Epileptic Seizure Detection using Deep Learning

Sharad Chitlangia
2017A8PS0472G
7th December 2018

Code available at https://github.com/Sharad24/Epileptic-Seizure-Detection
Github profile link - https://github.com/Sharad24/

Abstract

Epilepsy is the most common neurological disorder and an accurate forecast of seizures would help to overcome the patient's uncertainty and helplessness. Brain related disorders such as epilepsy can be diagnosed by analysing electroencephalograms (EEG). However, manual analysis of EEG data requires highly trained clinicians, and is a procedure that is known to have relatively low inter-rater agreement (IRA). Moreover, the volume of the data and the rate at which new data becomes available make manual interpretation a time-consuming, resource-hungry, and expensive process. On top of this, there is a dearth of trained and highly skilled neurophysiologists at every hospital, who can analyse EEG reports. In contrast, automated analysis of EEG data offers the potential to improve the quality of patient care by shortening the time to diagnosis and reducing manual error. Thus utilisation of technological developments like deep learning is important and can provide very good results. Deep Learning relies on data for which we use the TUH Abnormal EEG Corpus. 4 models each successively building on top of each other are described and trained on the corpus. The highest one achieves 86.4 classification accuracy. The model utilises recent advances in deep learning like 1 dimensional convolution(Speech tech), Recurrent units, Inception Modules(Image processing) and Densely connected convolutional neural networks.

Introduction

Electroencephalography (EEG) is an electrophysiological monitoring method to record electrical activity of the brain. It is typically noninvasive, with the electrodes placed along the scalp, although invasive electrodes are sometimes used such as in electrocorticography. It is frequently used for the diagnosis and management of various neurological conditions such as epilepsy, somnipathy, coma, encephalopathies, and others. Despite having lower spatial resolution than brain imaging techniques such as magnetic resonance imaging (MRI) and computed tomography (CT), EEG is a popular diagnostics tool among physicians due to its excellent temporal resolution, low cost, and noninvasive nature.

Generally, symptoms are not always guaranteed to be present in EEG data, but diagnosis of typical neurological disorders involves very long term monitoring of the patient. In this process a large amount of data is generated. This paired with the problem of there being a dearth of expert neurophysiology investigators makes way for creation of automated task based systems. It is these reasons why automatic interpretation of EEG by machine learning techniques has gained popularity in recent times.

The dataset used is known as the TUH Abnormal EEG Corpus, which is the largest of its type to date and freely available. Inspired by successes in time-domain signal classification, recurrent neural network (RNN) architectures using the raw EEG time-series signal as input were explored. This sets us apart from previous publications, in which the authors used both traditional machine learning algorithms such as k-nearest neighbour, random forests, and hidden markov models and modern deep learning techniques such as convolutional neural networks (CNN), however, did not use RNNs for this task.

It is shown that the combination of raw time series and RNNs eliminates the need to extract hand-crafted features and allows the classifier to automatically learn relevant patterns, surpassing their results. Taking inspiration from 1D convolution layers, gated recurrent units, inception modules, and densely connected networks, a deep gated RNN named ChronoNet which further increases accuracy.

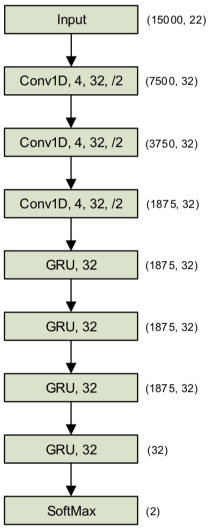
Material and Methods

All materials were acquired from the web. Since it was not possible to record seizure data on campus at the Cognitive Neuroscience Lab due to unavailability of Seizure patients, the TUH EEG corpus dataset was used.

All of the models described below are at: https://github.com/Sharad24/Epileptic-Seizure-Detection in the file ChronoNet.ipynb. All the models were trained on the Google Cloud Platform.

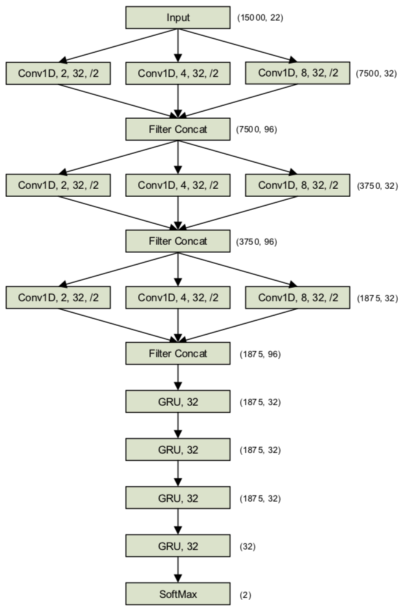
The basics of the deep learning models and submodules used is explained in literature survey.

Convolutional Gated Recurrent Neural Network(CRNN)



Since our data is of the time series type, having recurrent units is a must. Taking inspiration from speech technology based developments in deep learning, applying a 1 dimensional convolutional neural network before feeding into the recurrent neural network would be better. This helps in the sense that the 1 dimensional convolution acts like a feature extractor for the recurrent neural network. The other advantages are that the size of time series is reduced. This reduces the computation time taken by Gated Recurrent Unit as it is the most computationally expensive unit in the neural network. An accuracy of 76% is achieved through this.

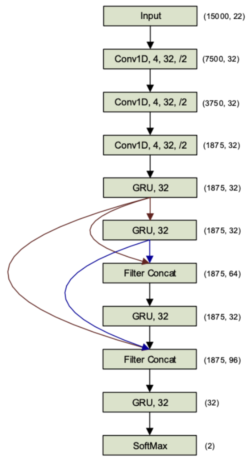
Inception Convolution Gated Recurrent Unit(ICRNN)



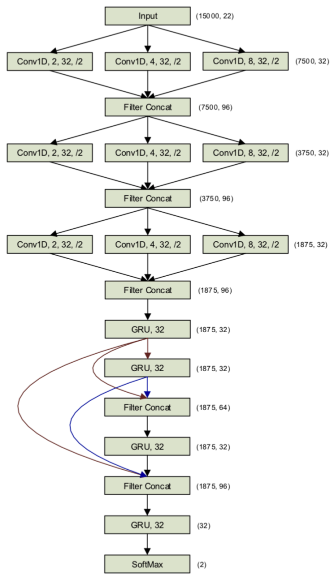
In the previous C-RNN architecture, each Conv1D layer had the capability to extract local information at only one time scale determined by a single fixed filter size, limiting the flexibility of the model. Since the rate of change of information in a time series depends on the task at hand, the filter size for each Conv1D layer would have to be hand-picked to fit the particular data. Inception Modules can be applied here since they are able to capture representations at varying levels.

This helps in the sense that EEG data that we have might not only correspond to high correlational data with 1 time sized filter. Having multiple filters of different sizes help us capture all the information at different sizes. An accuracy of 78% was achieved with this.

Convolutional Densely Connected Gated Recurrent Neural Network(CDRNN)



Since the neural network has been growing in size, it is important that we employ some techniques for ease in gradient flow i.e., dense connections. Dense connections are employed in the recurrent units. This should make sense as our recurrent units are computing outputs over a series length of 1875(max). It is clear that without dense connections we encounter the problem of gradient flow between layers. The neural network is now immune to the problem of degradation of learning.
This model achieves an accuracy of 83.4%. It is clear that dense connections are massively helpful visible from the massive increase in accuracy.



Inception Convolutional Densely Connected Gated Recurrent Neural Network(ICDRNN)(ChronoNet)

Combining ideas from inception, recurrent units and dense connections is definitely the next step. So replacing single 1 dimensional filters in convolutional densely connected recurrent neural network with varying sizes is employed. Here inception modules help capture features in EEG signal of varying sizes. This can be helpful since in epilepsy brain waves increase in amplitude as time progresses(during a seizure). After inception modules recurrent units are employed. The recurrent units have residual connections in them so that there is no problem faced in gradient flow. The recurrent units are very helpful since seizures do not happen at 1 instant. The EEG data shows that the seizures last over a period of time. This is where recurrent units help. This model achieves an accuracy of 86.2%.

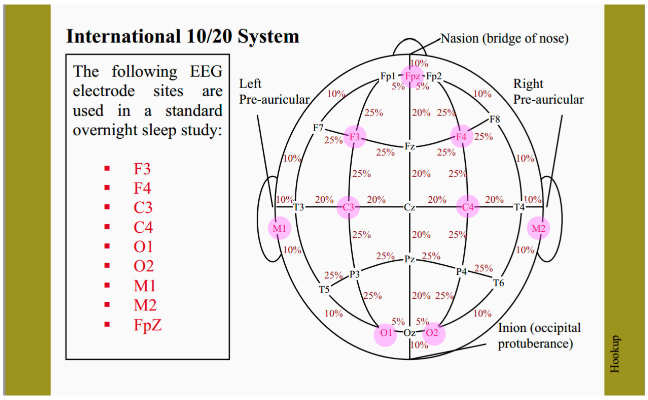
Results and Discussion

Data

The TUH - EEG Corpus was used. It has sessions recorded for epilepsy diagnosed patients since 2002. The total amount of sessions is 16000. The total sessions having seizures in them is 852. The fraction of seizure sessions is 0.05. Compared to other corpuses this is a huge number since recording EEG during seizures is in itself a very big task. The total amount of signal data is about 850 hours.

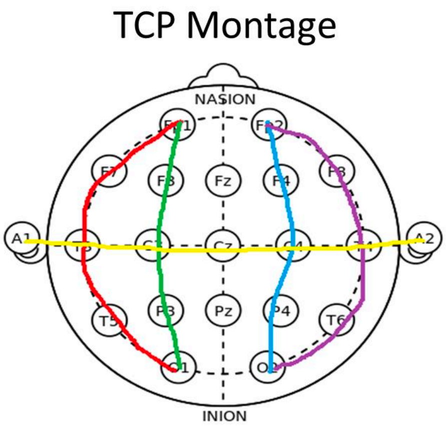
Data Preparation

EEG recordings in the TUH EEG Corpus was done according to the internationally acclaimed 10/20 electrode placement system. The "10" and "20" refer to the fact that the actual distances between adjacent electrodes are either 10% or 20% of the total front–back or right–left distance of the skull. For example, a measurement is taken across the top of the head, from the nasion to inion. Most other common measurements ('landmarking methods') start at one ear and end at the other, normally over the top of the head.



The montage system utilised was the TCP montage(Transverse Central Parietal). The image on the right shows how data is created for this particular montage. In the dataset, the electrode values with respect to a reference electrode are given.

So, Fp1-F7=(Fp1-ref)-(F7-ref).



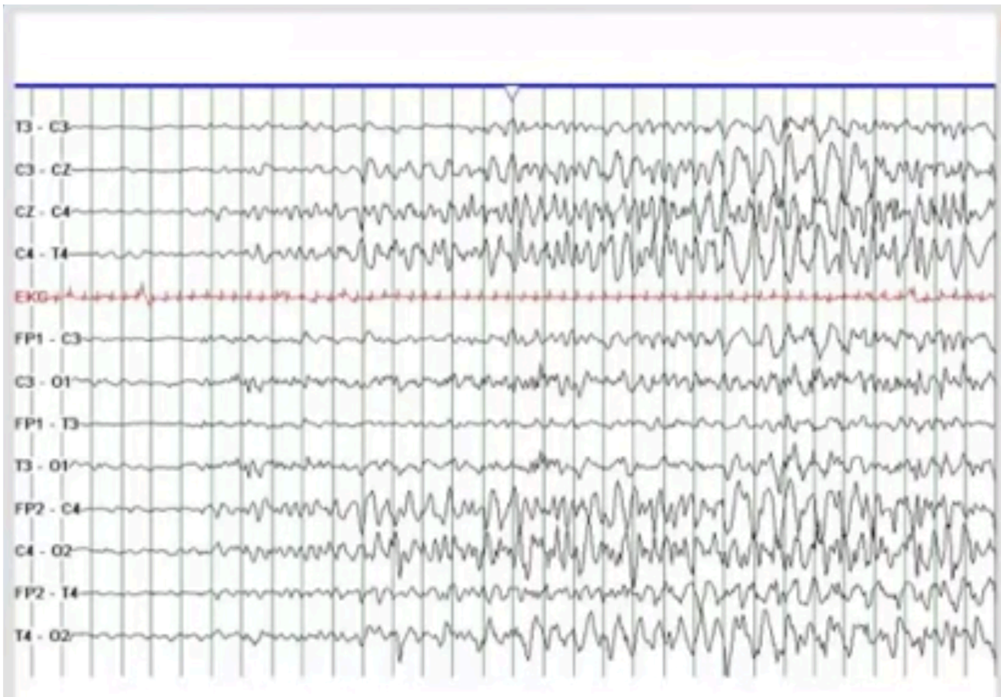
In a comparative original study done on the dataset, it is shown that the most amount of seizures were recorded in the first 60 seconds of each file. Normal files have recordings for around 1162 seconds samples at 250 Hz. Hence input to the neural network is only 250x15=15,000 time values. This massively reduces training time for neural network since it increases the amount of useful data.

Results

We used the dataset described above to train the four deep recurrent neural network architectures presented just above. Networks were trained using the adaptive moment estimation optimization algorithm with a learning rate of 0.001. We used a batch size of 64 and trained the networks for 500 epochs. Results are shown in the table given below:

Model	Training Accuracy	Testing Accuracy
CRNN	76.1%	75.6%
ICRNN	78.7%	76.2%
CDRNN	83.8%	82.1%
ICDRNN(ChronoNet)	86.4%	85.3%

The amount of data held out for testing was 25% of the whole dataset. It is seen that hypotheses while developing the models are all correct. Dense connections turn out to be an important factor in increasing the accuracy by a massive 6%. This primarily happens because of the training data time sequence being so long(15000 values). Even inception modules increase the accuracy by 2%. This is obvious since it is very clear that brain waves in seizures in epilepsy increase with time as shown below. Here we can see that initially the patient in theta wave with EEG amplitude very small. Then the amplitude increases by a factor of 10 to 15 and frequency goes down 1 -3 Hz which is delta waves. This plot is from the actual data and represents the first 30 seconds of eeg file. Each long line across the plot indicates 1 second has passed from the previous long line.



Literature Survey

Epilepsy is commonly known as the seizure disorder. If a patient has 2 or more seizures without any other medical illness or reason being a cause of it, the patient is diagnosed with Epilepsy. It is a common misconception to think that if someone has seizures they have Epilepsy. The other reasons can be as simple as hyperglycemia. Patients with Epilepsy are normally characterised by having abnormal EEG apart from getting seizures on a regular basis. The regularity varies from person to person. It can be anywhere from a few lasting for a couple of hours to a double digit number lasting for only a few minutes each day. 50 million people worldwide suffer from epilepsy and for approximately 30 % of them the disease cannot be sufficiently controlled by medication. Only 50 % of the patients who undergo respective surgery keep seizure free. For all remaining patients the uncertainty and unpredictability of seizures belongs to the most severe disabilities. Although seizures cannot be completely prevented, a reliable forecast of their occurrence would help to overcome the helplessness of affected patients and would significantly improve their quality of life

Seizures

A seizure is a sudden, uncontrolled electrical disturbance in the brain. Based on their intensiveness there are two types of seizures, convulsive and non-convulsive. By having abnormal EEG, we mean that the patient is in a state called as non-convulsive status epilepticus. This is a very mild type of seizure where the patient becomes drowsy. A family member or a close friend is only able to tell this on looking at the patient with attention. Seizures affecting only a part of the brain/ hemisphere are called as Focal/Partial Seizures. They either affect the whole hemisphere or only a particular lobe of a particular hemisphere of the brain. The other category is called as generalised seizures. These king of seizures affect both the regions of the hemisphere of the brain. Tonic Seizures are characterised by the fact that the person experiencing them has their muscles stiffened. The person would normally fall in such a seizure. Atonic Seizures are seizures in which the muscles become relaxed. Clonic seizures are the ones causing convulsions. Patients are also seen sometimes to get Tonic followed by Clonic seizures. Normally if a seizure lasts for more than 5 minutes it is termed as a medical emergency. Patients are sometimes treated with Benzodiazepines. These are drugs which enhance the production of GABA. GABA receptors capture GABA which is an inhibitory neurotransmitter. Enhancing GABA means that there are more inhibitory transmitters in the brain. So neural activity is reduced and hence the seizure ends. This is a common problem in patients as this is genetic. Other medication include anti-convulsants during Clonic Seizures and prescription of Ketogenic diet.

Demographics

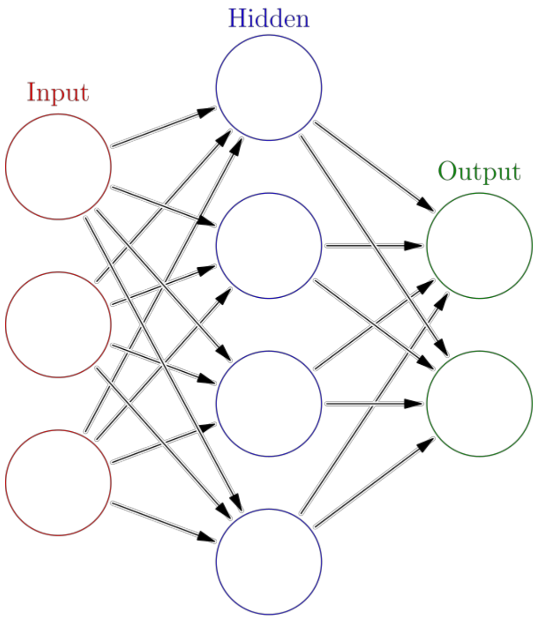
According to a report by Apollo hospitals, around 1 million cases in India are reported every year [10]. 50 million people suffer from it worldwide[11].

Machine Learning

Machine learning is the science of getting computers to act without being explicitly programmed. Machine learning algorithms build a mathematical model of sample data, known as training data, in order to make predictions or decisions without being explicitly programmed to perform the task.

Artificial Neural Networks

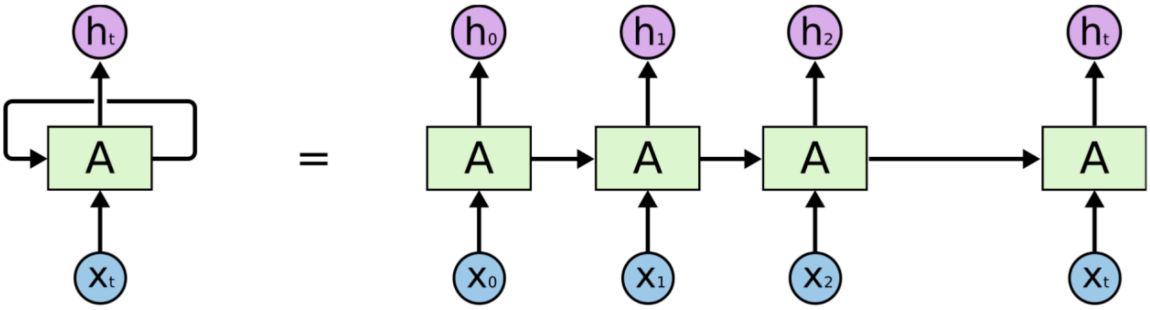
Artificial Neural Networks are computing systems vaguely inspired by the biological neural networks that constitute animal brains. They work as powerful function approximators and their depth, width (i.e., the architecture of the neural network can be varied on the basis of the task in hand).



The steps involved in training a typical neural network model would be defining the architecture based on the task, preprocessing the data before sending the data into the neural network, forward propagation(computing output value using the neural network and the preprocessed data), computing a loss based on the 'label' value we already have from our dataset, calculating gradients for each weight in the neural network and then doing it again. This whole process is repeated until the neural network optimises itself such that for a given input it outputs the correct value.

Recurrent Neural Network

These are a type of neural network which are optimum when having time series data. There are recurrent connections in between nodes of a neural network, similar to a feedback loop in sequential circuits in digital electronics.
Every hidden state in a recurrent neural network depends on its previous state. After computing all the hidden states it is multiplied by a weight matrix and then a non linear activation function is applied over it. This is our output representation from one layer. Multiple layer like these stacked together depending on what kind of architecture is desirable.



One problem that this basic recurrent neural network faces is the vanishing gradient problem. When computing the gradients for the weights in a layer, it is obvious that the gradient of a particular node will depend on the node just before it(when looking from end to start). While computing the gradient of a node in a neural network we see that the gradient of the activation is taken before as well which over here is the softmax function. The softmax function is a function from the exponential family. So the gradient update gets stuck after a point because for even a very high change in the input the weight update is not big because of infinitesimally small change in the value of the exponential function.

Gated Recurrent Unit

These were introduced in the first place to solve the problem of the vanishing gradients we just encountered. A gated recurrent unit (GRU) produces the current value of hidden state ht by performing a linear interpolation between an intermediate candidate hidden state h˜ t derived from Equation (2) and the previous value of hidden state ht−1. A GRU employs two gates: an update gate zt controlling the extent to which the previous state will be overwritten and a reset gate rt deciding how much of the previous state should be forgotten while computing the candidate hidden state. More formally, the GRU model can be presented in the following mathematical form:

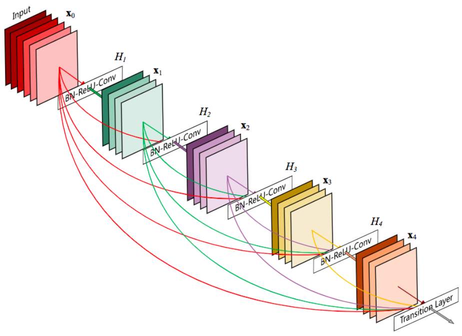
$$
\begin{aligned}
h_t &= (1 - z_t) \odot h_{t-1} + z_t \odot \tilde{h}_t \\
\tilde{h}_t &= g(W_h x_t + U_h(r_t \odot h_{t-1}) + b_h) \\
z_t &= \sigma(W_z x_t + U_z h_{t-1} + b_z) \\
r_t &= \sigma(W_r x_t + U_r h_{t-1} + b_r),
\end{aligned}
$$

Inception Module

This is submodule of a very bigger system of architectures known as Convolutional Neural Network. A typical convolution neural network can be described by a filter matrix running across an image to calculate the final output. The operation is a simple dot product between the area of the image encompassed by the filter and the filter. It is very clear that the resolution of the image reduces here. These filter matrices help capture spatial features. For example in a edge detecting convolutional neural network, filter edges in its matrix will compute high dot products where ever it sees an edge in the image over which the filter is convolved.
The idea behind the introduction of inception modules was to capture varying levels of spatial resolution at the same time. Since our filter sizes once specified cannot be changed, having filters of varying sizes helps. This is because it might be possible that in the input image the edges are not only of a small particular size, there are edges of large sizes as well. In such a problem having filters of different sizes help since both filters work together to compute the final output. Typical sizes from an inception module are 1x1, 3x3, 5x5, where after they have computed the output they are concatenated together.

DenseNet

When making huge architectures, final representations are computed by a lot of transformation over the input. When learning for such a neural network happens, the gradients have to pass through a lot of layers. Here, the gradients get stuck at a particular value and thence no learning actually happens. For this reason Residual functions have excelled in learning over convolution neural networks.



The gradients, as clearly visible can ‘fly’, through these residual type connections and hence it is possible to get away with the problem of gradient flow.

Future Scope

- Training of models on other datasets.
- Applications of Wavelet transforms for Preprocessing of data. This is because seizure data is highly correlational to wavelets of their same size and dimension. This is described in these papers -
    - [16] Epileptic seizure detection from EEG signal using Discrete Wavelet

  Transforms.
    - [17] Analysis of Epileptic Seizures in Wavelet Domain - Semantic Scholar
    - [18] Wavelet-based EEG processing for computer-aided seizure detection
- Wavelet Analysis: Analysis of data using wavelet transforms. Developing of models and methods with/without machine learning for detection on EEG data.
- Fourier transforms: Using Fourier transforms to transform input signal into feature rich representations for neural network. This is inspired from the commonly used mel-spectogram based transformation used for speech technology.

References

1 - TUH Abnormal EEG Corpus - https://www.isip.piconepress.com/projects/tuh_eeg/
2 - https://arxiv.org/pdf/1612.08083
3 - https://arxiv.org/abs/1506.00019
4 - https://www.cs.unc.edu/~wliu/papers/GoogLeNet.pdf
5 - https://arxiv.org/abs/1608.06993
6 - https://arxiv.org/pdf/1802.00308.pdf
7 - https://www.youtube.com/watch?v=uwB8AZ2vK1g
8 - https://www.wikipedia.org/
9 - Gazzaniga, M.S., R.B. Ivry & G.R. Mangun. Cognitive neuroscience: the biology of the mind (4th ed., New York: Norton, 2008). ISBN: 978-0-393-11136-1.
10 - https://www.google.com/
11 - https://arxiv.org/pdf/1811.00915.pdf
12 - P. Kwan and M. J. Brodie, “Early identification of refractory epilepsy.” The New England journal of medicine, vol. 342, pp. 314–319, Feb. 2000.
13 - J. de Tisi, G. S. Bell et al., “The long-term outcome of adult epilepsy surgery, patterns of seizure remission, and relapse: a cohort study.” Lancet (London, England), vol. 378, pp. 1388–1395, Oct. 2011.
14 - F. Mormann, R. G. Andrzejak et al., “Seizure prediction: the long and winding road,” Brain, vol. 130, no. 2, pp. 314–333, 2006.
15 - A. Schulze-Bonhage and A. Kühn, “Unpredictability of seizures and the burden of epilepsy,” Seizure prediction in epilepsy: from basic mechanisms to clinical applications, 2008.
16 - https://ieeexplore.ieee.org/document/6883868/
17 - https://pdfs.semanticscholar.org/d121/17194617408a15dd7baed9e2a46a364adfc9.pdf
18 - https://www.sciencedirect.com/science/article/pii/S1059131115000138