Neural Networks 105 (2018) 104–111

Contents lists available at ScienceDirect

Neural Networks

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

Convolutional neural networks for seizure prediction using intracranial and scalp electroencephalogram

Nhan Duy Truong [a,b,c], Anh Duy Nguyen [b], Levin Kuhlmann [d,e,f], Mohammad Reza Bonyadi [g,h], Jiawei Yang [i], Samuel Ippolito [c], Omid Kavehei [a,b,*]

[a] School of Electrical and Information Engineering, University of Sydney, Sydney, NSW 2006, Australia
[b] Nano-Neuro-inspired Research Laboratory, School of Electrical and Information Engineering, University of Sydney, Sydney, NSW 2006, Australia
[c] School of Engineering, Royal Melbourne Institute of Technology, Melbourne, VIC 3000, Australia
[d] Centre for Human Psychopharmacology, Swinburne University, Hawthorn, VIC 3122, Australia
[e] Neuroengineering Laboratory, Department of Electrical and Electronic Engineering, University of Melbourne, Parkville, VIC 3010, Australia
[f] Department of Medicine, St. Vincent's Hospital Melbourne, University of Melbourne, Parkville, VIC 3010, Australia
[g] Centre for Advanced Imaging, University of Queensland, St. Lucia, QLD 4072, Australia
[h] Optimization and Logistics Group, University of Adelaide, Adelaide, SA 5005, Australia
[i] Nanochap Electronics and Wenzhou Medical University, 268 Xueyuan West Rd., Wenzhou, China

ARTICLE INFO

Article history:
Received 23 January 2018
Received in revised form 24 April 2018
Accepted 26 April 2018
Available online 7 May 2018

Keywords:
Seizure prediction
Convolutional neural network
Machine learning
Intracranial EEG
Scalp EEG

ABSTRACT

Seizure prediction has attracted growing attention as one of the most challenging predictive data analysis efforts to improve the life of patients with drug-resistant epilepsy and tonic seizures. Many outstanding studies have reported great results in providing sensible indirect (warning systems) or direct (interactive neural stimulation) control over refractory seizures, some of which achieved high performance. However, to achieve high sensitivity and a low false prediction rate, many of these studies relied on handcraft feature extraction and/or tailored feature extraction, which is performed for each patient independently. This approach, however, is not generalizable, and requires significant modifications for each new patient within a new dataset. In this article, we apply convolutional neural networks to different intracranial and scalp electroencephalogram (EEG) datasets and propose a generalized retrospective and patient-specific seizure prediction method. We use the short-time Fourier transform on 30-s EEG windows to extract information in both the frequency domain and the time domain. The algorithm automatically generates optimized features for each patient to best classify preictal and interictal segments. The method can be applied to any other patient from any dataset without the need for manual feature extraction. The proposed approach achieves sensitivity of 81.4%, 81.2%, and 75% and a false prediction rate of 0.06/h, 0.16/h, and 0.21/h on the Freiburg Hospital intracranial EEG dataset, the Boston Children's Hospital-MIT scalp EEG dataset, and the American Epilepsy Society Seizure Prediction Challenge dataset, respectively. Our prediction method is also statistically better than an unspecific random predictor for most of the patients in all three datasets.

© 2018 Elsevier Ltd. All rights reserved.

1. Introduction

Advances in data mining and machine learning in the past few decades have attracted significantly more attention to the application of these techniques in detective and predictive data analytics, especially in health care, medical practices, and biomedical engineering (Bou Assi, Nguyen, Rihana, & Sawan, 2017; Freestone, Karoly, & Cook, 2017; Freestone et al., 2015; Kuhlmann, Grayden, & Cook, 2017; Kuhlmann, Grayden, Wendling, & Schiff, 2015; Sinha et al., 2017; Xiao, Wang, Iasemidis, Wong, & Chaovalitwongse, 2017). While the body of available proven knowledge lacks a convincing and comprehensive understanding of the sources of epileptic seizures, some early studies showed the possibility of predicting seemingly unpredictable seizures (Rogowski, Gath, & Bental, 1981; Salant, Gath, & Henriksen, 1998). Along with continuous improvements in recording electroencephalogram (EEG) signals, there have been an increasing number of EEG-based techniques for seizure prediction (Szostak, Grand, & Constandinou, 2017). There have been some articles on seizure prediction using the Freiburg

* Corresponding author at: School of Electrical and Information Engineering, University of Sydney, Sydney, NSW 2006, Australia.
E-mail addresses: duy.truong@sydney.edu.au (N.D. Truong), ngduyanhece@gmail.com (A.D. Nguyen), lkuhlmann@swin.edu.au (L. Kuhlmann), reza@cai.uq.edu.au (M.R. Bonyadi), jiaweiy@nanochap.com (J. Yang), samuel.ippolito@rmit.edu.au (S. Ippolito), omid.kavehei@sydney.edu.au (O. Kavehei).

https://doi.org/10.1016/j.neunet.2018.04.018
0893-6080/© 2018 Elsevier Ltd. All rights reserved.

Hospital dataset (University of Freiburg, 2003). For example, the dynamical similarity index, effective correlation dimension, and increments of accumulated energy were used as features (Maiwald, Winterhalder, Aschenbrenner-Scheibe, Voss, Schulze-Bonhage, & Timmer, 2004). The dynamical similarity index yielded the highest performance, with sensitivity of 42% and false prediction rate (FPR) less than 0.15/h. The mean phase coherence and lag synchronization index of 32-s sliding EEG windows were used as features for seizure prediction (Winterhalder et al., 2006). This approach achieved sensitivity of 60% and FPR of 0.15/h. The approach was further improved by combined use of bivariate empirical mode decomposition and Hilbert-based mean phase coherence as additional features (Zheng, Wang, Li, Bao, & Wang, 2014). As a result, sensitivity was increased beyond 70%, while FPR dropped below 0.15/h. A lightweight approach based on spike rate achieved 75.8% sensitivity and FPR of 0.09/h (Li, Zhou, Yuan, & Liu, 2013). By use of the synchronization information, a method based on phase-match error of two consecutive epochs and variation within each epoch resulted in 95.4% sensitivity and FPR of 0.36/h (Parvez & Paul, 2017). Another synchrony-based approach used the mean phase coherence between each pair of channels calculated over multiple window lengths as an indicator of incoming seizure onset (Kuhlmann et al., 2010).

Frequency bands of the power spectrum of each channel were used as a feature for seizure prediction (Park, Luo, Parhi, & Netoff, 2011). These features were then fed to a support vector machine (SVM) classifier to learn the differences between preictal and interictal instances. This method was tested with the Freiburg Hospital dataset, and achieved sensitivity of 98.3% and FPR of 0.29/h. A similar approach with additional features which are spectral power ratios between different frequency bands achieved sensitivity exceeding 98% and FPR less than 0.05/h (Zhang & Parhi, 2016). However, this approach relied on tailoring features for each patient independently, hence offering reduced generalization as a result. Differently from the two approaches described, Aarabi and He (2014) applied a Bayesian inversion of power spectral density and then applied a rule-based decision to perform the seizure prediction task. This approach was tested with the Freiburg Hospital dataset, with sensitivity of 87.07% and FPR of 0.2/h. Aarabi and He (2017) recently extracted six univariate and bivariate features, including correlation dimension, correlation entropy, noise level, Lempel-Ziv complexity, largest Lyapunov exponent, and nonlinear interdependence, and achieved a comparable sensitivity of 86.7% and lower FPR of 0.126/h. On the basis of the assumption that future events depend on a number of previous events, a multiresolution N-gram on amplitude patterns was used as features (Eftekhar, Juffali, El-Imad, Constandinou, & Toumazou, 2014). After optimization of the feature set per patient, this method yielded a high sensitivity of 90.95% and a low FPR of 0.06/h on the Freiburg Hospital dataset. Recently, the dynamics of EEG was captured by use of 64 fuzzy rules to estimate the trajectory of each sliding EEG window on a Poincaré plane (Sharif & Jafari, 2017). Principal component analysis was used to reduce interrelated features before classification by an SVM. This work achieved sensitivity of more than 91% and FPR below 0.08/h on the Freiburg Hospital dataset.

Patient-specific feature engineering techniques have been successful in seizure prediction tasks by achieving perfect sensitivity (100%) and a very low false alarm rate: 0.05/h (Zhang & Parhi, 2016) or 0/h (Mirowski, LeCun, Madhavan, & Kuzniecky, 2008). Such techniques, however, use numerous preengineered features, selected manually, for each patient, and require lots of resources (e.g., subject domain experts) and time. For example, Mirowski et al. (2008) used six different feature extraction methods and three machine learning algorithms. Zhang and Parhi (2016) used 44 features and a set of 91 cost-sensitive linear SVM classifiers to search for the optimal single features or feature combinations that perform best for each patient. For both of these approaches, not only is the best combination of features and classifiers not known for each patient, but an optimal feature set and classifier may be suboptimal in the future because of the dynamic changes in the brain.

Because of the drawbacks of feature engineering techniques, a generalized approach for seizure prediction is highly beneficial. In this work, we use a convolutional neural network (CNN) for seizure prediction. The main contributions of this work are as follows: (1) we propose an efficient method to preprocess raw EEG data into a form suitable for a CNN; (2) we propose a guideline to help the CNN perform well with the seizure prediction task with minimum feature engineering; and (3) we provide an algorithm that works well across multiple datasets; namely, the Freiburg Hospital dataset (University of Freiburg, 2003), the Boston Children's Hospital (CHB)-MIT dataset (Shoeb, 2009), and the American Epilepsy Society Seizure Prediction Challenge (Kaggle) dataset (Kaggle, 2014). The third main contribution will also reveal factors that describe (unrealistically) high performance of other seizure prediction methods. This confounder is mitigated here by the consideration of numerous datasets.

2. Proposed method

2.1. Datasets

Three datasets were used in this work: the Freiburg Hospital dataset (University of Freiburg, 2003), the CHB-MIT dataset (Shoeb, 2009), and the American Epilepsy Society Seizure Prediction Challenge (Kaggle) dataset (Brinkmann et al., 2016). These three datasets are summarized in Table 1. The Freiburg Hospital dataset consists of intracranial EEG (iEEG) recordings of 21 patients with intractable epilepsy. Because of lack of availability of the dataset, we are able to use data from only 13 patients. A sampling rate of 256 Hz was used to record iEEG signals. In this dataset, there are six recording channels from six selected contacts, where three of them are from epileptogenic regions and the other three are from the remote regions. For each patient, there is at least 50 min of preictal data and 24 h of interictal data. More details about the Freiburg Hospital dataset can be found in Maiwald et al. (2004).

The CHB-MIT dataset contains scalp EEG (sEEG) data from 23 pediatric patients with 844 h of continuous sEEG recording and 163 seizures. The sEEG signals were captured with use of 22 electrodes at a sampling rate of 256 Hz (Shoeb, 2009). We define interictal periods as being between at least 4 h before seizure onset and 4 h after seizure end. In this dataset, there are cases where multiple seizures occur close to each other. For the seizure prediction task, we are interested in predicting the leading seizures. Therefore for seizures that are less than 30 min from the previous seizure, we consider them as only one seizure and use the onset of the leading seizure as the onset of the combined seizure. Besides, we consider only patients with fewer than 10 seizures per day for the prediction task because it is not very critical to perform the task for patients having a seizure every 2 h on average. With these definitions and considerations, there are 13 patients with sufficient data (at least three leading seizures and 3 h of interictal recording).

The American Epilepsy Society Seizure Prediction Challenge dataset has iEEG data from five dogs and two patients with 48 seizures and 627.7 h of interictal recording (Brinkmann et al., 2016). Intracranial EEG (iEEG) canine data were recorded from 16 implanted electrodes with a sampling rate of 400 Hz. Recorded iEEG data from the two patients were from 15 depth electrodes (patient 1) and 24 subdural electrodes (patient 2) at a sampling rate of 5 kHz. Preictal and interictal 10-min segments were extracted by the organizers. Specifically, for each lead seizure, six preictal segments were extracted from 66 min to 5 min before seizure onset with 10 s apart. Interictal segments were randomly selected at least 1 week from any seizure.

Table 1
Summary of the three datasets used in this work.

Dataset	EEG type	No. of patients	No. of channels	No. of seizures	Interictal hours
Freiburg Hospital	Intracranial	13 patients	6	59	311.4
Boston Children's Hospital–MIT	Scalp	13 patients	22	64	209
American Epilepsy Society Seizure Prediction Challenge (Kaggle)	Intracranial	5 dogs, 2 patients	16	48	627.7

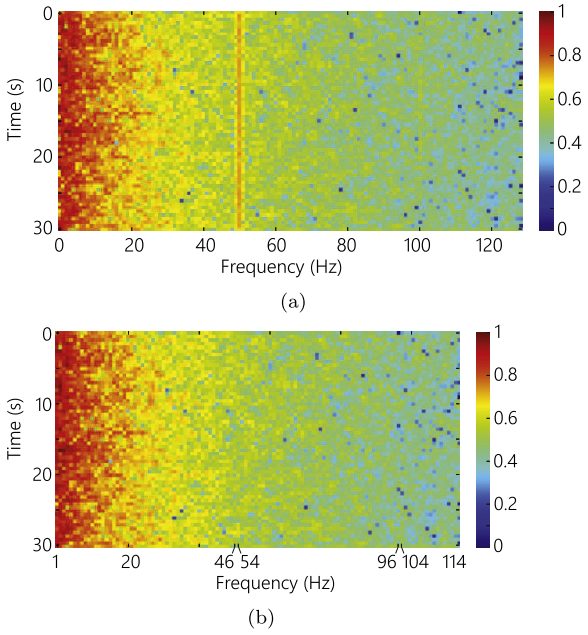

Fig. 1. (a) Example short-time Fourier transform of a 30-s window. (b) Same window after removal of line noise.

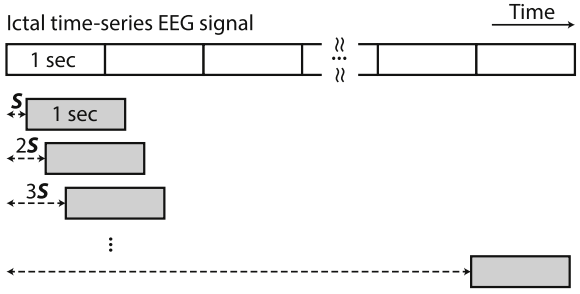

Fig. 2. Generate extra preictal segments to balance the training dataset by sliding a 30-s window along the time axis at every step $\boldsymbol{S}$ over preictal signals. $\boldsymbol{S}$ is chosen per subject so that there are a similar number of samples per class (preictal or interictal) in the training set.

2.2. Preprocessing

Since a two-dimensional CNN is used in this work, it is necessary to convert raw EEG data into a matrix (i.e., image-like format). The conversion must be able to keep the most important information from the EEG signals. Wavelet and Fourier transforms are commonly used to convert time-series EEG signals into image shape (Brinkmann et al., 2016; Khan, Marcuse, Fields, Swann, & Yener, 2017). They are also used as an effective feature extraction method for seizure detection and prediction. In this work, we use the short-time Fourier transform (STFT) to translate raw EEG signals into a two-dimensional matrix composed of frequency and time axes. We use an EEG window length of 30 s. Most of the EEG recordings were contaminated by power line noise at 50 Hz (see Fig. 1a) for the Freiburg Hospital dataset and 60 Hz for the CHB-MIT dataset. In the frequency domain, it is convenient to effectively remove the power line noise by excluding components in the frequency ranges of 47–53 Hz and 97–103 Hz for a power line frequency of 50 Hz and components in the frequency ranges of 57–63 Hz and 117–123 Hz for a power line frequency of 60 Hz. The DC component (at 0 Hz) was also removed. Fig. 1b shows the STFT of a 30-s window after removal of power line noise.

One challenge in many classification tasks is the imbalance of the dataset; that is, more instances in one class than in others (Branco, Torgo, & Ribeiro, 2016). Seizure prediction also encounters this issue; for example, in the Freiburg Hospital dataset, the interictal-to-preictal ratio per patient ranges from 9.5:1 to 15.9:1. To overcome this, we generate more preictal segments by using an overlapped sampling technique during the training phase. In particular, we create extra preictal samples for training by sliding a 30-s window along the time axis at every step $\boldsymbol{S}$ over preictal time-series EEG signals (see Fig. 2). $\boldsymbol{S}$ is chosen per subject so that we have a similar number of samples per class (preictal or interictal) in the training set.

2.3. Convolutional neural network

CNNs have been used extensively for computer vision and natural language processing (Krizhevsky, Sutskever, & Hinton, 2012; Sainath, Mohamed, Kingsbury, & Ramabhadran, 2013). In this work, we use a CNN with three convolution blocks as described in Fig. 3. Each convolution block consists of a batch normalization, a convolution layer with a rectified linear unit activation function, and a max pooling layer. The batch normalization ensures the inputs to the convolution layer have zero mean and unit variance. The first convolution layer has 16 $n \times 5 \times 5$ kernels, where n is the number of EEG channels, with stride of $1 \times 2 \times 2$. The next two convolution blocks have 32 and 64 convolution kernels, respectively, and both have a kernel size of 3×3, stride of 1×1, and max pooling over a 2×2 region. Following the three convolution blocks are two fully connected layers with sigmoid activation and output sizes of 256 and 2, respectively. The former fully connected layer uses a sigmoid activation function, while the latter uses a softmax activation function. Both of the fully connected layers have a dropout rate of 0.5. Our model is implemented in Python 2.7 with use of Keras 2.0 with a Tensorflow 1.4.0 backend. The model was configured to run in parallel on four NVIDIA K80 graphics cards.

Because of the limited available datasets, it is important to prevent the CNN from overfitting the data. First, we keep the CNN architecture simple and shallow as described above (Ba & Caruana, 2014). Second, we propose an approach to prevent overfitting during training of the neural network. A common practice is to randomly split 20% of the training set for use as a validation set. After each training epoch, a loss and/or accuracy is calculated with respect to the validation set to check if the network starts to overfit the training set. This approach works well with datasets where time information is not involved (e.g., images for the classification task). For seizure prediction, we need to use samples from a period different from that of those during training to monitor if the model starts to overfit the data. In this work, we select 25% of later samples from preictal and interictal recordings in the training set for monitoring and the rest for training (see Fig. 4).

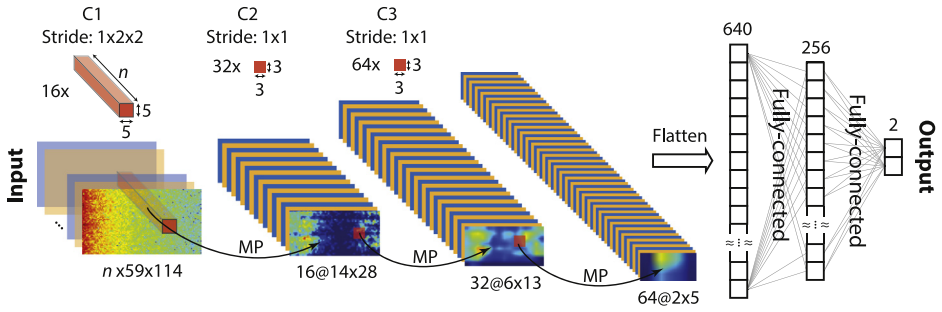

Fig. 3. Convolutional neural network architecture. This illustration is applied to the Freiburg Hospital and Boston Children's Hospital-MIT datasets. For the American Epilepsy Society Seizure Prediction Challenge dataset, the feature sizes are different because of the different recording sampling rate. Short-time Fourier transforms of 30-s windows of raw EEG signals are input. There are three convolution blocks, named *C1*, *C2*, and *C3*. Each block consists of a batch normalization, a convolution layer with a rectified linear unit (ReLU) activation function, and a max pooling layer. For simplicity, max pooling layers are not shown and are noted as *MP*. For C1, there are 16 $n \times 5 \times 5$ kernels, where n is the number of EEG channels, with stride of $1 \times 2 \times 2$. ReLU activation is applied on convolution results before they are subsampled by a max pooling over a $1 \times 2 \times 2$ region. The same steps are applied in C2 and C3 except the convolution kernel size is 3×3, stride is 1×1, and max pooling size is 2×2. Blocks C2 and C3 have 32 and 64 convolution kernels, respectively. Features extracted by the three convolution blocks are flattened and connected to two fully connected layers with output sizes of 256 and 2, respectively. The former fully connected layer uses a sigmoid activation function, while the latter uses a soft-max activation function. Both of the fully connected layers have a dropout rate of 0.5.

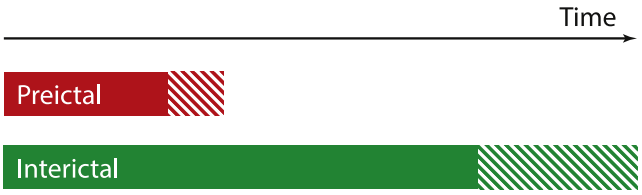

Fig. 4. Practice to prevent the convolutional neural network from overfitting the data during training. Twenty-five percent of later samples (diagonal lines) from preictal and interictal recordings in the training dataset are used for monitoring and the rest are used for training.

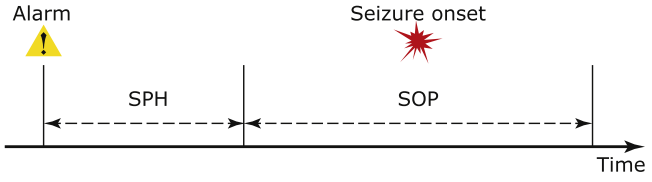

Fig. 5. Definition of the seizure occurrence period (SOP) and the seizure prediction horizon (SPH). For a correct prediction, a seizure onset must be after the SPH and within the SOP.

2.4. Postprocessing

It is common to have isolated false positives during interictal periods. These isolated false predictions can be effectively reduced by use of a discrete-time Kalman filter (Park et al., 2011). In this work, we propose a simple method, called k-of-n, in which an alarm is set only if at least k predictions among the last n predictions were positive. Our experiments showed that $k = 8$ and $n = 10$ are good choices for the purpose of efficient prediction. This means that if during the last 300 s at least 240 s led to a positive prediction, then the alarm is set.

2.5. System evaluation

The seizure prediction horizon (SPH) and seizure occurrence period (SOP) need to be defined before performance metrics such as sensitivity and FPR are estimated. In this work, we follow the definitions of the SOP and SPH proposed by Maiwald et al. (2004) (see Fig. 5). The SOP is the interval where the seizure is expected to occur. The period between the alarm and the beginning of the SOP is the SPH. For a correct prediction, a seizure onset must be after the SPH and within the SOP. Likewise, a false alarm occurs when the prediction system returns a positive result but no seizure occurs during the SOP. When an alarm occurs, it will last until the end of the SOP. Sensitivity is defined as the percentage of seizures correctly predicted divided by the total number of seizures. The FPR is defined as the number of false alarms per hour.

Regarding clinical use, the SPH must be long enough to allow sufficient intervention or precautions (SPH is also called *intervention time*; Bou Assi et al., 2017). In contrast, the SOP should be not too long to reduce the patient's anxiety. Inconsistency in defining the SPH and SOP make the benchmarking among methods difficult and confusing. Park et al. (2011) reported using an SPH of 30 min, but from their explanation what they were implicitly using was an SPH of 0 min and an SOP of 30 min (i.e., if an alarm occurs at any point within 30 min before seizure onset, it is considered a successful prediction). Similarly, Zhang and Parhi (2016) provided a different definition of the SPH: the interval between the alarm and seizure onset.

The metrics used to test the proposed approach are sensitivity and FPR with an SPH of 5 min and an SOP of 30 min. To have a robust evaluation, we follow a leave-one-out cross-validation approach for each subject. If a subject has N seizures, $(N - 1)$ seizures will be used for training, and the remaining seizure will be used for validation. This round is done N times, so all seizures will be used for validation exactly once. Interictal segments are randomly split into N parts. $(N - 1)$ parts are used for training and the remaining part is used for validation. The $(N - 1)$ parts are further split into monitoring and training sets to prevent overfitting as depicted in Fig. 4.

We also compare the prediction performance of our approach with that of an unspecific random predictor. Given an FPR, the probability to raise an alarm in an SOP can be approximated by Schelter et al. (2006)

$$P \approx 1 - e^{-\mathrm{FPR}\cdot\mathrm{SOP}}. \tag{1}$$

Therefore the probability of predicting at least m of M independent seizures by chance is given by

$$p = \sum_{i \geq m} \binom{M}{i} P^i (1 - P)^{M-i}. \tag{2}$$

We calculated p for each patient by using the FPR of that patient and the number of seizures (m) predicted by our method. If p is less than 0.05, we can conclude that our prediction method is significantly better than a random predictor at a significance level of 0.05.

Table 2
Seizure prediction results obtained with the Freiburg Hospital interictal EEG dataset. The model was executed twice, and average results with standard deviations were reported. The seizure occurrence period (SOP) was 30 min and the seizure prediction horizon (SPH) was 5 min. The p value was calculated for the worst case for each patient; that is, with minimum sensitivity and maximum false prediction rate (FPR). Our seizure prediction approach achieves significantly better performance than an unspecific random predictor for all patients except Pat14, where the convolutional neural network results are only marginally better than the random predictor's.

Patient	No. of seizures	Interictal hours	Sensitivity (%)	FPR (/h)	p
Pat1	4	23.9	100 ± 0.0	0.00 ± 0.00	<0.001
Pat3	5	23.9	100 ± 0.0	0.00 ± 0.00	<0.001
Pat4	5	23.9	100 ± 0.0	0.00 ± 0.00	<0.001
Pat5	5	23.9	40 ± 0.0	0.13 ± 0.00	0.032
Pat6	3	23.8	100 ± 0.0	0.00 ± 0.00	<0.001
Pat14	4	22.6	50 ± 0.0	0.27 ± 0.00	0.078
Pat15	4	23.7	100 ± 0.0	0.02 ± 0.02	<0.001
Pat16	5	23.9	80 ± 0.0	0.17 ± 0.13	0.001
Pat17	5	24	80 ± 0.0	0.00 ± 0.00	<0.001
Pat18	5	24.8	100 ± 0.0	0.00 ± 0.00	<0.001
Pat19	4	24.3	50 ± 0.0	0.16 ± 0.00	0.033
Pat20	5	24.8	60 ± 0.0	0.04 ± 0.00	<0.001
Pat21	5	23.9	100 ± 0.0	0.00 ± 0.00	<0.001
Total	59	311.4	81.4 ± 0.0	0.06 ± 0.00	

Table 3
Seizure prediction results obtained with the Boston Children's Hospital-MIT scalp EEG dataset. The model was executed twice, and average results with standard deviations were reported. The seizure occurrence period (SOP) was 30 min and the seizure prediction horizon (SPH) was 5 min. The p value was calculated for the worst case for each patient; that is, with minimum sensitivity and maximum false prediction rate (FPR). Our seizure prediction approach achieves significantly better performance than an unspecific random predictor for all patients except Pat9, where the convolutional neural network results are only marginally better than the random predictor's.

Patient	No. of seizures	Interictal hours	Sensitivity (%)	FPR (/h)	p
Pat1	7	17	85.7 ± 0.0	0.24 ± 0.00	<0.001
Pat2	3	22.9	33.3 ± 0.0	0.00 ± 0.00	<0.001
Pat3	6	21.9	100 ± 0.0	0.18 ± 0.00	<0.001
Pat5	5	13	80 ± 20	0.19 ± 0.03	0.010
Pat9	4	12.3	50 ± 0.0	0.12 ± 0.12	0.067
Pat10	6	11.1	33.3 ± 0.0	0.00 ± 0.00	0.025
Pat13	5	14	80 ± 0.0	0.14 ± 0.00	<0.001
Pat14	5	5	80 ± 0.0	0.40 ± 0.00	0.004
Pat18	6	23	100 ± 0.0	0.28 ± 0.02	<0.001
Pat19	3	24.9	100 ± 0.0	0.00 ± 0.00	<0.001
Pat20	5	20	100 ± 0.0	0.25 ± 0.05	<0.001
Pat21	4	20.9	100 ± 0.0	0.23 ± 0.09	<0.001
Pat23	5	3	100 ± 0.0	0.33 ± 0.00	<0.001
Total	64	209	81.2 ± 1.5	0.16 ± 0.00	

Table 4
Seizure prediction results obtained with the American Epilepsy Society Seizure Prediction Challenge dataset. The model was executed twice, and average results with standard deviations were reported. The seizure occurrence period (SOP) was 30 min and the seizure prediction horizon (SPH) was 5 min. The p value was calculated for the worst case for each participant; that is, with minimum sensitivity and maximum false prediction rate (FPR). Our seizure prediction approach achieves significantly better performance than an unspecific random predictor for four of five dogs and for Pat1.

Participant	No. of seizures	Interictal hours	Sensitivity (%)	FPR (/h)	p
Dog1	4	80	50 ± 0.0	0.19 ± 0.02	0.053
Dog2	7	83.3	100 ± 0.0	0.04 ± 0.03	<0.001
Dog3	12	240	58.3 ± 0.0	0.14 ± 0.09	<0.001
Dog4	14	134	78.6 ± 0.0	0.48 ± 0.07	<0.001
Dog5	5	75	80 ± 0.0	0.08 ± 0.01	<0.001
Pat1	3	8.3	100 ± 0.0	0.42 ± 0.06	0.009
Pat2	3	7	66.7 ± 0.0	0.86 ± 0.00	0.693
Total	48	627.7	75 ± 0.0	0.21 ± 0.04	

3. Results

In this section, we test our approach with three datasets: (1) the Freiburg Hospital iEEG dataset, (2) the CHB-MIT sEEG dataset, and (3) the American Epilepsy Society Seizure Prediction Challenge iEEG dataset. An SOP of 30 min and an SPH of 5 min were used in our calculating all metrics in this work. Each fold of the leave-one-out cross-validation was executed twice, and average results with standard deviations were reported. Table 2 summarizes the seizure prediction results for the Freiburg Hospital iEEG dataset. Prediction sensitivity is 81.4% (i.e., 48 of 59 seizures are successfully predicted). The FPR is very low at 0.06/h. Our method achieves a similar sensitivity of 81.2% on the CHB-MIT sEEG dataset but with a higher FPR of 0.16/h (see Table 3). This is reasonable since sEEG recordings tend to be noisier than sEEG ones. For the American Epilepsy Society Seizure Prediction Challenge dataset, the overall sensitivity is 75% and FPR is 0.21/h (see Table 4). It is important to note that our approach works comparably with both iEEG and sEEG recordings without any denoising techniques except power line noise removal.

Table 5
Benchmarking of recent seizure prediction approaches and this work.

Year	Authors	Dataset	Feature	Classifier	Same FE[a]	No. of seizures	Sensitivity (%)	FPR (/h)	SOP	SPH
2004	Maiwald et al. (2004)	FB, 21 patients	Dynamical similarity index	Threshold crossing	Yes	88	42	< 0.15	30 min	2 min
2006	Winterhalder et al. (2006)	FB, 21 patients	Phase coherence, lag synchronization	Threshold crossing	No	88	60	0.15	30 min	10 min
2011	Park et al. (2011)	FB, 18 patients	Univariate spectral power	SVM	Yes	80	98.3	0.29	30 min	0[b]
2013	Li et al. (2013)	FB, 21 patients	Spike rate	Threshold crossing	Yes	87	72.7	0.11	50 min	10 s
2014	Zheng et al. (2014)	FB, 10 patients	Mean phase coherence	Threshold crossing	No	50	> 70	< 0.15	30 min	10 min
2014	Eftekhar et al. (2014)	FB, 21 patients	Multiresolution N-gram	Threshold crossing	No	87	90.95	0.06	20 min	10 min
2014	Aarabi and He (2014)	FB, 21 patients	Bayesian inversion of power spectral density	Rule-based decision	Yes	87	87.07	0.20	30 min	10 s
2016	Zhang and Parhi (2016)	FB, 18 patients	Power spectral density ratio	SVM	No	80	100	0.03	50 min	0[b]
2016	Zhang and Parhi (2016)	MIT, 17 patients	Power spectral density ratio	SVM	No	76	98.68	0.05	50 min	0[b]
2017	Parvez and Paul (2017)	FB, 21 patients	Phase-match error, deviation, fluctuation	LS-SVM	Yes	87	95.4	0.36	30 min	0[b]
2017	Sharif and Jafari (2017)	FB, 19 patients	Fuzzy rules on Poincaré plane	SVM	Yes	83	91.8–96.6	0.05–0.08	15min	2–42 min
2017	Aarabi and He (2017)	FB, 10 patients	Univariate and bivariate features	Rule-based decision	Yes	28	86.7	0.126	30 min	10 s
2017	Khan et al. (2017)	MIT, MSSM	Wavelet transform	CNN	Yes	131	87.8	0.14	10 min	0[b]
2017	This work	FB, 13 patients	Short-time Fourier transform	CNN	Yes	59	81.4	0.06	30 min	5 min
2017	This work	MIT, 13 patients	Short-time Fourier transform	CNN	Yes	64	81.2	0.16	30 min	5 min
2017	This work	Kaggle[c], 5 dogs, 2 patients	Short-time Fourier transform	CNN	Yes	48	75	0.21	30 min	5 min

CNN, convolutional neural network, FB, Freiburg Hospital intracranial EEG dataset; FE, feature engineering; FPR, false prediction rate; LS, least squares; MIT, Massachusetts Institute of Technology scalp EEG dataset; MSSM, Mount Sinai Hospital dataset (intracranial EEG); SPH, seizure prediction horizon; SOP, seizure occurrence period; SVM, support vector machine.

[a] Same FE across all patients. "No" means feature engineering is tailored for each patient.

[b] The authors implicitly used zero SPH and disregarded clinical considerations, and hence the results could be overestimated.

[c] American Epilepsy Society Seizure Prediction Challenge dataset (intracranial EEG).

Table 5 demonstrates a benchmark of recent seizure prediction approaches and this work. It is complicated to tell which approach is the best because each approach was tested with one dataset that is limited in the amount of data. In other words, one approach may perform well on one dataset and poorly on another. Therefore we added an extra indicator on whether the same feature engineering or feature set is applied across all patients to evaluate generalization of each method. From a clinical perspective, it is desirable to have a long enough SPH to allow effective therapeutic intervention and/or precautions. The SOP, however, should be short to minimize the patient's anxiety (Maiwald et al., 2004). Some studies that implicitly used zero SPH disregarded clinical considerations, and hence could have overestimated the prediction accuracy. The approach proposed by Park et al. (2011) achieved a very high sensitivity of 98.3% and FPR of 0.29/h in testing with 18 patients from the Freiburg Hospital dataset. Our method yields a lower sensitivity of 81.4% but a much better FPR of 0.06/h. It is nontrivial to note that the SPH was implicitly set to zero, which means prediction at a time close to or at seizure onset can be counted as a successful prediction. Likewise, research conducted by Parvez and Paul (2017) and Zhang and Parhi (2016) also implied the use of zero SPH, which will not be compared directly with our results. Among the rest of the studies listed in Table 5, Eftekhar et al. (2014) had a very good prediction sensitivity of 90.95% and a low FPR of 0.06/h for an SOP of 20 min and an SPH of 10 min. They fine-tuned the feature set for each patient to achieve the maximum performance. This, however, leads to the need for adequate expertise and time to perform the feature engineering for a new dataset. Sharif and Jafari (2017) applied the same set of features to all patients and performed classification using an SVM. This approach achieved a high sensitivity of 91.8–96.6% and a low FPR of 0.05–0.08 in testing with the Freiburg Hospital iEEG dataset. However, no studies have reported successful use of a similar approach on sEEG signals.

4. Discussion

Information extracted from EEG signals in frequency and time (synchronization) domains has been used widely to predict seizures. We proposed a novel way to exploit both frequency and time aspects of EEG signals without handcrafted feature engineering. The STFT of an EEG window has two dimensions; namely, the frequency and the time. A two-dimensional convolution filter was slid throughout the STFT to collect the changes in both the frequency and the time of EEG signals. The filter weights are automatically adjusted during the training phase and the CNN acts like a feature extraction method in an automatic fashion.

Khan et al. (2017) used the continuous wavelet transform (CWT) as a preprocessing step and used the wavelet transform of raw EEG signals as input to a CNN. In this section, we implement the same CWT and compare it with the STFT in terms of seizure prediction performance. Following Khan et al. (2017), we apply a set of ten scales from 2^0 to 2^9 and the Mexican-hat mother wavelet, then downsample the time axis of the wavelet transform so that the final dimension is $n \times 10 \times 128$. Here we use the area under the receiver operating characteristic curve (AUC) as a comparison criterion instead of sensitivity and FPR. The AUC is a threshold-free metric, so it can be used to directly compare the performance of different methods. The results are illustrated in Fig. 6. With

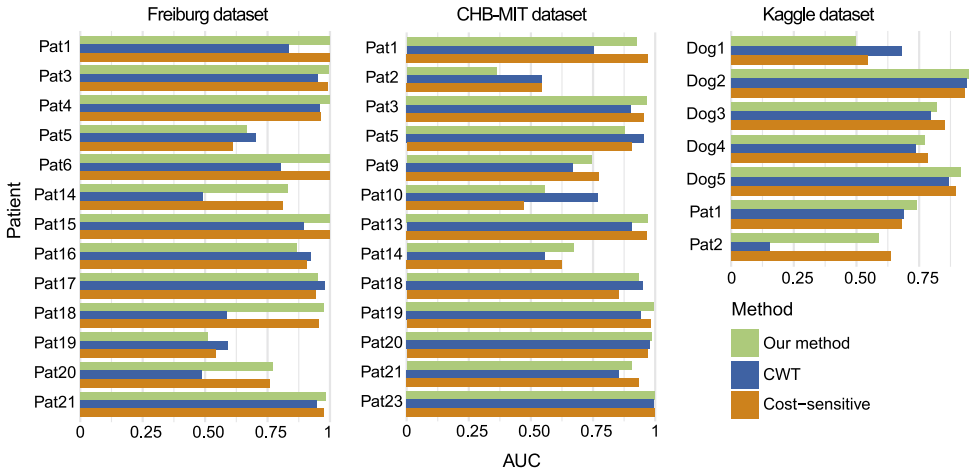

Fig. 6. Comparison among our method, the preprocessing step using the continuous wavelet transform (CWT) (Khan et al., 2017), and cost-sensitive learning. AUC, area under the receiver operating characteristic curve; CHB, Boston Children's Hospital.

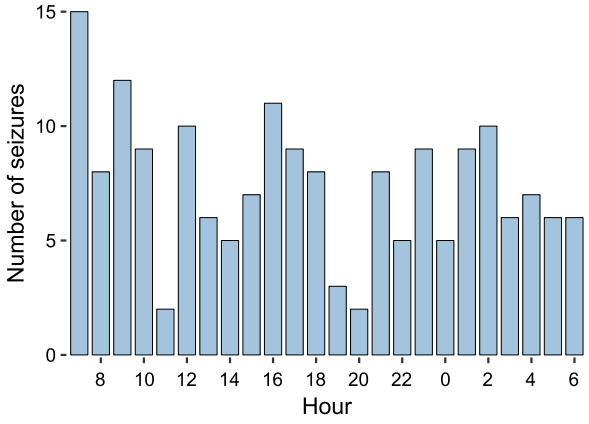

Fig. 7. Number of seizures versus time of day across patients for the Boston Children's Hospital-MIT dataset. Most seizures occur in the early morning. Two lower peaks occur around 4 p.m. and 2 a.m.

use of the Wilcoxon signed-rank test on the three datasets with a significance level of 0.05, the STFT is significantly better than the CWT, with $p = 0.0135$.

We used the oversampling technique to overcome the imbalance of the datasets. Cost-sensitive learning has been used widely in the literature for the same purpose (Branco et al., 2016). We applied cost-sensitive learning by changing the cost function in a way that the misclassification cost of preictal samples is multiplied by the ratio of interictal samples to preictal samples for each patient. We used STFT as the preprocessing step for cost-sensitive learning. The results are illustrated in Fig. 6. Although our oversampling technique does not result in a significant improvement as compared with cost-sensitive learning when applied on the three datasets, we argue that our oversampling technique is a more intuitive way to address the overfitting problem caused by the imbalance of time-series datasets.

Tables 2 and 3 show that our prediction method is significantly superior to an unspecific random predictor for all patients except Pat14 in the Freiburg Hospital dataset and Pat9 in the CHB-MIT dataset. It is worth remembering that the Freiburg Hospital dataset consists of iEEG recordings and the CHB-MIT dataset consists of sEEG recordings. In other words, our method works well with both types of EEG signals. Regarding the American Epilepsy Society Seizure Prediction Challenge dataset, our method results in significantly better performance than a random predictor for four of five dogs (see Table 4) and for Pat1.

As seizure characteristics may change over time, calibration of the seizure prediction algorithm is necessary. Minimum feature engineering has a great advantage in that it does not require an expert to carefully extract and select the optimum features for the prediction task. Hence it allows faster and more frequent updates so that patients are able to benefit the most from the seizure prediction algorithm. Also, minimum feature engineering allows seizure prediction to be available to more patients. Since the feature extraction task is undertaken by the CNN, neurophysiologists and clinical staff can spend more time in monitoring and recording EEG signals for diagnostic purposes and/or training data collection.

Our method can be further improved by non-EEG data such as information on the time when seizures occur. Epileptic seizures have been shown to have biases in distribution over time at various intervals that can be as long as 1 year or as short as 1 h (Griffiths & Fox, 1938). Importantly, Griffiths and Fox (1938) showed that there were more incidences of seizure around sunrise, noon, and midnight in their dataset of 101 patients with 39,929 seizures. However, this pattern is patient specific. Adopting the same observation, Karoly et al. (2017) leveraged this pattern to significantly improve their seizure forecasting system. Unfortunately, the three datasets investigated in this article are not large enough to assess if the time of day information is useful because the maximum recording period per patient was 3 days. Nevertheless, it is still worth seeing how incidences of seizure are distributed over the day across patients in the CHB-MIT dataset, the only dataset from which we can access the time of seizure occurrence. On the basis of the CHB-MIT data, the greatest incidence occurs in the early morning, and there are two lower peaks around 4 pm and 2 am (see Fig. 7).

5. Conclusion

Seizure prediction capability has been studied and improved over the last four decades. A perfect prediction is not yet available, but with current prediction performance it appears possible to provide patients with a warning so they can take some precautions for their safety. We proposed a novel approach of using CNNs with minimum feature engineering. The proposed approach showed its good generalization in working well with both iEEG and sEEG data. This gives more patients the opportunity to possess a seizure prediction device that can help them have a more manageable life.

Acknowledgments

This research was supported by Sydney Informatics Hub, funded by the University of Sydney via Core Research Facilities. The authors appreciate Benjamin H. Brinkmann from Mayo Systems Electrophysiology Laboratory for providing information on some unlabeled datasets. O. Kavehei acknowledges support provided by a 2018 Early Career Research grant from the Faculty of Engineering and Information Technology, University of Sydney (Grant no ECR2018Kavehei). L. Kuhlmann acknowledges support from National Health and Medical Research Council project grants APP1065638 and APP1130468. J. Yang acknowledges the National Natural Science Foundation of China for financial support under grant 61501332.

References

Aarabi, A., & He, B. (2014). Seizure prediction in hippocampal and neocortical epilepsy using a model-based approach. *Clinical Neurophysiology*, *125*(5), 930–940.

Aarabi, A., & He, B. (2017). Seizure prediction in patients with focal hippocampal epilepsy. *Clinical Neurophysiology*, *128*, 1299–1307. http://dx.doi.org/10.1016/j.clinph.2017.04.026.

Ba, J., & Caruana, R. (2014). Do deep nets really need to be deep? In Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, & K. Q. Weinberger (Eds.), *Advances in neural information processing systems* (pp. 2654–2662).

Bou Assi, E., Nguyen, D. K., Rihana, S., & Sawan, M. (2017). Towards accurate prediction of epileptic seizures: a review. *Biomedical Signal Processing and Control*, *34*, 144–157. http://dx.doi.org/10.1016/j.bspc.2017.02.001.

Branco, P., Torgo, L., & Ribeiro, R. P. (2016). A survey of predictive modeling on imbalanced domains. *ACM Computing Surveys*, *49*(2), 31. http://dx.doi.org/10.1145/2907070.

Brinkmann, B. H., Wagenaar, J., Abbot, D., Adkins, P., Bosshard, S. C., Chen, M., et al. (2016). Crowdsourcing reproducible seizure forecasting in human and canine epilepsy. *Brain*, *139*(6), 1713–1722. http://dx.doi.org/10.1093/brain/aww045.

Eftekhar, A., Juffali, W., El-Imad, J., Constandinou, T. G., & Toumazou, C. (2014). Ngram-derived pattern recognition for the detection and prediction of epileptic seizures. *PLoS One*, *9*(6), e96235. http://dx.doi.org/10.1371/journal.pone.0096235.

Freestone, D. R., Karoly, P. J., & Cook, M. J. (2017). A forward-looking review of seizure prediction. *Current Opinion in Neurology*, *30*(2), 167–173. http://dx.doi.org/10.1097/WCO.0000000000000429.

Freestone, D. R., Karoly, P. J., Peterson, A. D. H., Kuhlmann, L., Lai, A., Goodarzy, F., et al. (2015). Seizure prediction: science fiction or soon to become reality? *Current Neurology and Neuroscience Reports*, *15*(11), 73. http://dx.doi.org/10.1007/s11910-015-0596-3.

Griffiths, G. M., & Fox, J. T. (1938). Rhythm in epilepsy. *The Lancet*, *232*(5999), 409–416. http://dx.doi.org/10.1016/S0140-6736(00)41614-4.

Kaggle (2014). American Epilepsy Society Seizure Prediction Challenge. URL https://www.kaggle.com/c/seizure-prediction.

Karoly, P. J., Ung, H., Grayden, D. B., Kuhlmann, L., Leyde, K., Cook, M. J., et al. (2017). The circadian profile of epilepsy improves seizure forecasting. *Brain*, *140*(8), 2169–2182. http://dx.doi.org/10.1093/brain/awx173.

Khan, H., Marcuse, L., Fields, M., Swann, K., & Yener, B. (2017). Focal onset seizure prediction using convolutional networks. *IEEE Transactions on Biomedical Engineering*, *PP*(99), 1. http://dx.doi.org/10.1109/TBME.2017.2785401.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems*, 1097–1105.

Kuhlmann, L., Freestone, D., Lai, A., Burkitt, A. N., Fuller, K., Grayden, D. B., et al. (2010). Patient-specific bivariate-synchrony-based seizure prediction for short prediction horizons. *Epilepsy Research*, *91*(2), 214–231. http://dx.doi.org/10.1016/j.eplepsyres.2010.07.014.

Kuhlmann, L., Grayden, D. B., & Cook, M. J. (2017). Introduction. *International Journal of Neural Systems*, *27*(01), 1702001. http://dx.doi.org/10.1142/S0129065717020014.

Kuhlmann, L., Grayden, D. B., Wendling, F., & Schiff, S. J. (2015). The role of multiple-scale modelling of epilepsy in seizure forecasting. *Journal of Clinical Neurophysiology*, *32*(3), 220–226. http://dx.doi.org/10.1097/WNP.0000000000000149.

Li, S., Zhou, W., Yuan, Q., & Liu, Y. (2013). Seizure prediction using spike rate of intracranial EEG. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *21*(6), 880–886. http://dx.doi.org/10.1109/TNSRE.2013.2282153.

Maiwald, T., Winterhalder, M., Aschenbrenner-Scheibe, R., Voss, H. U., Schulze-Bonhage, A., & Timmer, J. (2004). Comparison of three nonlinear seizure prediction methods by means of the seizure prediction characteristic. *Physica D: Nonlinear Phenomena*, *194*(3–4), 357–368. http://dx.doi.org/10.1016/j.physd.2004.02.013.

Mirowski, P. W., LeCun, Y., Madhavan, D., & Kuzniecky, R. (2008). Comparing SVM and convolutional networks for epileptic seizure prediction from intracranial EEG. In *2008 IEEE workshop on machine learning for signal processing* (pp. 244–249). http://dx.doi.org/10.1109/MLSP.2008.4685487.

Park, Y., Luo, L., Parhi, K. K., & Netoff, T. (2011). Seizure prediction with spectral power of EEG using cost-sensitive support vector machines. *Epilepsia*, *52*(10), 1761–1770. http://dx.doi.org/10.1111/j.1528-1167.2011.03138.x.

Parvez, M. Z., & Paul, M. (2017). Seizure prediction using undulated global and local features. *IEEE Transactions on Biomedical Engineering*, *64*(1), 208–217. http://dx.doi.org/10.1109/TBME.2016.2553131.

Rogowski, Z., Gath, I., & Bental, E. (1981). On the prediction of epileptic seizures. *Biological Cybernetics*, *42*(1), 9–15. http://dx.doi.org/10.1007/BF00335153.

Sainath, T. N., Mohamed, A.-R., Kingsbury, B., & Ramabhadran, B. (2013). Deep convolutional neural networks for LVCSR. In *2013 IEEE international conference on acoustics, speech and signal processing* (pp. 8614–8618). http://dx.doi.org/10.1109/ICASSP.2013.6639347.

Salant, Y., Gath, I., & Henriksen, O. (1998). Prediction of epileptic seizures from two-channel EEG. *Medical & Biological Engineering & Computing*, *36*(5), 549–556. http://dx.doi.org/10.1007/BF02524422.

Schelter, B., Winterhalder, M., Maiwald, T., Brandt, A., Schad, A., Schulze-Bonhage, A., et al. (2006). Testing statistical significance of multivariate time series analysis techniques for epileptic seizure prediction. *Chaos. An Interdisciplinary Journal of Nonlinear Science*, *16*(1), 13108. http://dx.doi.org/10.1063/1.2137623.

Sharif, B., & Jafari, A. H. (2017). Prediction of epileptic seizures from EEG using analysis of ictal rules on Poincaré plane. *Computer Methods and Programs in Biomedicine*, *145*, 11–22. http://dx.doi.org/10.1016/j.cmpb.2017.04.001.

Shoeb, A. H. (2009). *Application of machine learning to epileptic seizure onset detection and treatment* (Ph.D. thesis), Massachusetts Institute of Technology.

Sinha, N., Dauwels, J., Kaiser, M., Cash, S. S., Brandon Westover, M., Wang, Y., et al. (2017). Predicting neurosurgical outcomes in focal epilepsy patients using computational modelling. *Brain*, *140*(2), 319–332. http://dx.doi.org/10.1093/brain/aww299.

Szostak, K. M., Grand, L., & Constandinou, T. G. (2017). Neural interfaces for intracortical recording: requirements, fabrication methods, and characteristics. *Frontiers in Neuroscience*, *11*, 665. http://dx.doi.org/10.3389/fnins.2017.00665.

University of Freiburg, 2003. EEG Database at the Epilepsy Center of the University Hospital of Freiburg, Germany. URL http://epilepsy.uni-freiburg.de.

Winterhalder, M., Schelter, B., Maiwald, T., Brandt, A., Schad, A., Schulze-Bonhage, A., et al. (2006). Spatio-temporal patient–individual assessment of synchronization changes for epileptic seizure prediction. *Clinical Neurophysiology*, *117*(11), 2399–2413. http://dx.doi.org/10.1016/j.clinph.2006.07.312.

Xiao, C., Wang, S., Iasemidis, L., Wong, S., & Chaovalitwongse, W. A. (2017). An adaptive pattern learning framework to personalize online seizure prediction. *IEEE Transactions on Big Data*, *PP*(99), 1. http://dx.doi.org/10.1109/TBDATA.2017.2675982.

Zhang, Z., & Parhi, K. K. (2016). Low-complexity seizure prediction from iEEG/sEEG using spectral power and ratios of spectral power. *IEEE Transactions on Biomedical Circuits and Systems*, *10*(3), 693–706. http://dx.doi.org/10.1109/TBCAS.2015.2477264.

Zheng, Y., Wang, G., Li, K., Bao, G., & Wang, J. (2014). Epileptic seizure prediction using phase synchronization based on bivariate empirical mode decomposition. *Clinical Neurophysiology*, *125*(6), 1104–1111. http://dx.doi.org/10.1016/j.clinph.2013.09.047.