University Library



A gateway to Melbourne's research publications

Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:
Kiral-Kornek, I;Roy, S;Nurse, E;Mashford, B;Karoly, P;Carroll, T;Payne, D;Saha, S;Baldassano, S;O'Brien, T;Grayden, D;Cook, M;Freestone, D;Harrer, S

Title:
Epileptic Seizure Prediction Using Big Data and Deep Learning: Toward a Mobile System

Date:
2018-01-01

Citation:
Kiral-Kornek, I., Roy, S., Nurse, E., Mashford, B., Karoly, P., Carroll, T., Payne, D., Saha, S., Baldassano, S., O'Brien, T., Grayden, D., Cook, M., Freestone, D. & Harrer, S. (2018). Epileptic Seizure Prediction Using Big Data and Deep Learning: Toward a Mobile System. Ebiomedicine, 27, pp.103-111. https://doi.org/10.1016/j.ebiom.2017.11.032.

Persistent Link:
https://hdl.handle.net/11343/234004

License:
CC BY-NC-ND

EBioMedicine 27 (2018) 103–111

Contents lists available at ScienceDirect

EBioMedicine

journal homepage: www.ebiomedicine.com

Research Paper

Epileptic Seizure Prediction Using Big Data and Deep Learning: Toward a Mobile System

Isabell Kiral-Kornek [a,1], Subhrajit Roy [a,1], Ewan Nurse [a,b], Benjamin Mashford [a], Philippa Karoly [a,b], Thomas Carroll [a,b], Daniel Payne [a,b], Susmita Saha [a], Steven Baldassano [b], Terence O'Brien [b], David Grayden [c], Mark Cook [b], Dean Freestone [b,*], Stefan Harrer [a,*]

[a] IBM Research – Australia, 204 Lygon Street, 3053 Carlton, VIC, Australia
[b] The University of Melbourne, 3010 Parkville, VIC, Australia
[c] Department of Biomedical Engineering, The University of Melbourne, 3010 Parkville, VIC, Australia

ARTICLE INFO

Article history:
Received 17 August 2017
Received in revised form 16 November 2017
Accepted 28 November 2017
Available online 12 December 2017

Keywords:
Epilepsy
Seizure prediction
Artificial intelligence
Deep neural networks
Mobile medical devices
Precision medicine

ABSTRACT

Background: Seizure prediction can increase independence and allow preventative treatment for patients with epilepsy. We present a proof-of-concept for a seizure prediction system that is accurate, fully automated, patient-specific, and tunable to an individual's needs.
Methods: Intracranial electroencephalography (iEEG) data of ten patients obtained from a seizure advisory system were analyzed as part of a pseudoprospective seizure prediction study. First, a deep learning classifier was trained to distinguish between preictal and interictal signals. Second, classifier performance was tested on held-out iEEG data from all patients and benchmarked against the performance of a random predictor. Third, the prediction system was tuned so sensitivity or time in warning could be prioritized by the patient. Finally, a demonstration of the feasibility of deployment of the prediction system onto an ultra-low power neuromorphic chip for autonomous operation on a wearable device is provided.
Results: The prediction system achieved mean sensitivity of 69% and mean time in warning of 27%, significantly surpassing an equivalent random predictor for all patients by 42%.
Conclusion: This study demonstrates that deep learning in combination with neuromorphic hardware can provide the basis for a wearable, real-time, always-on, patient-specific seizure warning system with low power consumption and reliable long-term performance.
© 2017 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

Epilepsy is singularly unusual among other serious neurological conditions because seizures are brief and infrequent, so that for at least 99% of the time patients are unaffected by seizure activity. Although seizure activity is infrequent, the disability caused by epilepsy can be significant due to the uncertainty around the occurrence and the consequences of the events. The constant uncertainty impairs the quality of life for these individuals. A recent survey confirmed that the majority of patients find this unpredictability to be the most debilitating aspect of epilepsy ("2016 Community Survey," 2016). There is an unmet need for a device that provides a warning when there is an increased risk of a seizure.

A warning system could support new treatment approaches and improve a patient's quality of life. For example, such a system could inform patients' daily routines and help them to avoid dangerous situations when at higher risk of seizure. Tracking fluctuations in seizure likelihood could also be used to titrate therapeutic interventions, reducing the time spent using anti-epileptic drugs or electrical stimulation.

Given the nature of epilepsy, there are undeniably technological and theoretical hurdles to creating a viable warning system for seizures; however, such a system is no longer considered impossible to build (Freestone et al., 2015; Mormann and Andrzejak, 2016). A key development has been the use of long-term electroencephalography (EEG) data. After a long-term clinical trial, Cook et al. were able to demonstrate success of an implantable recording system, seizure prediction algorithm, and handheld patient advisory device (Cook et al., 2013). Using this device, the group recorded a dataset that comprises a total of over 16 years of continuous intracranial electroencephalography (iEEG) recording and thousands of seizures. Cook et al. established the feasibility of seizure prediction in a clinical setting, and provided inspiration for the development of further seizure prediction algorithms (Freestone et al., 2017).

Despite the trial's success, there were also limitations (Elger and Mormann, 2013). While pre-seizure patterns in the iEEG data were

* Corresponding authors.
E-mail addresses: deanrf@unimelb.edu.au (D. Freestone), sharrer@au.ibm.com (S. Harrer).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.ebiom.2017.11.032
2352-3964/© 2017 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

extracted in an automated fashion, it was based on a limited and predefined set of features, which may be one reason that prediction was not possible for all patients. After the initial design phase, the algorithm was no longer tunable, making the system inflexible to patients' changing preferences regarding false alarm and missed seizure rates.

As preictal patterns are patient specific, no pre-determined set of features will be able to capture all possible preictal signatures. Therefore, standard feature engineering techniques are unsuitable for the creation of a generalizable predictor (Freestone et al., 2017). Instead of restricting the feature space a-priori, all data should be considered potentially relevant for recognizing preictal patterns – a task to which novel computational techniques are uniquely suited.

Deep learning, a machine learning technique, is a powerful computational tool that enables features to be automatically learnt from data (LeCun et al., 2015). Typically, deep learning is used to train a class of algorithms known as deep neural networks to perform specific tasks. The availability of big data has cemented the usefulness of deep learning for a diverse range of problems (LeCun et al., 2015). Applications range from self-driving cars via robotics to novel diagnostic and treatment options in medical imaging, healthcare, and genomics ("FACT SHEET," 2016; Gulshan et al., 2016; Litjens et al., 2017; Ratner, 2015; Stebbins, 2016). Recent open source seizure prediction competitions (Brinkmann et al., 2016; "Melbourne University AES/MathWorks/NIH Seizure Prediction | Kaggle," 2016) have shown that machine learning techniques are able to produce pre-eminent results, suggesting this method may provide a path to clinical translation of seizure prediction devices. However, the best performing algorithms in competitions often require an unrealistic amount of computing resources for a wearable device ("Melbourne University AES/MathWorks/NIH Seizure Prediction | Kaggle," 2016).

For seizure prediction to be implemented in a clinical device, it is necessary for algorithms to run on small, low-power technology. A number of recent advances in computing led to the development of sophisticated deep learning algorithms using ultra-low power chips (Furber, 2016). One example of such a chip is IBM's TrueNorth Neurosynaptic System (Esser et al., 2016; Merolla et al., 2014). TrueNorth is a specialized chip capable of implementing artificial neural networks in hardware and hence it is neuromorphic in nature. It is one of the most power-efficient chips to date, consuming <70 mW power at full chip utilization. The chip's neuromorphic technology allows for the deployment and testing of algorithms that were previously unrealizable in a clinically viable seizure warning system.

Seizure prediction has been established as clinically feasible and highly desirable for patients. In light of promising results (Brinkmann et al., 2016; Cook et al., 2013; Howbert et al., 2014), the development of a practical seizure warning device has been declared a grand challenge in epilepsy management ("Seizure Gauge Challenge," 2017). In this paper, we describe how deep learning and the TrueNorth processor can be leveraged to advance the task of patient-specific seizure prediction. Prediction results were benchmarked using data recorded during the trial undertaken by Cook and colleagues (Cook et al., 2013). The presented results address several limitations of this earlier study, and provide proof-of-concept for a deep learning system for seizure prediction.

2. Materials and Methods

The overall study design is shown in Fig. 1. The iEEG signal is recorded using intracranial electrodes (magenta circles). Annotated iEEG signals are processed by a deep neural network that is trained to distinguish between preictal and interictal signals. The resulting deep learning model is subsequently deployed onto the neuromorphic TrueNorth chip.

2.1. System Design Rationale

The objective of this study is the development, implementation, and evaluation of a clinically relevant seizure prediction system. In order for a system to be valuable to patients while being maintainable by clinicians, we defined the following goals:

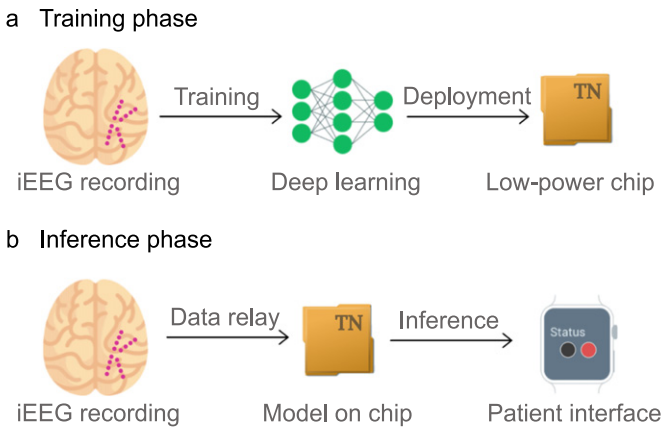


Fig. 1. Concept of seizure advisory system: a) Training phase: iEEG signal is recorded via intracranial electrodes (magenta circles indicate a possible configuration) and recordings are passed on to a deep learning network (green network graph). The model is subsequently deployed onto a TrueNorth chip. b) Inference phase: iEEG signal is recorded via intracranial electrodes (magenta circles) and recordings are passed on to the TrueNorth chip. Prediction of a seizure is indicated to the patient on a wearable device.

G1. The system needs to perform well and reliably across patients.
G2. The system needs to operate autonomously over long periods of time without a requirement for regular maintenance or reconfiguration by an expert.
G3. The system needs to allow for patients to set personal preferences with respect to sensitivity.
G4. The system must run in real-time on a low-power platform.

We addressed performance (G1) and long-term feasibility (G2) using deep learning, a technique that, in contrast to a more traditional feature engineering approach, does not rely on data analysis experts for the monitoring and adaptation of models. Unlike traditional computing systems that learn through instructions or explicit programming, deep learning algorithms learn from examples to automatically discriminate different classes of signals. In the context of a seizure prediction system, this is what allows the algorithm to distinguish between preictal and interictal data segments. By its nature, a system using an artificial neural network cannot only adjust to each individual patient's brain signals, but also to short- and long-term changes in the recording. It further allows for the integration of other patient-specific variables that have been shown to co-vary with seizure likelihood, such as time of day information. Moreover, a deep neural network can automatically learn to discriminate between different classes of data, for example, in this case, preictal and interictal.

Generally, a classification neural network such as the one used in our study will classify the signal on a sample-by-sample basis, leading to potentially very frequent but short alarms. In a real-time system, an additional processing layer is therefore required to balance the sensitivity of the system, number, and duration of alarms. In addition to forming the basis of system optimization, this processing layer also allows for instantaneous tuning of the system's sensitivity by the patient directly (G3).

Adaptation to changes of the signal over time (G2), as for example observed by Cook et al. (2013) were addressed in both processing layers.

Running a neural network classifier in a real-time environment requires specialized hardware (G4). TrueNorth is a highly power-

efficient and specialized chip. The network needed to be adapted to run on the TrueNorth chip.

2.2. Data

Data for the study were collected for a previous clinical trial of an implanted seizure advisory system (Cook et al., 2013). The iEEG of the enrolled patients were continuously recorded for up to two years using an implanted 16-electrode iEEG system. Data were reviewed by expert investigators and all seizures were annotated. Seizures were labelled as either clinical or clinically equivalent (see (Cook et al., 2013) for a detailed description of the data). In this study, we used the data of all ten patients that were included in the prospective trial conducted by Cook and colleagues (Cook et al., 2013). The data comprised a total of 16.29 years of iEEG signal and 2817 seizures. Sections of the data will shortly be made publicly available.

2.3. Learning Pre-seizure Patterns

Data segments were transformed into a time-frequency representation (spectrograms). Including information about the circadian patterns of seizure occurrence can improve prediction performance (Karoly et al., 2016, 2017); therefore, hour of day labels were also incorporated into the spectrograms.

The system operated in two distinct phases, as depicted in Fig. 1. During the *training* phase (Fig. 1a), previously labelled data were used to train a deep neural network to distinguish between preictal (defined as occurring within the 15 min before a seizure) and interictal (defined as anything that is neither preictal nor ictal) data. Training was performed on a dataset containing the same number of preictal and interictal samples to ensure unbiased learning of features. During the *inference* phase (Fig. 1b), the trained deep learning model was used to classify incoming data into preictal and interictal classes in a pseudoprospective and continuous manner using all data recorded after the training period.

Initially, two months of iEEG data, containing at least one seizure for each patient, was used for algorithm training and calibration. Following the initial training, a new model was trained after each month of incoming data. We found that non-stationarities in the iEEG signal (Sillay et al., 2013; Ung et al., 2017) were detrimental to the model's performance, which led us to devise a protocol in which data older than a certain number of months is excluded from the training set. The resulting model was used to make predictions for the following month of data. This procedure ensured that inference always occurred chronologically after training (i.e., pseudoprospective). For more details on data selection, data processing, and the training protocol we invite the reader to refer to Section 2 in the supplementary information.

Training and inference for all ten patients were developed using a high-performance computer. A full deployment onto the neuromorphic TrueNorth chip was undertaken for one patient to provide proof-of-concept of low-power system functionality.

2.4. Enabling Real-time Tunability

During system operation, the deep neural network classifies each incoming data segment as either preictal or interictal. To determine whether any given sample should lead to an alarm for the patient, an artificial leaky integrate-and-fire neuron was implemented. This ensured that alarms were only enabled if several preictal predictions were made in close temporal proximity. Parameters for this processing layer (firing threshold and leak of the neuron and length of an alarm) were optimized to yield the best performance as determined by an objective that could be set automatically or by a user. It is this layer that can give a patient or clinician control over which metric they want to prioritize. Unless stated otherwise, parameters were optimized to ensure optimal performance as defined below. Optimization occurred in a pseudoprospective manner and parameters were updated after each month of incoming data.

2.5. Performance Evaluation

To evaluate seizure prediction performance, we used the metrics introduced in (Cook et al., 2013). These were sensitivity (true positive seizure prediction rate), time in warning (TiW, total duration of a red-light indicator), and sensitivity improvement over chance (IoC). IoC is determined by comparing our system to a random predictor that spends an equal amount of time in warning and computing the difference of the achieved sensitivities. These metrics provide a clinically relevant indication of performance (Mormann et al., 2007). In this work, we report mean prediction scores, as well as monthly performance, starting after a short initial data collection phase. Results were obtained for three independent runs as per Section 2.2, for which we report the mean performance as well as the 95% confidence interval. A more detailed description of all metrics and their computation can be found in Section 3 in the supplementary information.

3. Results

3.1. Full System Implementation

Mean values and 95% confidence intervals were determined for three independent repetitions of training and inference. Inference results, starting in month 3, are shown in Fig. 2 for all patients and are summarized in Table 1.

The performance of the proposed system in terms of sensitivity, improvement over chance (IoC), and time in warning (TiW) are benchmarked against results reported by Cook and colleagues (Cook et al., 2013). The outcome is displayed in Fig. 2. Note that results reported by Cook et al. were derived from substantially fewer inference days (mean of 16.3% per patient), as indicated by the shorter horizontal blue lines in Fig. 2. On the other hand, we show our results on an average of 89.0% of the data per patients for the same cohort.

Performance was computed for three independent runs of the system as outlined in Section 2.2. Fig. 2 shows the average performance across time, as well as the monthly performance averaged across runs and 95% confidence intervals (CI). These results demonstrate that the system could be put into use for all patients after only two months of initial data acquisition.

For all patients, seizure prediction was significantly better than chance for most of the months that were evaluated. Significant IoC suggests that seizure prediction will be useful for patients in a clinical setting (Mormann and Andrzejak, 2016). The mean improvement over chance was 42·3% (standard deviation 13·6%) across all months and patients. Mean performances were significantly above chance for all patients as computed using the method described in (Snyder et al., 2008) ($p < 0{\cdot}0001$ for eight of the patients, $p < 0{\cdot}002$ for one patient, and $p < 0{\cdot}01$ for the remaining patient). The mean sensitivity was 68·6% (standard deviation 11·1%) and the system spent an average of 26·9% (standard deviation 14·4%) of the time in the warning state.

From a patient's perspective, it is of interest to investigate the durations of alarms and the prediction horizon (the expected wait time between alarm onset and seizure onset). Fig. 3 summarizes the statistics of the seizure forecasting alarms. Fig. 3a and b show the distributions of alarm time durations for true alarms and false alarms, respectively. For all patients, 75% of alarms lasted between 9 min and 5·9 h. False alarm periods were generally shorter than true positive alarm periods. Fig. 3c shows the distribution of seizure prediction horizons for each patient. It can be seen that the median prediction horizon was generally less than 1 h (with the exceptions of patients 9 and 15). The majority of prediction horizons (75%) were between 4·7 min and 4·8 h for all patients. The entire range of alarms lay between 1 min and 11·9 h.

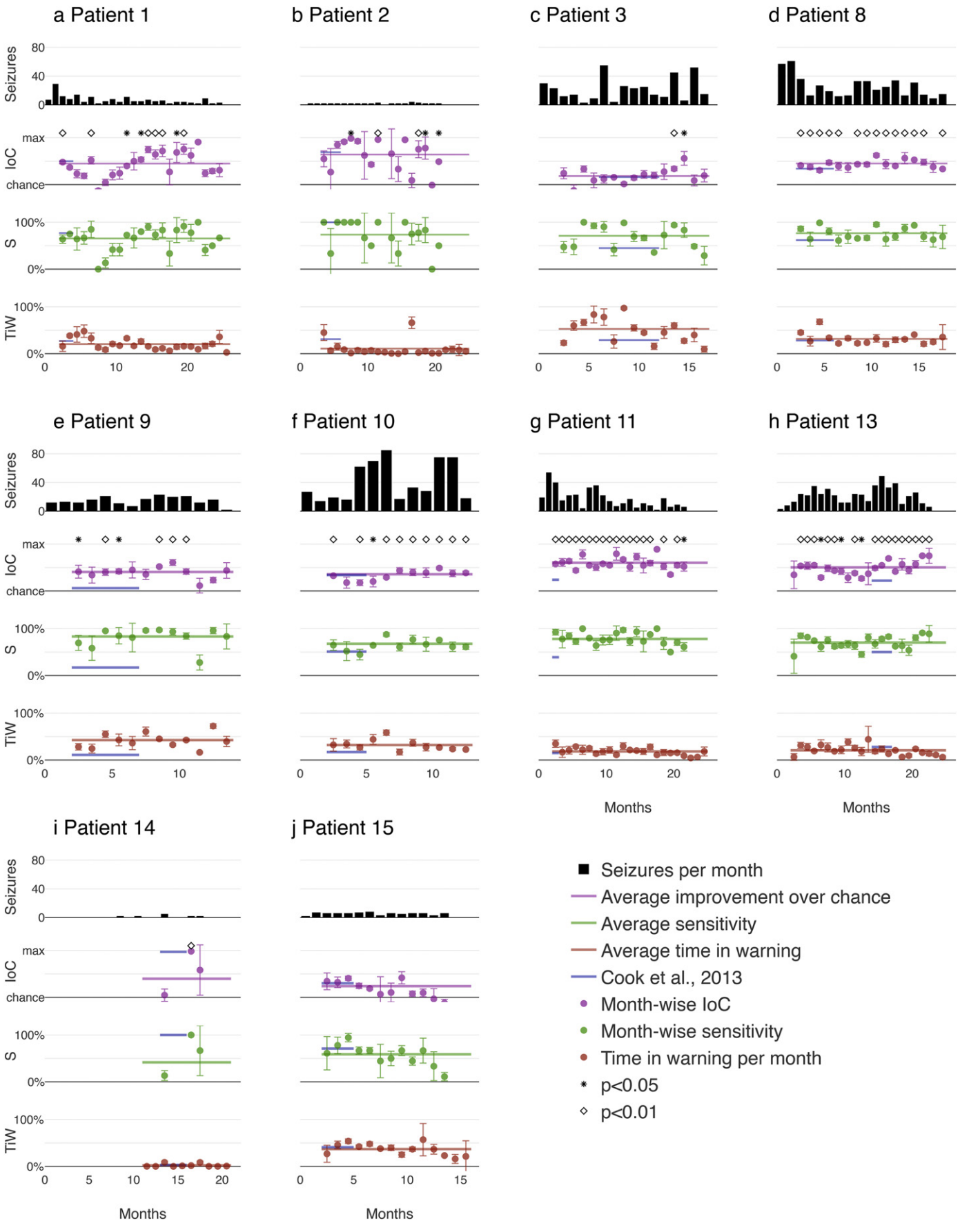

a Patient 1
b Patient 2
c Patient 3
d Patient 8
e Patient 9
f Patient 10
g Patient 11
h Patient 13
i Patient 14
j Patient 15
Seizures
80
40
0
max
IoC
chance
100%
S
0%
100%
TiW
0%
0
10
20
5
15
Months
Seizures per month
Average improvement over chance
Average sensitivity
Average time in warning
Cook et al., 2013
Month-wise IoC
Month-wise sensitivity
Time in warning per month
p<0.05
p<0.01

Table 1
Summary of long-term performance of seizure warning system for all patients: number of total seizures, seizure rates, and performance values (mean and standard deviation) over all months, as well as significance values for three individual runs. (IoC: improvement over chance, TiW: time in warning.)

Patient number	Total number of seizures	Seizure rate per month	Mean (std) IoC in %	Mean (std) Sensitivity in %	Mean (std) TiW in %	*p*-value (IoC) all runs
1	151	5·9	45·0 (21·1)	65·4 (19·3)	20·5 (11·5)	<0·00001
2	32	1·3	64·1 (43·2)	73·6 (23·4)	10·8 (16·8)	<0·00001
3	368	19·8	18·3 (14·8)	71·1 (23·3)	52·8 (23·7)	<0·00001
8	466	25·0	45·2 (8·9)	76·7 (13·1)	31·5 (12·1)	<0·00001
9	204	15·5	40·5 (14·3)	83·1 (19·8)	42·6 (15·2)	<0·00001
10	545	43·7	35·6 (10·9)	67·7 (13·7)	32·0 (11·4)	<0·00001
11	464	19·3	60·2 (8·8)	77·9 (11·1)	18·4 (7·9)	<0·00001
13	498	20·0	50·3 (11·4)	70·3 (12·2)	20·8 (9·9)	<0·00001
14	12	0·6	39·7 (46·7)	41·7 (46·3)	2·2 (3·5)	<0·01
15	77	5·0	23·9 (18·8)	58·8 (21·7)	36·9 (12·4)	<0·002

The distributions in Fig. 3 are computed using the raw system output. Results could potentially be improved by deploying patient-specific requirements, such as disabling the alarm for some period after the onset of a seizure or during sleep.

3.2. Case Study on Individual Tunability: Balancing Sensitivity and Time in Warning

Ultimately, whether a seizure prediction system is clinically useful may depend on a patient's individual preferences regarding sensitivity and duration and number of alarms. Our model allows for tuning of the system to account for clinicians' or patients' priorities by adjusting the relative weight between sensitivity and time in warning. Fig. 4 displays an example implementation of the tunable system for patient 9. Priority was given either to the largest possible improvement over chance (Fig. 4a), high sensitivity by weighing the relative importance of sensitivity to time in warning 3:1 (Fig. 4b), or low time spent in warning state by weighing the relative importance of sensitivity to time in warning 1:3 (Fig. 4c). In a real-world use-case scenario, a patient or clinician could easily set which metric they want to prioritize and to what extent by changing a single model parameter accessible through an interface, as indicated in the top row of Fig. 4.

Results summarized in Table 2 demonstrate the successful implementation of the prioritization functionality, with the prioritized metric being either at the highest in the case of improvement over chance (40·5%) and sensitivity (91·8%) or lowest in the case of time in warning (19·3%).

3.3. Deployment of the System Onto the Ultra-low Power TrueNorth Chip

After completing the pseudoprospective study, we demonstrated that the developed system can be deployed onto the ultra-low power neuromorphic TrueNorth chip by repeating the pseudoprospective study using the chip and exemplarily choosing patient 13. The results are shown in Fig. 5 and Table 3. The mean improvement over chance was 41·3%, which is 9% less than the mean results obtained on the high-performance computer (50·3%). Average sensitivity across all months was 71·7% with 31·7% time in warning.

4. Discussion

People with uncontrolled epilepsy often live with uncertainty about when a seizure is going to occur. It is this uncertainty that can lead to difficulties with daily life activities, such as driving, working, or even socialising ("2016 Community Survey," 2016), and may expose an individual to unnecessary danger. By presenting information about when a seizure is likely to happen, we hope to restore a degree of perceived control.

In this article, we have proposed a deep learning approach to seizure prediction that addresses many of the challenges identified from previous studies (Cook et al., 2013; Freestone et al., 2017; "Melbourne University AES/MathWorks/NIH Seizure Prediction | Kaggle," 2016; Ung et al., 2017). Our results provide a proof-of-concept of a robust, real-time, low-power seizure prediction system that can be configured by patients according to their needs and preferences. Prediction algorithms could be deployed for all patients after only two months of collecting iEEG data that contained at least one seizure per month, resulting in a system that can be beneficial to patients in minimal time. Our automatic retraining protocol makes accurate seizure prediction possible across patients despite long-term changes in brain signals. The developed prediction system can be run on an ultra-low power chip and can be integrated into a mobile system.

To evaluate our approach, we have benchmarked the proposed system against three studies, summarized in Table 4.

Both algorithms (Cook et al., 2013; Karoly et al., 2017), extract predefined features from the iEEG signal and use conventional machine learning techniques. In addition, Karoly et al. combine the classifier's output with insights obtained from circadian rhythms of seizures. (Karoly et al., 2017) also reported results on a classifier using patient-specific circadian information only which we included as well to demonstrate that our results cannot be attributed to the time of day feature only.

When measured based on improvement over chance, our proposed algorithm performs better than (Cook et al., 2013) for 6 out of 10 patients and better than (Karoly et al., 2017) for 6 out of 8 reported patients. It is important to note that results reported in (Cook et al., 2013) and (Karoly et al., 2017) were derived from substantially fewer inference days. While Cook et al. reported testing performance on 16.3% of the data (Cook et al., 2013), we show our results on 89.0% of the data for the same cohort. Using small amounts of data for testing makes it impossible to assess long-term performance and address the impact of changing brain signals. Moreover, while Karoly et al. have reported results for 62.2% of the data (Karoly et al., 2017), we use 94.4% of the data for the same subset of patients. The difference stems from Karoly et al. discarding 100 days to allow the signal to stabilize and using the following 100 days of data to train the algorithm (Karoly et al., 2017). This will lead to a longer waiting period for patients between implantation and operation of the device as a seizure predictor.

Fig. 2. Pseudoprospective long-term prediction results: Traces from top to bottom of each plot indicate the number of seizures per month, improvement over chance, sensitivity, and time in warning. Horizontal lines depict performance averaged across months. Dark blue lines represent performances and evaluation periods from the previous study published by Cook and colleagues (Cook et al., 2013), given here for comparison; note that direct comparisons of the values are not possible due to the difference in number of seizures in the test set, which are indicated by the lengths of the dark blue line segments representing the portions of the data that were used in (Cook et al., 2013) for each patient. Circles represent mean performance averaged over three runs. Vertical bars denote 95% confidence intervals. Month scales are different for each patient.

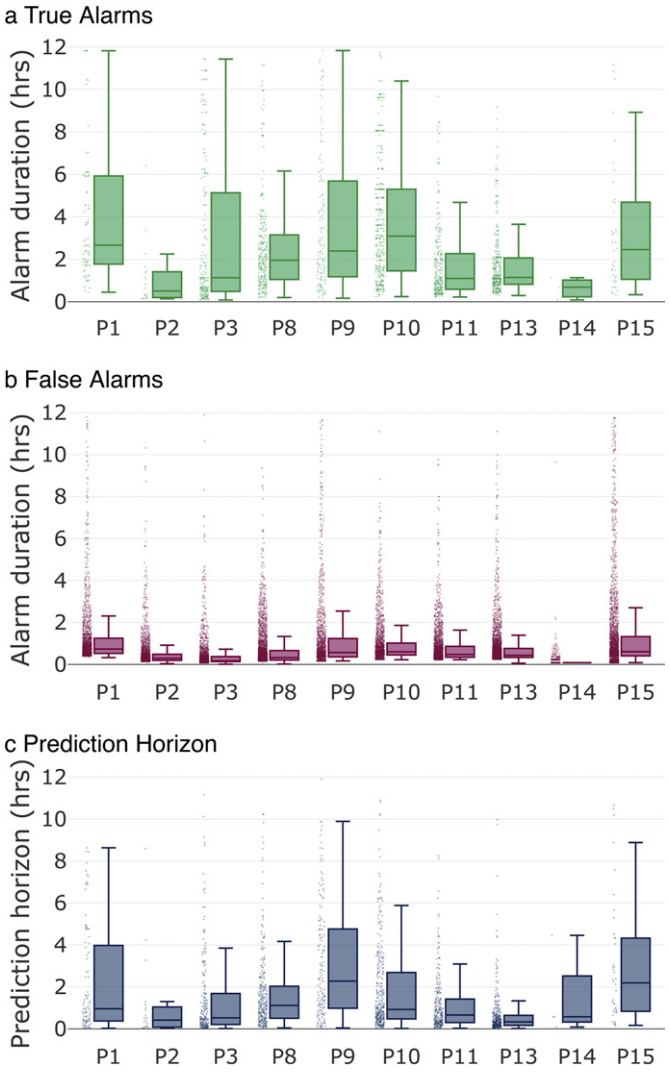


Fig. 3. Alarm duration and prediction horizon for all patients: Dots show all individual data points, solid lines indicate medians, box tops indicate 75th percentiles, box bottoms indicate 25th percentiles, whiskers indicate the span of the data after removal of outliers. a) Alarm durations of true alarms in hours. b) Alarm durations of false alarms in hours. c) Prediction horizons in hours. A prediction horizon is the time between alarm and seizure onset.

While Cook et al. showed that seizure prediction is possible using feature engineering techniques (Cook et al., 2013), the primary criticisms of their world-first, long-term study was varying prediction performance across patients and a lack of tunability to accommodate patient preferences (Elger and Mormann, 2013). This holds equally true for the algorithm described in (Karoly et al., 2017). Our proposed system includes algorithmic components to overcome all these limitations.

On the other hand, the system presented in (Cook et al., 2013) had the ability to forecast not only imminent seizures, but also times of low seizure likelihood for some patients. To implement this feature, a classifier trained to predict seizures was repurposed. However, this was only successful for four out of ten patients. This suggests that predicting times of low seizure risk requires a more sophisticated approach than simply applying the model that was used to detect high-risk periods. In our work, deep neural networks were trained specifically to detect times of high seizure likelihood. We expect that detecting times of low seizure likelihood will require additional models tuned to different signal features (i.e., the absence of preictal signal features does not necessarily equate to low risk). Therefore, an important extension to the current system is to include additional networks trained specifically to detect periods of safety.

In addition, a recent competition on Kaggle used a subset of the data used by us, achieving an area under the receiver operating characteristic (ROC) curve of up to 0.75 ("Melbourne University AES/MathWorks/NIH Seizure Prediction | Kaggle," 2016). However, there were limitations to these analyses. First, the results were based on a limited subset of recordings. Second, clinically relevant metrics such as sensitivity, improvement over chance, and time in warning were not reported in the contest. Third, winning algorithms relied on highly complex features that could not be deployed in currently available implantable or wearable devices. In contrast, the method presented here has requirements for power, size, and computation that can be realized within a wearable device.

Our results demonstrate that there is a preictal signature in iEEG data. Even though this is impossible for the human eye to detect, a deep neural network can create a model that correctly identifies this signature. We found this signature to be patient-specific and non-stationary. Following conventional non-deep learning techniques, a changing signal would require regular expert intervention to update the model. In contrast, the deep-learning pipeline we presented is designed to create models in an automatic fashion that can be updated without medical expert supervision.

Deep learning algorithms rely on large amounts of data to automatically learn or extract important features. The closer any dataset

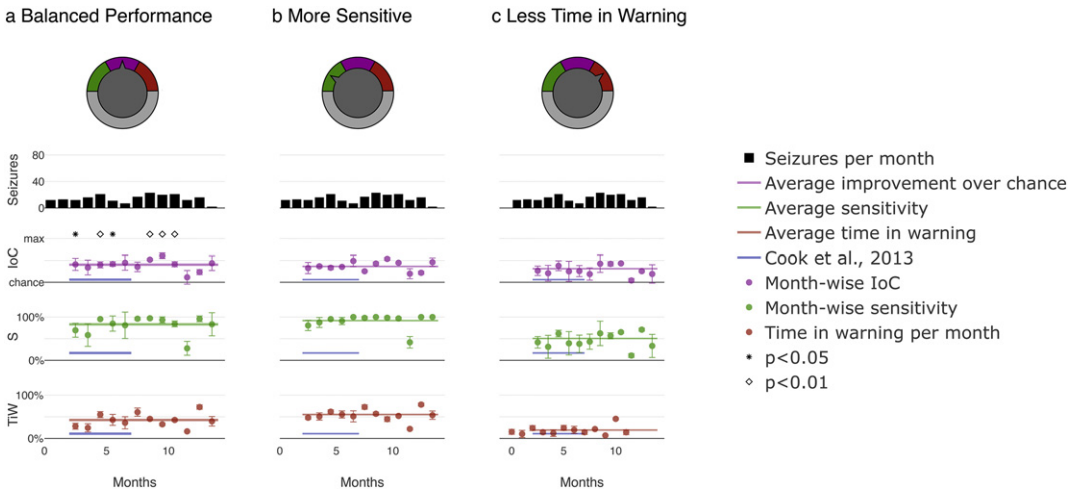


Fig. 4. Pseudoprospective long-term seizure prediction study for patient 9 - month-wise system output for the entire duration of the study and different priority settings: Customization was achieved through tuning the ratio of weights assigned to sensitivity and time in warning using a tuning factor. a) 'Balanced performance' prioritizes improvement over chance. b) 'More sensitive' prioritizes sensitivity over time in warning. c) 'Less time in warning' prioritizes time in warning. Note improved sensitivity in b and reduced time in warning in c. Plot details are described in Fig. 2.

Table 2
Pseudoprospective long-time seizure prediction study for patient 9: mean performance over all months using different relative weights of sensitivity to time in warning.

System mode	Mean (std) IoC in %	Mean (std) sensitivity in %	Mean (std) TiW in %	p-Value (IoC) all runs
Balanced performance S:TiW = 1:1	40·5 (14·3)	83·1 (19·8)	42·6 (15·2)	<0·00001
More sensitive S:TiW = 3:1	36·5 (11·5)	91·8 (15·2)	55·2 (12·6)	<0·00001
Less TiW S:TiW = 1:3	31·2 (13·3)	50·6 (17·0)	19·3 (9·5)	<0·00001

represents all possible variations of the signal to be identified, the better the model will perform under real-life conditions. This means that generally more data leads to better performance. In a typical deep-learning application, such as image recognition, data sets frequently contain at least thousands of unique training examples (LeCun et al., 2015). Despite having access to a uniquely large amount of iEEG data, the total number of training samples used in our study was still orders of magnitudes smaller. This makes deep learning-based iEEG classification a challenging task, especially for patients with small numbers of seizures. For example, Patient 14 (Fig. 2i) had a seizure rate of 0.58 seizures per month (see Table 1), which means that the initial model was trained on only four seizures. It is not surprising that the performance of our deep-learning model fell short of the performances of manually tuned algorithms for this patient. Although higher than for Patient 14, seizure rates for Patients 1, 2, and 15 were also far below the average of this study, which may explain why the performance of our deep learning model matched but did not exceed that of manually designed algorithms for these three patients. In the other six patients, the presented deep learning approach gave prediction performance increases compared to the results reported in (Cook et al., 2013). Patient 9 is an example of how deep learning technology allowed reliable and tunable seizure prediction where manually chosen algorithms previously failed for the algorithm presented in (Cook et al., 2013).

After validating the developed system on a high-performance computer, we deployed the prediction system onto the ultra-low power TrueNorth platform. To make data and network compatible with this chip, they needed to be converted to lower precision and spiking representations. This resulted in an updated network topology and caused a 9% drop in IoC performance (Fig. 5). To preserve the pseudoprospective study nature of our work, no further optimization of the TrueNorth

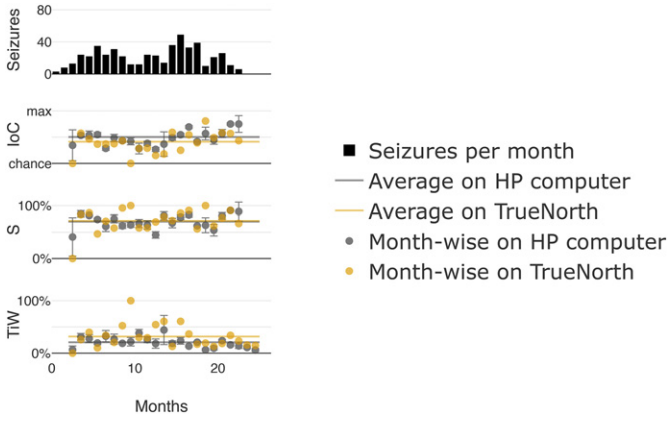


Fig. 5. Pseudoprospective long-term seizure prediction study for patient 13 on the TrueNorth chip: Traces from top to bottom represent the number of seizures per month, improvement over chance, sensitivity, and time in warning. Results on TrueNorth are represented in yellow alongside results obtained using a high-performance (HP) computer, as previously shown in Fig. 2 for the same patient for comparison, displayed here in gray.

Table 3
Performance benchmarking using TrueNorth for patient 13.

	High-performance computer	TrueNorth
Mean IoC in %	50·3	41·3
Mean sensitivity in %	70·3	71·7
Mean TiW in %	20·8	31·7

models was undertaken. A neural network architecture can be devised that is specifically optimized for implementation on the TrueNorth processor for potential use in a future clinical study. Suitably trained and optimized networks implemented in TrueNorth have been shown to deliver at or near state-of-the-art accuracy at very low power and high throughput for a variety of problems (Esser et al., 2016).

Our results show that the proposed framework can produce useful seizure prediction after as little as two months of system deployment. Results demonstrate that implementation of our algorithms on the low-power TrueNorth chip is feasible, thereby enabling a device that consumes minimal power and ensuring long battery life (Esser et al., 2016; Merolla et al., 2014). Convolutional neural networks deployed onto TrueNorth utilized approximately 50% of the chip's processing capability at a power consumption of <40 mW, which is comparable to the power consumption of a hearing aid. Any further specifications of a real device would rely on the exact purpose, such as advising patients when they are about to have a seizure or closed-loop treatment, intervention, or prevention of seizures. Moreover, developing a real device will allow for optimization of TrueNorth-specific neural networks.

Designing a seizure prediction device comes with the inherent risk of making incorrect assumptions about the performance that is best for a patient. Each user will have different needs and preferences for device performance (Freestone et al., 2017). For example, a patient may prefer different settings during night and day. Our algorithm allows for instantaneous and easy adjustment. A patient or their clinician will be able to prioritize high sensitivity or low time in warning to suit their needs and circumstances. We believe that giving a patient direct control over the sensitivity and time in warning will improve the usefulness of a seizure advisory system. The characteristics of alarm times and prediction horizons are also of great relevance to real-life system practicality, as shown in Fig. 3. A prediction horizon of, on average, one hour across all patients (see Fig. 3c) is long enough to allow for a patient to adjust their behavior and short enough to not be too disruptive over a long period of time. Note, however, that for some patients, true positive alarms lasted as long as 6 h (see Fig. 3a). One possible reason for these prolonged alarm durations is that the brain may be entering a highly excitable state (Kimiskidis et al., 2015), leading to repeated preictal predictions even long before a seizure is ultimately triggered. Note also that distributions in Fig. 3 are computed using the raw system output. Results could potentially be improved by additional processing, such as disabling the alarm for some period after the onset of a seizure or during sleep, by including patient-specific requirements, or by incorporating environment-aware sensors that may track possible patient-specific seizure triggers.

The real-time seizure prediction system presented could be applied to a closed-loop therapeutic device for titrating therapies such as neuromodulation or acute drug delivery. This has the benefit of reducing the quantity of drugs or therapeutic stimulation delivered to the patient, thus reducing the treatment burden and side effects (Sun and Morrell, 2014). For this application, alarm durations are less important. However, knowing the prediction horizon may still be crucial for optimal administration of drugs and, even though we observed a spread in prediction horizon lengths, this information may be useful to better guide dosage of anti-epileptic drugs.

Further relevant information sources can be readily incorporated into deep neural networks. For example, signals such as electrocardiogram (Fujiwara et al., 2016), weather patterns (Rakers et al., 2017), biomarkers (Nadler, 2003), or interictal spike rate (Li et al., 2013) may all

Table 4
Comparison to other studies: the improvement over chance and percentage of days used in the inference phase for each patient. We compare the performance of our system with results reported in (Cook et al., 2013) and (Karoly et al., 2017).

Patient number	This study, mean IoC in %	This study, % tested	Cook et al., mean IoC in %	Cook et al., % tested	Karoly et al., mean IoC in %	Karoly et al., % tested	Circadian only, mean IoC in % (Karoly et al., 2017)	Circadian only, % tested (Karoly et al., 2017)
1	45·0	96·1	60	6·1	34	74·0	7	74·0
2	64·1	87·6	69	10·4	–	–	–	–
3	18·3	94·6	16	34·8	26	64·2	7	64·2
8	45·2	94·6	34	20·1	48	64·2	30	64·2
9	40·5	92·4	6	29·8	34	49·4	17	49·4
10	35·6	92·0	34	31·3	35	46·5	19	46·5
11	60·2	95·8	24	4·1	43	72·3	28	72·3
13	50·3	96·0	22	11·1	48	73·2	33	73·2
14	39·7	47·2	97	12·2	–	–	–	–
15	23·9	93·6	30	3·3	19	57·1	30	57·1

be relevant to predicting seizure onset. In our work, we included temporal information in form of time of day to account for the patient-specific seizure distribution reported in (Karoly et al., 2016, 2017). Extending the current predictive system to incorporate these additional inputs and data types is the focus of ongoing work.

We have demonstrated the feasibility of implementing a real-time and ultra-low power solution on the TrueNorth chip. TrueNorth, due to its unique capabilities – namely its neuromorphic architecture, small size, and low power consumption – could provide the mobile processing power for a wearable seizure prediction or intervention system. We have proposed a deep learning approach to seizure prediction, which addresses many of the challenges identified from previous analyses using the same data (Freestone et al., 2017; Karoly et al., 2017). This study is one of the largest pseudoprospective seizure prediction studies undertaken to date. Therefore, this study may serve as a benchmark for new work exploring deep learning enabled seizure prediction. We expect advances in data processing, network design, and specialized hardware to shape the future of epilepsy research. We hope this study will motivate and guide further development of seizure prediction and intervention systems.

Acknowledgements

The authors thank Patrick Kwan, Roger Traub, Ajay Royyuru for fruitful discussions, the IBM SyNAPSE Team at IBM Research – Almaden that developed the TrueNorth chip, hardware systems, end-to-end software ecosystem, and educational materials, as well as Jianbin Tang, Antonio Jimeno Yepes and Hidemasa Muta from IBM Research – Australia for guidance on TrueNorth programming and for providing IBM Softlayer IT support.

Funding Sources

IBM employed all IBM authors of this article. The University of Melbourne employed or provided scholarships to all University of Melbourne authors of this article with funding from National Health and Medical Research Council (1065638), Australia. The corresponding authors Stefan Harrer and Dean Freestone declare that they had full access to all the data in the study and that they had final responsibility for the decision to submit for publication.

Conflicts of Interest

The authors declare they have no competing interests.

Author Contributions

IKK, SR, EN, BM, SH, PK, SB, DF, MC, DG prepared and curated data for neural network processing.

IKK, SR, EN, BM, SH, PK, DF, MC, DG designed the pseudo-prospective study.

IKK, SR, BM, TC, DP designed and implemented neural network systems.

BM programmed and ran neural network models on the neuromorphic chip.

IKK, SR, EN, PK, SS, BM, SH analyzed data.

IKK, SR, EN, PK, BM, DP, SS, SH, DF, MC, TOB, DG interpreted results.

IKK, SR, EN, PK, BM, SB, SH, DF, MC, DG wrote the paper.

IKK, SR, PK produced figures and tables.

IKK and SR contributed equally to the presented work and their names are listed alphabetically.

Appendix A. Supplementary Data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ebiom.2017.11.032.

References

2016 Community Survey [WWW Document], 2016. http://www.epilepsy.com/sites/core/files/atoms/files/community-survey-report-2016%20V2.pdf, Accessed date: 5 May 2017.

Brinkmann, B.H., Wagenaar, J., Abbot, D., Adkins, P., Bosshard, S.C., Chen, M., Tieng, Q.M., He, J., Muñoz-Almaraz, F.J., Botella-Rocamora, P., et al., 2016. Crowdsourcing reproducible seizure forecasting in human and canine epilepsy. Brain 139, 1713–1722.

Cook, M.J., O'Brien, T.J., Berkovic, S.F., Murphy, M., Morokoff, A., Fabinyi, G., D'Souza, W., Yerra, R., Archer, J., Litewka, L., 2013. 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. 12, 563–571.

Elger, C.E., Mormann, F., 2013. Seizure prediction and documentation-two important problems. Lancet Neurol. 12:531–532. https://doi.org/10.1016/S1474-4422(13)70092-9.

Esser, S.K., Merolla, P.A., Arthur, J.V., Cassidy, A.S., Appuswamy, R., Andreopoulos, A., Berg, D.J., McKinstry, J.L., Melano, T., Barch, D.R., di Nolfo, C., Datta, P., Amir, A., Taba, B., Flickner, M.D., Modha, D.S., 2016. Convolutional networks for fast, energy-efficient neuromorphic computing. Proc. Natl. Acad. Sci. 113:11441–11446. https://doi.org/10.1073/pnas.1604850113.

FACT SHEET, 2016. At Cancer Moonshot Summit, Vice President Biden Announces New Actions to Accelerate Progress Toward Ending Cancer As We Know It [WWW Document]. whitehouse.gov https://obamawhitehouse.archives.gov/the-press-office/2016/06/28/fact-sheet-cancer-moonshot-summit-vice-president-biden-announces-new (accessed 6.28.17).

Freestone, D.R., Karoly, P.J., Peterson, A.D., Kuhlmann, L., Lai, A., Goodarzy, F., Cook, M.J., 2015. Seizure prediction: science fiction or soon to become reality? Curr. Neurol. Neurosci. Rep. 15, 73.

Freestone, D.R., Karoly, P.J., Cook, M.J., 2017. A forward-looking review of seizure prediction. Curr. Opin. Neurol. 30:167–173. https://doi.org/10.1097/WCO.0000000000000429.

Fujiwara, K., Miyajima, M., Yamakawa, T., Abe, E., Suzuki, Y., Sawada, Y., Kano, M., Maehara, T., Ohta, K., Sasai-Sakuma, T., et al., 2016. Epileptic seizure prediction based on multivariate statistical process control of heart rate variability features. IEEE Trans. Biomed. Eng. 63, 1321–1332.

Furber, S., 2016. Large-scale neuromorphic computing systems. J. Neural Eng. 13:51001. https://doi.org/10.1088/1741-2560/13/5/051001.

Gulshan, V., Peng, L., Coram, M., Stumpe, M.C., Wu, D., Narayanaswamy, A., Venugopalan, S., Widner, K., Madams, T., Cuadros, J., Kim, R., Raman, R., Nelson, P.C., Mega, J.L., Webster, D.R., 2016. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. JAMA 316:2402. https://doi.org/10.1001/jama.2016.17216.

Howbert, J.J., Patterson, E.E., Stead, S.M., Brinkmann, B., Vasoli, V., Crepeau, D., Vite, C.H., Sturges, B., Ruedebusch, V., Mavoori, J., et al., 2014. Forecasting seizures in dogs with naturally occurring epilepsy. PLoS One 9, e81920.

Karoly, P.J., Freestone, D.R., Boston, R., Grayden, D.B., Himes, D., Leyde, K., Seneviratne, U., Berkovic, S., O'Brien, T., Cook, M.J., 2016. Interictal spikes and epileptic seizures: their relationship and underlying rhythmicity. Brain 139, 1066–1078.

Karoly, P.J., Ung, H., Grayden, D.B., Kuhlmann, L., Leyde, K., Cook, M.J., Freestone, D.R., 2017. The circadian profile of epilepsy improves seizure forecasting. Brain 140 (8), 2169–2182.

Kimiskidis, V.K., Koutlis, C., Tsimpiris, A., Kälviäinen, R., Ryvlin, P., Kugiumtzis, D., 2015. Transcranial magnetic stimulation combined with EEG reveals covert states of elevated excitability in the human epileptic brain. Int. J. Neural Syst. 25:1550018. https://doi.org/10.1142/S0129065715500185.

LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521:436–444. https://doi.org/10.1038/nature14539.

Li, S., Zhou, W., Yuan, Q., Liu, Y., 2013. Seizure prediction using spike rate of intracranial EEG. IEEE Trans. Neural Syst. Rehabil. Eng. 21, 880–886.

Litjens, G., Kooi, T., Bejnordi, B.E., Setio, A.A.A., Ciompi, F., Ghafoorian, M., van der Laak, J.A.W.M., van Ginneken, B., Sánchez, C.I., 2017. A Survey on Deep Learning in Medical Image Analysis. ArXiv170205747 Cs. Melbourne University AES/MathWorks/NIH Seizure Prediction | Kaggle WWW Document, 2016. https://www.kaggle.com/c/melbourne-university-seizure-prediction, Accessed date: 14 June 2017.

Merolla, P.A., Arthur, J.V., Alvarez-Icaza, R., Cassidy, A.S., Sawada, J., Akopyan, F., Jackson, B.L., Imam, N., Guo, C., Nakamura, Y., Brezzo, B., Vo, I., Esser, S.K., Appuswamy, R., Taba, B., Amir, A., Flickner, M.D., Risk, W.P., Manohar, R., Modha, D.S., 2014. A million spiking-neuron integrated circuit with a scalable communication network and interface. Science 345:668–673. https://doi.org/10.1126/science.1254642.

Mormann, F., Andrzejak, R.G., 2016. Seizure prediction: making mileage on the long and winding road. Brain 139, 1625–1627.

Mormann, F., Andrzejak, R.G., Elger, C.E., Lehnertz, K., 2007. Seizure prediction: the long and winding road. Brain 130, 314–333.

Nadler, J.V., 2003. The recurrent mossy fiber pathway of the epileptic brain. Neurochem. Res. 28:1649–1658. https://doi.org/10.1023/A:1026004904199.

Rakers, F., Walther, M., Schiffner, R., Rupprecht, S., Rasche, M., Kockler, M., Witte, O.W., Schlattmann, P., Schwab, M., 2017. Weather as a risk factor for epileptic seizures: A case-crossover study. Epilepsia 58 (7), 1287–1295.

Ratner, M., 2015. IBM's Watson Group signs up genomics partners. Nat. Biotechnol. 33: 10–11. https://doi.org/10.1038/nbt0115-10.

Seizure Gauge Challenge [WWW Document], 2017. Epilepsy Foundation. http://www.epilepsy.com/accelerating-new-therapies/epilepsy-innovation-institute/2017-challenge, Accessed date: 28 June 2017.

Sillay, K.A., Rutecki, P., Cicora, K., Worrell, G., Drazkowski, J., Shih, J.J., Sharan, A.D., Morrell, M.J., Williams, J., Wingeier, B., 2013. Long-term measurement of impedance in chronically implanted depth and subdural electrodes during responsive Neurostimulation in humans. Brain Stimulat. 6:718–726. https://doi.org/10.1016/j.brs.2013.02.001.

Snyder, D.E., Echauz, J., Grimes, D.B., Litt, B., 2008. The statistics of a practical seizure warning system. J. Neural Eng. 5, 392.

Stebbins, M., 2016. Moonshot: Accelerating the Pace of Cancer Research, Cancer Moonshot 2016 [WWW Document]. https://medium.com/cancer-moonshot/moonshot-accelerating-the-pace-of-cancer-research-be1efaa81dd2 Accessed 13 December 2017.

Sun, F.T., Morrell, M.J., 2014. Closed-loop neurostimulation: the clinical experience. Neurotherapeutics 11, 553–563.

Ung, H., Baldassano, S., Bink, H., 2017. Intracranial EEG fluctuates over months after implanting electrodes in human brain. J. Neural Eng. 14, e056011.