

SEIZURE DETECTION USING LEAST EEG CHANNELS BY DEEP CONVOLUTIONAL NEURAL NETWORK

Mustafa Talha Avcu^{12#}, Zhuo Zhang^{2*}, Derrick Wei Shih Chan³

¹Electrical & Electronics Engineering, Bilkent University, Turkey

²Neural & Biomedical Engineering, Institute for Infocomm Research, Singapore

³Dept of Paediatrics, Neurology Service, KK Women's and Children's Hospital, Singapore

ABSTRACT

This work aims to develop an end-to-end solution for seizure onset detection. We design the SeizNet, a Convolutional Neural Network for seizure detection. To compare SeizNet with traditional machine learning approach, a baseline classifier is implemented using spectrum band power features with Support Vector Machines (BPsvm). We explore the possibility to use the least number of channels for accurate seizure detection by evaluating SeizNet and BPsvm approaches using all channels and two channels settings respectively. EEG Data is acquired from 29 pediatric patients admitted to KK Woman's and Children's Hospital who were diagnosed as typical absence seizures. We conduct leave-one-out cross validation for all subjects. Using full channel data, BPsvm yields a sensitivity of 86.6% and 0.84 false alarm (per hour) while SeizNet yields overall sensitivity of 95.8 % with 0.17 false alarm. More interestingly, two channels SeizNet outperforms full channel BPsvm with a sensitivity of 93.3% and 0.58 false alarm. We further investigate interpretability of SeizNet by decoding the filters learned along convolutional layers. Seizure-like characteristics can be clearly observed in the filters from third and forth convolutional layers.

1. INTRODUCTION

Monitoring brain activity through EEG is critical for epilepsy diagnosis. To capture seizure events that may occur sparsely, neurologists have to visually scan vast amount of EEG data. The process is extremely time consuming and may be subjective due to inter observer variance. Computer aided seizure detection approach would serve as valuable clinical tool for the scrutiny of EEG data in an objective and much more efficient manner.

Traditional machine learning approaches for seizure detection usually composite three stages: data pre-processing to eliminate artifacts, followed by feature extraction and decision-making. A number of features have been identified

to describe the behavior of seizures, including those based on time-domain, frequency-domain, time-frequency analysis, wavelet features and chaotic features such as entropy etc. A pioneer work was presented in [1], which created subject-specific seizure onset detection model using hand-crafted features extracted from the raw EEG data followed by classification. The subject-specific model reaches a sensitivity of 96% and false alarm rate of 0.08 per hour on CHB-MIT dataset by using SVM over a combination of spectral, spatial and temporal features.

The extracted features are believed to contain discriminative information for systems to differentiate seizure from non-seizure states. Feature space are highly compact as compare to raw EEG data space which is critical when computing power is limited. Furthermore, the features may bring interpretability for a machine learning system. However, there are limitations for such seizure detection methods. Firstly, extracting features from raw data may induce information loss. Secondly, the standard power band analysis splitting the spectrum bands (delta 0-4Hz, theta 4-8Hz etc.) could not take into consideration individual variance of spectrum distributions. Finally, hand crafted feature extraction brings extra computational complexity for real-time applications.

Deep learning (DL) solves such problems by representation learning which enables computer to learn high level features from raw data without human interference. Recent advances in DL such as batch normalization [2], dropout [3] and various new network structures have largely prompted the applications for DL in real world problems and some have achieved near human-like performance. Various DL approaches have been proposed for seizure detection. Hugle *et al.* [4] proposed a Convolutional neural network (CNN) designed for implantable microcontroller by using only 4 electrodes selected a priori by expert. Ullah and colleagues [5] developed a pyramid CNN, and a typical 13-layer CNN model is created in [6], both using Bonn University database [7] which has only 3.27 hours of EEG data. The latter deep CNN structure has about 100k parameters, however neither dropout nor batch normalization is used to regularize the deep CNN. As a result, they could not reach the state-of-art performance

*Corresponding author. #The work was carried out under SIPGA scholarship from Agency for Science, Technology and Research (A*STAR), Singapore.

achieved by machine learning approaches using hand crafted features.

In this work, we aim to develop an end-to-end solution for seizure onset detection. A CNN structure called seizNet is carefully designed to enable an efficient and effective representative learning for seizure onset detection, equipped with dropout and batch normalization to prevent overfitting for a more generalized solution. We explore the possibility to use least number of channels for accurate detection across subjects. Finally, we attempt to interpret the model by discovering signatures hidden in the filters from different convolutional layers.

2. METHOD

2.1. Baseline method – BPsvm

We develop a SVM-based classifier using hand crafted spectrum band power features, called BPsvm. Frequency spectrum components within the 0-25 Hz band is considered, as suggested by [8]. In this study, we preserve the 5-second epoch for analysis across different approaches. To obtain features in higher resolution, we split the 5-second epoch into 5 1-second windows for spectrum transformation and band power feature extraction. In Shueb’s work [1], 8 bands from 0.5-24 Hz are chosen. As our window size is 1 second, the lowest frequency we can analyze is 1 Hz, therefore sub-bands are defined as [1-3, 3-6, 6-9, 9-12, 12-15, 15-18, 18-21, 21-24] Hz. Spectrum band power feature in the sub-band signals on every 1-second are calculated and then concatenated into one instance for every 5-second epoch. Afterwards, SVM classifier is trained based on extracted features by using radial basis function (RBF) kernel.

2.2. Deep learning method – SeizNet

We develop a deep CNN network named SeizNet for end-to-end seizure detection solution. Comparing to [6], SeizNet contains additional dropout layers [3] and batch normalization [2] after every convolution layer. Such layers are designed to avoid model overfitting. Unlike the typical usage of dropout which is only after fully connected layer, we used dropout in various parts of the model as it is indeed suggested by the inventors of dropout [3]. The number of filter at each convolution layer is multiplied by two every time like VGGNet [9]. It enables SeizNet to have less number of filters at low levels in which filters learn basic shapes, while to have more filters at the higher levels where filters are capable of grasping sophisticated patterns. As an activation function ReLU is used and other hyper-parameters of the model such as number of filters and filter sizes at each layer as well as number of unit in the fully connected layer are cross-validated over a broad range. Detailed architecture of SeizNet can be found in table 1. The total number of parameters for SeizNet-

2chn and SeizNet-18chn are 200,592 and 201,872 respectively, both include 240 non-trainable parameters.

Table 1. SeizNet Architecture

	Layer	Output
Input	$(1000 \times n^*)$	$(1 \times 1000 \times n)$
Conv 1	$8 \times \text{Conv}2D(1 \times 10)$	$(1 \times 991 \times 8)$
	MaxPool2D (1×2)	$(1 \times 495 \times 8)$
	Dropout(0.2)	$(1 \times 495 \times 8)$
Conv 2	$16 \times \text{Conv}2D(1 \times 10)$	$(1 \times 486 \times 16)$
	MaxPool2D (1×2)	$(1 \times 243 \times 16)$
	Dropout (0.2)	$(1 \times 243 \times 16)$
Conv 3	$32 \times \text{Conv}2D(1 \times 10)$	$(1 \times 224 \times 32)$
	MaxPool2D (1×2)	$(1 \times 112 \times 32)$
	Dropout (0.2)	$(1 \times 112 \times 32)$
Conv 4	$64 \times \text{Conv}2D(1 \times 10)$	$(1 \times 93 \times 64)$
	MaxPool2D (1×2)	$(1 \times 46 \times 64)$
	Dropout (0.2)	$(1 \times 46 \times 64)$
Flatten	Flatten	(2944)
	Dense (50)	50
	Dropout (0.5)	(50)
Output	Dense (2)	(2)

*n=2 for 2-channel data, n=18 for 18-channel data

We conduct leave-one-subject-out cross validation to evaluate our model. For training, we use Adam [10] optimizer with a learning rate of $4.1e - 3$, binary cross entropy loss function and batch size of 128. We explored early-stopping [11] approach by randomly selecting 20% of training data as our validation set. However, it is observed that SeizNet does not overfit due to the effect of aggressive regularization, namely dropout and batch normalization applied after every convolution layer. Therefore, 100 iterations have been decided to execute without any validation split. It also enables us to take advantage of all the data which possibly affects the performance since deep learning approaches for BCI problems are often deprived of big datasets.

2.3. Filter Decoding for SeizNet interpretation

CNNs extract spatial features hierarchically in a modular way throughout convolution layers, the filters decompose the input space to set a mapping between abstract features and labels. Decoding such filters help us discern what decomposed components are, thereby how CNNs work. One technique to fathom the characterization of hidden units is visualizing sample inputs that maximize the selected units. A pioneer method, Activation Maximization (AM) proposed by Erhan in 2009 [12] has turned this technique into an optimization problem yielding artificial inputs that maximally activate any chosen hidden unit/units by gradient ascent rather than selecting from the data set which is shown to be problematic and inadequate in the sense of leading a conclusion. Furthermore, its consistency has been shown with different initializations producing mostly same salient features at the input [12].

To date, AM has been widely used to decode abstract spatial filters and to make qualitative interpretations of CNNs. It was first applied by [13] to the AlexNet [14], known as first

modern CNN architecture as well as to the others such as VGGNet [9], GoogleNet [15] etc.

3. EXPERIMENT AND RESULT

3.1. Seizure EEG data and Pre-processing

Data used in this study is from KK Women's and Children's Hospital, Singapore. IRB was acquired from the hospital review board. EEG data of 29 pediatric patients diagnosed with typical absence seizures are included in this study. The data are extracted from Nikon Kohden EEG-1200K and EEG-9100K recording systems (reading setting: Cal Voltage=50 μ V, HFF=70Hz, LFF=0.53Hz, Sensitivity=7 – 10 μ V/mm, sampling rate=200 or 500Hz). The length of patients' EEG recordings range from 25 to 66 minutes. In total the data contains 1037.6 minutes of EEG recording with 24.95 minutes seizure data distributed among 120 seizure onsets. EEG data is down-sampled to 200 Hz across all subjects. Data from all channels for each subject is z-normalized. Window size of 5 second is chosen and preserved for all the methods in order to obtain a conclusive comparison based on performance metrics. A common problem in CNN networks for seizure detection is that datasets are often imbalanced meaning interictal phases outnumber the ictal phases by a wide margin and it has been shown that imbalanced datasets lead to statistically significant performance drop in CNN architectures [16]. To overcome this issue, a data augmentation method during pre-preprocessing is preferred rather than undersampling or oversampling. To increase the number of ictal phases, sliding is applied with different overlapping proportions according to existence or absence of seizure. While shifting with 5 seconds (no overlapping) is implemented to create interictal class, 0.075 second shifting is used for ictal class to create balanced input for the SeizNet. For BPsvm, however, no such technique is applied since SVM is shown to be robust against imbalanced datasets [17].

3.2. Experiment Settings and Performance Metrics

We compared result of four experimental settings including 18-channel SVM, 18-channel CNN, 2-channel SVM and 2-channel CNN respectively. Performance of the seizure detection algorithms are assessed with sensitivity and false alarm rate by the community [18] and often extended with latency in order to make more comprehensive analysis across detector algorithms. The definitions are described as follows:

Sensitivity(%): Proportion of seizures correctly detected

False alarm rate(fp/h): Number of false positive seizures per hour

Latency(second): Delay between electrographic onset and detection

While BPsvm yields determinate results, seizNet models produce different result in every round, due to random initialization. To evaluate the result objectively, ten tests have been carried out for both seizNet-2-chn and seizNet-18-chn models, and most frequent result which statistically corresponds

to the mode of sensitivity and false alarm is chosen as a final result for each subject.

3.3. Results

Results obtained from different experimental settings can be found in Table 2. In both BPsvm and seizNet, models using 18-chn reduce false alarms with a boost in sensitivity compare to the 2-chn models. But to our surprise, SeizNet-2-chn model despite using much less number of channels, outperforms BPsvm-18-chn model for all performance metrics measured.

Table 2. Comparison of Performance Metrics

Model	BPsvm		SeizNet	
	2-chn	18-chn	2-chn	18-chn
Seizure detected	104/120	108/120	112/120	115/120
Sensitivity (%)	86.6%	90%	93.3%	95.8%
False alarms	33	14	10	3
FAR* (fp/h)	1.91	0.81	0.58	0.17
Mean Latency(sec)	4.42	3.75	3.26	3.80

*FAR=false alarm rate

Detailed schematic illustrating number of missed seizures and false alarms for each subject can be found in Fig 1. It is observed that for subject 2 and subject 23, BPsvm-2chn could identify all the seizures while not triggering more false alarm than seizNet-2chn which could not find all the seizures for the mentioned subjects.

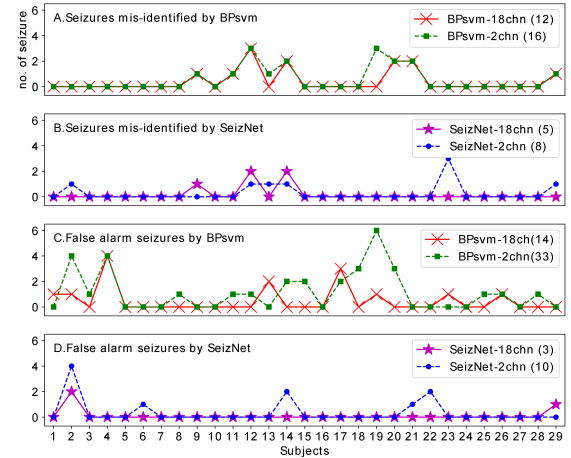


Fig. 1. Mis-identified and false alarm seizures across subjects

3.4. Filters in SeizNet

To understand what representation features are learned in seizNet, we use Activation Maximization method explained thoroughly in Section 2.3. It enables us to visualize the low level and high level features that help to reveal the feature hierarchy throughout convolution layers of SeizNet. In this work, AM is implemented with keras-vis library [19]. Unlike the baseline method in which activation maximization

loss is based on only model weights, the library provides two kind of regularization terms, namely LP norm and total variation added to the loss in order to enforce natural image prior. Default values are preserved for the weights of these regularization terms. Only input range is changed according to our pre-processing, and it is set to (-10,10). Finally, seed is initialized with random values.

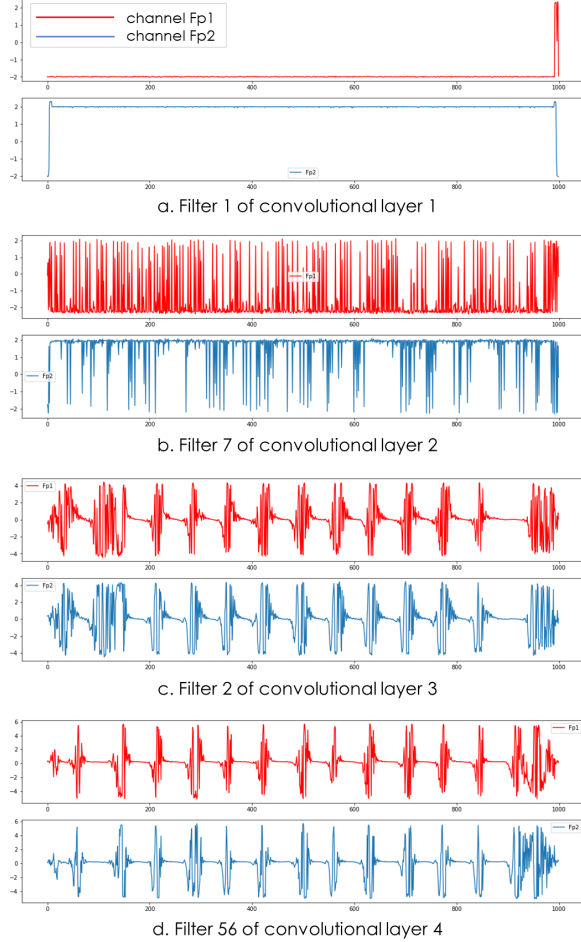


Fig. 2. Filters from 4 convolutional layers

In image classification problems, filters at the first convolutional layer usually encode the direction and color, in other words channels. Hierarchically moving along the CNN, more and more complex features are found at the higher layers that are indeed combination of features at the lower levels [12].

In SeizNet, the first layer filters present very basic shapes as shown in Fig 2.a. While interpreted as color encoding from the perspective of image analysis, for SeizNet the filters can be thought to encode EEG channel information, due to the fact that there are filters that have high and constant value for the channel 1 and low and constant value for the channel 2 and vice versa. Another possible interpretation could come from EEG montage perspective where bi-polar signal represents the difference of two electrodes, since the filters indeed subtract the channels one another.

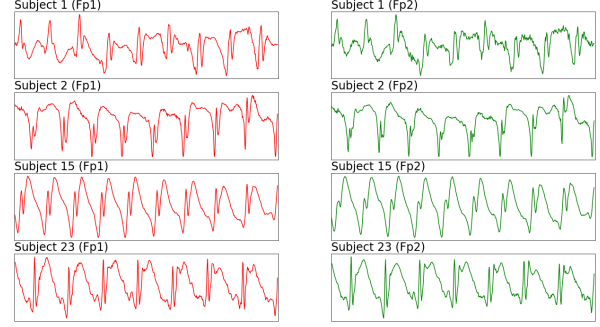


Fig. 3. Examples of seizure waveforms (3 seconds)

In the second convolution layer, various kind of filters have been found but nothing meaningful to us has been observed. However, from the third convolution layer onwards we observe substantial characteristics of seizures, therefore our hypothesis for the second layer is that it serves as the middle man to bring basic information in the lower layers into complex like-seizure signals in the higher layers

Characteristics of the seizures are observed at the third convolution layer and become clearer at the last convolution layer. It can be inferred that SeizNet has learned the fact that absence seizures create periodic and 3 Hz signals. Nonetheless, clearly SeizNet focuses on spike-and-wave happening three times in one second and try to capture it rather than capturing a whole shape of a seizure. One possible reason behind is that seizure patterns often vary from subject to subject as can be seen from the Fig 3, therefore it is reasonable for filters of a generalized model to learn the common and salient characteristics of seizures.

4. DISCUSSION AND CONCLUSION

It has been observed from the overall results that SeizNet is better than BPsvm in terms of sensitivity and false alarm. It shows that CNN models are more suitable for generalized models unlike the SVM which is indeed often implemented in subject-specific models. Nevertheless, as observed in Fig 1, frequency domain features can be more discriminative than the features extracted from time-domain for some subjects. This is also consistent with the fact that even EEG experts are to check the spectrogram in some cases in order to finalize their decision. An interesting discovery in this study is that SeizNet model trained by only 2 channels is able to outperform traditional approach trained with full scalp EEG data.

End-to-end approach is more favorable for developing real time seizure detection systems as it eliminates feature extraction, which can be a burden for real time signal processing. A solution using data merely from 2 channels makes the approach even more adoptable for light-weight, home based seizure monitoring system.

5. REFERENCES

- [1] Ali Hossam Shoeb, *Application of machine learning to epileptic seizure onset detection and treatment*, Ph.D. thesis, Massachusetts Institute of Technology, 2009.
- [2] Sergey Ioffe and Christian Szegedy, “Batch normalization: Accelerating deep network training by reducing internal covariate shift,” *arXiv preprint arXiv:1502.03167*, 2015.
- [3] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov, “Dropout: a simple way to prevent neural networks from overfitting,” *The Journal of Machine Learning Research*, vol. 15, no. 1, pp. 1929–1958, 2014.
- [4] Maria Hügler, Simon Heller, Manuel Watter, Manuel Blum, Farrokh Manzouri, Matthias Dümpele, Andreas Schulze-Bonhage, Peter Woias, and Joschka Boedecker, “Early seizure detection with an energy-efficient convolutional neural network on an implantable microcontroller,” *arXiv preprint arXiv:1806.04549*, 2018.
- [5] Ihsan Ullah, Muhammad Hussain, Hatim Aboalsamh, et al., “An automated system for epilepsy detection using eeg brain signals based on deep learning approach,” *Expert Systems with Applications*, vol. 107, pp. 61–71, 2018.
- [6] U Rajendra Acharya, Shu Lih Oh, Yuki Hagiwara, Jen Hong Tan, and Hojjat Adeli, “Deep convolutional neural network for the automated detection and diagnosis of seizure using eeg signals,” *Computers in biology and medicine*, vol. 100, pp. 270–278, 2018.
- [7] Ralph G Andrzejak, Klaus Lehnertz, Florian Mormann, Christoph Rieke, Peter David, and Christian E Elger, “Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state,” *Physical Review E*, vol. 64, no. 6, pp. 061907, 2001.
- [8] J. Gotman, JR Ives, and P Gloor, “Frequency content of eeg and emg at seizure onset: possibility of removal of emg artefact by digital filtering,” *Clinical Neurophysiology*, vol. 52, no. 6, pp. 626–639, 1981.
- [9] Karen Simonyan and Andrew Zisserman, “Very deep convolutional networks for large-scale image recognition,” *arXiv preprint arXiv:1409.1556*, 2014.
- [10] Diederik P Kingma and Jimmy Ba, “Adam: A method for stochastic optimization,” *arXiv preprint arXiv:1412.6980*, 2014.
- [11] Yuan Yao, Lorenzo Rosasco, and Andrea Caponnetto, “On early stopping in gradient descent learning,” *Constructive Approximation*, vol. 26, no. 2, pp. 289–315, 2007.
- [12] Dumitru Erhan, Yoshua Bengio, Aaron Courville, and Pascal Vincent, “Visualizing higher-layer features of a deep network,” *University of Montreal*, vol. 1341, no. 3, pp. 1, 2009.
- [13] Karen Simonyan, Andrea Vedaldi, and Andrew Zisserman, “Deep inside convolutional networks: Visualising image classification models and saliency maps,” *arXiv preprint arXiv:1312.6034*, 2013.
- [14] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton, “Imagenet classification with deep convolutional neural networks,” in *Advances in neural information processing systems*, 2012, pp. 1097–1105.
- [15] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich, “Going deeper with convolutions,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 1–9.
- [16] Mateusz Buda, Atsuto Maki, and Maciej A Mazurkowski, “A systematic study of the class imbalance problem in convolutional neural networks,” *Neural Networks*, vol. 106, pp. 249–259, 2018.
- [17] Rehan Akbani, Stephen Kwek, and Nathalie Japkowicz, “Applying support vector machines to imbalanced datasets,” in *European conference on machine learning*. Springer, 2004, pp. 39–50.
- [18] Alexandros T Tzallas, Markos G Tsipouras, Dimitrios G Tsalikakis, Evangelos C Karvounis, Loukas Astrakas, Spiros Konitsiotis, and Margaret Tzaphlidou, “Automated epileptic seizure detection methods: a review study,” in *Epilepsy-histological, electroencephalographic and psychological aspects*. InTech, 2012.
- [19] Raghavendra Kotikalapudi and contributors, “keras-vis,” <https://github.com/raghakot/keras-vis>, 2017.