase relative to Algorithm 3 of 0.004 on average, but this difference was not statistically significant. While the kaggle.com competition was designed to select the single best detection algorithm, it is possible that optimal detection over diverse patient cohorts may be achieved through combination or stacking of individual models.

Discussion

This seizure detection competition yielded a set of highly successful classification algorithms. These algorithms provide an immediate improvement in seizure detection over the current industry standard, and represent a new benchmark for personalized seizure detection. The performance of the top three algorithms was consistent between the contest and the validation study, indicating that these methods provide robust seizure detection not limited to patients in the dataset used for development.

The field of seizure detection has been hampered by a lack of reproducible, properly validated algorithms. While many seizure detection algorithms have been published (Tzallas et al., 2012; Orosco et al., 2013), few studies have been carried out to assess the reproducibility of algorithm performance across multiple datasets (Varsavsky et al., 2011; Orosco et al., 2013). Consequently, there is significant concern that many of these algorithms suffer from overfitting of the dataset used for development, and do not offer genuine advancements in detection technology. In this study, we present a framework for direct, objective comparison of algorithms on a common dataset. We have addressed concerns regarding algorithm robustness by demonstrating the efficacy of these algorithms on validation data from patients not included in the original study. We have shown that the algorithms produced in this competition detect seizure activity with high accuracy regardless of subject species (human and canine), recording method (high-density EMU and implanted device recordings), or recording institution (HUP, Mayo Clinic, NeuroVista). Seizure start and end times were annotated by several different clinicians without loss of classification accuracy. The original competition dataset remains open for submission, so that any seizure detection algorithm can be validated on these data and compared to the competitors' performances. We have taken the additional step to supply the source code for the top three algorithms along with a custom pipeline to facilitate application of these algorithms to any EEG dataset. This pipeline can be easily used by researchers to assess the efficacy of these algorithms in their own data for functional detection of seizures or as a benchmark for comparison of novel detection methods (https://github.com/sbaldassano/seizuredetection).

Accurately detecting the beginning of seizure activity is critical for developing effective neuroresponsive devices designed to intervene before seizure propagation. The NeuroVista cohort used in this validation study provides an unprecedented use–case opportunity to directly evaluate algorithm performance in human patients with continuously recording implanted devices. We have demonstrated highly accurate early seizure detection in human implanted device recordings (early seizure AUCs of 0.966, 0.925, and 0.970 for Algorithms 1, 2, and 3, respectively), and shown that it is possible to achieve high specificity while tuned for high early seizure sensitivity. It is also important to consider that these results were achieved despite inherent subjectivity in the marking of seizure EEC (Wilson et al., 2003; Halford et al., 2013), which may introduce dataset variability across seizures and clinician markers. While these algorithms must be further validated on a larger database of human recordings, these results represent a promising step toward improvement in the efficacy of neuroresponsive devices.

Automated seizure detection algorithms may also address the clinical and research burden of manual record review. Manual identification of seizure epochs is cumbersome due to high patient volumes at major academic centres and the long length of recordings from ambulatory or implanted devices. The detection methods presented in this study offer high sensitivity and high specificity seizure detection for all subjects included in the competition and the validation study [validation study AUC scores (mean ± SEM) of 0.983 ± 0.004, 0.982 ± 0.004, 0.983 ± 0.004]. Using the provided data hosting services and analysis pipeline, these algorithms can be immediately applied to clinical and research datasets to decrease the existing labour burden.

Structuring the problem of seizure detection as a kaggle.com competition allowed us to engage with a diverse group of data scientists, many of whom have little to no neuroscience or clinical experience. This competition provided a unique opportunity to leverage the signal processing and machine learning expertise of these scientists to address an important problem in neuroscience research. Crowdsourcing solutions using this platform yielded rapid advancements in state-of-the-art detection technology by providing research bandwidth far exceeding that of a single laboratory. These results were achieved in an extremely short time and at greatly reduced cost compared to prolonged multi-year research efforts. The success of this competition demonstrates how sharing high quality data results in the creation of powerful translational tools for clinical patient care. Open access to data and methods is essential for producing reproducible research and comparing performance of novel algorithms. To facilitate

corroboration of results and encourage further research, all datasets used in the competition and validation study are hosted on http://ieeg.org.

There are several methodological considerations to address related to the competition and this research. One critical issue in development of robust algorithms is prevention of overfitting of the training data. While use of a private leader board in the kaggle.com competition mitigates this concern to a degree, the data used for the public and private leader boards are from the same patients and therefore similar in nature. It is possible that allowing multiple submissions to the public leader board may result in inappropriately high-variance solutions. Further, it is possible that teams may use statistical analyses of the unlabelled testing data during method development in a manner not translatable to prospective algorithm use. We address the issue of potential overfitting by evaluating the algorithms in the validation study, using data from 18 patients not included in the competition. Algorithm performance is consistent between the studies on average, but there are a few validation study subjects in which performance showed a modest decline. It is important to note that algorithm performance may be dependent on cohort-specific variables such as clinical patient scenario and methods of data collection or seizure start time annotation.

Another consideration relates to ease of algorithm implementation in an implantable device. The kaggle.com competition did not place any requirements on algorithm runtime, computational requirements, or code structure. The winning classification algorithms we present may need to be optimized or otherwise altered to be compatible with the relatively limited computational power of a compact implantable device. The validation study was carried out using a Rackform R331.v5 server (Silicon Mechanics) with 32 Intel Xeon E5-2698v3 (2.3 GHz) cores and 256 GB memory. Computation times (training, testing) for each algorithm were benchmarked using subject NV1 (696 ictal and 5180 interictal training clips, 5804 testing clips) (Algorithm 1: 75 s, 28 s; Algorithm 2: 3076 s, 618 s; Algorithm 3: 1092 s, 939 s).

These algorithms were able to accurately model seizure behaviour over many individual subjects. This patient-specific modelling requires verification of several training seizures by a neurologist before detection can be automated. However, preliminary work indicates that only a few training seizures are required for high-accuracy detection (Supplementary Fig. 3), suggesting that algorithms may be implemented with limited neurologist input. While these algorithms may not be suitable for short, inpatient studies during which few seizures are recorded, such as presurgical patient evaluation, they offer utility for extended inpatient monitoring (e.g. long-term EEG monitoring, neurointensive care unit observation) as well as for ambulatory or implanted device recordings. It is also important to consider that seizure detection using standard (scalp) EEG presents unique challenges. Compared to intracranial recordings, standard EEG is complicated by more poorly localized signals (Ray et al., 2007), increased biological noise (Scheer et al., 2006), and attenuation of higher-frequency signal components by tissue (Pfurtscheller and Cooper, 1975). Further work must be carried out to validate the performance of these algorithms on standard EEG to assess their range of clinical applicability.

This study only included seizures with at least a 4-h lead of seizure-free activity. Clustered seizures represent related physiologic events and tend to have highly correlated morphology on EEG. As a result, these seizures are more easily modelled than leading seizures, and are typically removed from performance benchmarking to prevent biasing of results (Litt et al., 2001; Cook et al., 2013). Early seizure behaviour may also be easier to identify in the context of postictal EEG suppression from a previous seizure due to decreased background activity (Esteller et al., 2005). In contrast to closed-loop applications using continuous data, these algorithms are unable to leverage knowledge of recent seizures, a powerful feature for detecting subsequent, clustered seizure events (Dudek and Staley, 2011). This study methodology provides conservative estimates of performance relative to those expected in practice. Further work using extended, continuous recordings must be carried out to directly compare algorithm performance to that of existing closed-loop devices.

The performance metric used for algorithm ranking relies on the area under the ROC curves for seizure and early seizure classification. This metric provides the most objective assessment of algorithm efficacy over the full range of functionality. However, in the application of seizure detection from continuous iEEG it is important to limit the number of false positive detections in order to achieve acceptable positive predictive values. Therefore it could be argued that a more appropriate metric would be sensitivity at high specificity, or the area under a restricted segment of the ROC curve. While we included such a metric in the validation study, it was not used for initial algorithm rankings in the competition and may have resulted in selection of different winning teams.

Finally, for the purpose of the competition, the datasets were composed of 1-s clips randomized from seizure and interictal events, discarding any temporal relationships between clips. Performance of these algorithms can only be improved by incorporating knowledge of sequential clips in a real-world situation.

We present this crowdsourced experiment as a successful example of a new collaborative research paradigm rooted in the principles of data and code sharing and experimental reproducibility. The authors firmly believe that changing the incentive structure to stimulate similar projects across many fields is essential to accelerate progress and eliminate waste and redundancy in research. Our group continues to aggressively work with funding agencies and academic and industry partners to harness the 'power of the crowd' in further translational neuroengineering research.

It is important to note that crowdsourced research may present challenges that must be considered during project

design. Effective crowdsourcing requires incentivization of large numbers of qualified entrants, most of whom will not receive cash prizes. As the primary motivation of these entrants is often the opportunity to develop creative solutions to interesting problems, the competition must address a compelling need to foster participation. In addition, while this competition produced rapid advancements in seizure detection technology, experience gained in traditional, prolonged research efforts was essential to framing the problem in a tractable manner for participants who may not be experts in epilepsy or neuroscience. Such experience is necessary to design fair and objective performance metrics to compare competitors' solutions while maximizing the utility of winning solutions. Crowdsourced research can also raise concerns regarding intellectual property and licensing of solutions. In this competition, we required that winning solutions be made available under an Open Source Initiative approved license to facilitate implementation by other researchers and clinicians.

We have successfully conducted a large-scale, online, seizure detection competition using open access datasets from canines and humans. The structure of the kaggle.com competition allows for direct comparison of algorithms on a common dataset. This competition yielded many novel solutions to the problem of personalized seizure detection. The top three algorithms were successfully validated on unseen datasets including long-running human implanted device recordings. We have provided open source code for each of these algorithms and an application pipeline to facilitate translational use by researchers and clinicians. The rapid progress afforded by crowdsourcing algorithm development provides further evidence for a need for open access data and methods to ensure transparency and reproducibility in research.

Supplementary material

Supplementary material is available at *Brain* online.

References

Abend NS, Dlugos DJ, Hahn CD, Hirsch LJ, Herman ST. Use of EEG monitoring and management of non-convulsive seizures in critically ill patients: a survey of neurologists. Neurocrit Care 2010; 12: 382–9.

Abend NS, Topjian AA, Gutierrez-Colina AM, Donnelly M, Clancy RR, Dlugos DJ. Impact of continuous EEG monitoring on clinical management in critically ill children. Neurocrit Care 2011; 15: 70–5.

Acharya UR, Sree SV, Chattopadhyay S, Yu W, Ang PCA. Application of recurrence quantification analysis for the automated identification of epileptic EEG signals. Int J Neural Syst 2011; 21: 199–211.

Alkan A, Koklukaya E, Subasi A. Automatic seizure detection in EEG using logistic regression and artificial neural network. J Neurosci Methods 2005; 148: 167–76.

Altenburg J, Vermeulen RJ, Strijers RLM, Fetter WPF, Stam CJ. Seizure detection in the neonatal EEG with synchronization likelihood. Clin Neurophysiol 2003; 114: 50–5.

Brinkmann BH, Wagenaar J, Abbot D, Adkins P, Bosshard SC, Chen M, et al. Crowdsourcing reproducible seizure forecasting in human and canine epilepsy. Brain 2016; 139: 1713–22.

Bühlmann P. Bagging, boosting and ensemble methods. In: Handbook of computational statistics. Berlin, Heidelberg: Springer Berlin Heidelberg; 2012. p. 985–1022.

Casson AJ, Yates DC, Patel S, Rodriguez-Villegas E. Algorithm for AEEG data selection leading to wireless and long term epilepsy monitoring. Conf Proc IEEE Eng Med Biol Soc 2007; 2007: 2456–9.

Chabolla DR, Murro AM, Goodman RR, Barkley GL, Worrell GA, Drazkowski JF, et al. Treatment of Mesial temporal lobe epilepsy with responsive hippocampal stimulation by the RNS™ neurostimulator. In: Annual meeting of the American Epilepsy Society; 2006; San Diego, CA. AES 2006 Annual Meeting Abstract Database. AESnet.org.

Chkhenkeli SA, Šramka M, Lortkipanidze GS, Rakviashvili TN, Bregvadze ES, Magalashvili GE, et al. Electrophysiological effects and clinical results of direct brain stimulation for intractable epilepsy. Clin Neurol Neurosurg 2004; 106: 318–29.

Chung JM, Meador K, Eisenschenk S, Ghacibeh GA, Vergara DT, Eliashiv DS, et al. Utility of invasive ictal EEG recordings in presurgical evaluation of patients with medically refractory temporal lobe epilepsy and normal MRI. Int J Epilepsy 2015; 2: 66–71.

Cook MJ, O'Brien TJ, Berkovic SF, Murphy M, Morokoff A, Fabinyi G, et al. Prediction of seizure likelihood with a long-term, implanted seizure advisory system in patients with drug-resistant epilepsy: a first-in-man study. Lancet Neurol 2013; 12: 563–71.

Cui L, Bozorgi A, Lhatoo SD, Zhang G-Q, Sahoo SS. EpiDEA: extracting structured epilepsy and seizure information from patient discharge summaries for cohort identification. AMIA Annu Symp Proc 2012; 2012: 1191–200.

D'Alessandro M, Vachtsevanos G, Esteller R, Echauz J, Cranstoun S, Worrell G, et al. A multi-feature and multi-channel univariate selection process for seizure prediction. Clin Neurophysiol 2005; 116: 506–16.

Davis KA, Sturges BK, Vite CH, Ruedebusch V, Worrell G, Gardner AB, et al. A novel implanted device to wirelessly record and analyze continuous intracranial canine EEG. Epilepsy Res 2011; 96: 116–22.

Davis KA, Ung H, Wulsin D, Wagenaar J, Fox E, Patterson N, et al. Mining continuous intracranial EEG in focal canine epilepsy: relating interictal bursts to seizure onsets. Epilepsia 2016; 57: 89–98.

Dudek FE, Staley KJ. Seizure probability in animal models of acquired epilepsy: a perspective on the concept of the preictal state. Epilepsy Res 2011; 97: 324–31.

Esteller R, Echauz J, D'Alessandro M, Worrell G, Cranstoun S, Vachtsevanos G, et al. Continuous energy variation during the seizure cycle: towards an on-line accumulated energy. Clin Neurophysiol 2005; 116: 517–26.

Esteller R, Echauz J, Tcheng T, Litt B, Pless B. Line length: an efficient feature for seizure onset detection. In: 2001 Conference Proceedings of the 23rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE; 2001. p. 1707–10.

Gardner AB, Krieger AM, Vachtsevanos G, Litt B. One-class novelty detection for seizure analysis from intracranial EEG. J Mach Learn Res 2006; 7: 1025–44.

Geurts P, Ernst D, Wehenkel L. Extremely randomized trees. Mach Learn 2006; 63: 3–42.

Gotman J. Automatic recognition of epileptic seizures in the EEG. Electroencephalogr Clin Neurophysiol 1982; 54: 530–40.

Gutierrez-Colina AM, Topjian AA, Dlugos DJ, Abend NS. Electroencephalogram monitoring in critically ill children: indications and strategies. Pediatr Neurol 2012; 46: 158–61.

Halford JJ, Schalkoff RJ, Zhou J, Benbadis SR, Tatum WO, Turner RP, et al. Standardized database development for EEG epileptiform transient detection: EEGnet scoring system and machine learning analysis. J Neurosci Methods 2013; 212: 308–16.

Hodaie M, Wennberg RA, Dostrovsky JO, Lozano AM. Chronic anterior thalamus stimulation for intractable epilepsy. Epilepsia 2002; 43: 603–8.

Kharbouch A, Shoeb A, Guttag J, Cash SS. An algorithm for seizure onset detection using intracranial EEG. Epilepsy Behav 2011; 22 (Suppl 1): S29–35.

Kwan P, Brodie MJ. Early identification of refractory epilepsy. N Engl J Med 2000; 342: 314–9.

Litt B, Esteller R, Echauz J, D'Alessandro M, Shor R, Henry T, et al. Epileptic seizures may begin hours in advance of clinical onset: a report of five patients. Neuron 2001; 30: 51–64.

Liu HS, Zhang T, Yang FS. A multistage, multimethod approach for automatic detection and classification of epileptiform EEG. IEEE Trans Biomed Eng 2002; 49: 1557–66.

Morrell MJ. Antiepileptic medications for the treatment of epilepsy. Semin Neurol 2002; 22: 247–58.

Morrell MJ. Responsive cortical stimulation for the treatment of medically intractable partial epilepsy. Neurology 2011; 77: 1295–304.

Orosco L, Correa AG, Laciar E. Review: a survey of performance and techniques for automatic epilepsy detection. J Med Biol Eng 2013; 33: 526.

Osorio I, Frei MG, Giftakis J, Peters T, Ingram J, Turnbull M, et al. Performance reassessment of a real-time seizure-detection algorithm on long ECoG series. Epilepsia 2002; 43: 1522–35.

Pfurtscheller G, Cooper R. Frequency dependence of the transmission of the EEG from cortex to scalp. Electroencephalogr Clin Neurophysiol 1975; 38: 93–6.

Pradhan N, Sadasivan PK, Arunodaya GR. Detection of seizure activity in EEG by an artificial neural network: a preliminary study. Comput Biomed Res 1996; 29: 303–13.

Raghunathan S, Gupta SK, Ward MP, Worth RM, Roy K, Irazoqui PP. The design and hardware implementation of a low-power real-time seizure detection algorithm. J Neural Eng 2009; 6: 56005.

Ramgopal S, Thome-Souza S, Jackson M, Kadish NE, Sánchez Fernández I, Klehm J, et al. Seizure detection, seizure prediction, and closed-loop warning systems in epilepsy. Epilepsy Behav 2014; 37: 291–307.

Ray A, Tao JX, Hawes-Ebersole SM, Ebersole JS. Localizing value of scalp EEG spikes: a simultaneous scalp and intracranial study. Clin Neurophysiol 2007; 118: 69–79.

Scheer HJ, Sander T, Trahms L. The influence of amplifier, interface and biological noise on signal quality in high-resolution EEG recordings. Physiol Meas 2006; 27: 109–17.

So EL. Integration of EEG, MRI, and SPECT in localizing the seizure focus for epilepsy surgery. Epilepsia 2000; 41: S48–54.

Sun FT, Morrell MJ. Closed-loop neurostimulation: the clinical experience. Neurotherapeutics 2014; 11: 553–63.

Sun FT, Morrell MJ, Wharen RE. Responsive cortical stimulation for the treatment of epilepsy. Neurotherapeutics 2008; 5: 68–74.

Temko A, Thomas E, Marnane W, Lightbody G, Boylan G. EEG-based neonatal seizure detection with support vector machines. Clin Neurophysiol 2011; 122: 464–73.

Tzallas AT, Tsalikakis DG, Karvounis EC, Astrakas L, Tzaphlidou M, Tsipouras MG, et al. Automated epileptic seizure detection methods: a review study. In: Stevanovic D, editor. Epilepsy - Historical, Electroencephalographic, and Psychological Aspects. Rijeka, Croatia: INTECH Open Access Publisher; 2012.

Varsavsky A, Mareels I, Cook M. Epileptic seizures and the EEG: measurement, models, detection and prediction. Boca Raton, FL USA: Taylor & Francis; 2011.

Wagenaar JB, Brinkmann BH, Ives Z, Worrell GA, Litt B. A multimodal platform for cloud-based collaborative research. In: 2013 6th International IEEE/EMBS Conference on Neural Engineering (NER). IEEE; 2013. p. 1386–9.

Wagenaar JB, Worrell GA, Ives Z, Matthias D, Litt B, Schulze-Bonhage A. Collaborating and sharing data in epilepsy research. J Clin Neurophysiol 2015; 32: 235–9.

White AM, Williams PA, Ferraro DJ, Clark S, Kadam SD, Dudek FE, et al. Efficient unsupervised algorithms for the detection of seizures in continuous EEG recordings from rats after brain injury. J Neurosci Methods 2006; 152: 255–66.

Wiebe S, Blume WT, Girvin JP, Eliasziw M. A randomized, controlled trial of surgery for temporal-lobe epilepsy. N Engl J Med 2001; 345: 311–18.

Wilson SB, Scheuer ML, Emerson RG, Gabor AJ. Seizure detection: evaluation of the Reveal algorithm. Clin Neurophysiol 2004; 115: 2280–91.

Wilson SB, Scheuer ML, Plummer C, Young B, Pacia S. Seizure detection: correlation of human experts. Clin Neurophysiol 2003; 114: 2156–64.