# Seizure Detection through Time Series Modeling

**Yakun Wang**
Center for Data Science
New York University
New York, NY 10011
yw3918@nyu.edu

**Liangzhi Li**
Center for Data Science
New York University
New York, NY 10011
ll3399@nyu.edu

**Shizhan Gong**
Center for Data Science
New York University
New York, NY 10011
sg5722@nyu.edu

## Abstract

Epilepsy is one of the most common neurological disorder, characterized by recurrent seizure activities. The diagnosis of epilepsy is usually made by a trained neurologist but can be difficult to be made in the early stages. Studies have shown that it is possible to build an automated seizure detection system using EEG data. In this project, we implement three different models: AR, HMM and RNN, to address this problem. We use the publicly available EEG data from Bonn University to build the models. We preprocess the data via Gaussian Process Probabilistic Amplitude Demodulation, train models in three approaches, and test their performance. Our result shows that RNN model performed best and achieves 0.98 accuracy on the test set.

## 1 Introduction

Epilepsy is a chronic neurological disorder that may cause brief electrical disturbances in the brain. It is characterized by the occurrence of recurrent seizures in the brain. Approximately 1% of the world population suffer from epilepsy. During a seizure, the patient may experience abnormal behaviors, symptoms, and sensations, even loss of consciousness in some extreme cases. The chronic unprovoked seizures will cause a broad range of debilitating medical and social consequences for epileptic patients, and will significantly affect their normal lives. Although there are several treatments that can help, this condition can't be completely cured. Thus, early detection of seizure is very important to clinical care.

Brain activity during a seizure differs significantly from the normal state, regarding frequency and pattern of neural firing [1]. This makes it possible to build an automated seizure detection system by detecting specific patterns of brain activity so that medication control can be taken at earlier stage, and so that the normal lives of epileptic patients will be less affected. Since epilepsy is a condition related to the electrical activity of the brain, electroencephalogram (EEG) signal can be used to detect seizures. Traditionally, the gold standard for diagnosis of epilepsy is continuous EEG monitoring along with video monitoring of the patient, which usually require in-patient admission. EEG signals are analyzed by trained neurophysiologist. The process is not only slow, but more importantly, not always accurate, making diagnosis in the early stages of the disease a challenging task. Additionally, occurrence of seizures are usually unpredictable. Hence, automated seizure detection systems are recommended to screen for seizures during long-term EEG recordings.

32nd Conference on Neural Information Processing Systems (NeurIPS 2018), Montréal, Canada.

The objective of this project is to build a model to detect seizure using EEG data.Our project uses three different approaches to address this problem: ARMA, HMM and RNN models. We first preprocessed the given EEG data via Gaussian Process Probabilistic Amplitude Demodulation (GP-PAD) , and then fit EEG data into the three models. We found that RNN model achieved the highest accuracy.

## 2 Related Work

Previous work in algorithms for automated seizure detection systems dates back to early 1970s. Over the last few decades, multiple algorithms have been developed to address this problem. Some research focuses on adult epileptic patients, and others focuses on neonatal seizures. In any case, the goal of using EEG data to detect seizures is similar. Most work include three phases: data preprocessing, feature extraction and classification. Raw EEG data are usually noisy, so some techniques of signal processing should be applied before feeding them into algorithms. Then, majority of algorithms work by analyzing EEG signals to extract relevant information to classify an episode of epileptic seizure from background EEG activity. Frequency analysis methods include Fourier transform and wavelet transform, which transform the temporal data into other domains. Finally, a classifier will be implemented to classify whether the input EEG sequence is seizure or not, using the extracted features.

There are a wide variety of EEG patterns that characterize a seizure. For example, Srinivasan et al. [2] proposed a neural-network-based automated epileptic seizure detection system that uses approximate entropy (ApEn) as input feature. Adeli et al. [3] presented a wavelet-chaos methodology for analysis of EEG for detection of seizures and epilepsy. In this study, the nonlinear dynamics of the original EEG data are quantified in the form of the correlation dimension (CD, representing system complexity) and the largest Lyapunov exponent (LLE, representing system chaoticity). Faul et al. [4] used the properties of Gaussian Process (GP) probabilistic models to develop two methods, highlighting the presence of seizures from EEG signals. They believe that GP models have attractive advantages over parametric and neural network modeling due to small number of tunable parameters, ability of being trained on relatively small training set, and provision of a measure of prediction certainty. Mousavi et al. [1] presented a method for epilepsy detection based on autoregressive (AR) estimation of EEG signals. They determined the optimal order of AR model by BIC and extract AR parameters of EEG signals and their sub-bands (created with the help of wavelet decomposition) based on the AR model. Finally they used multilayer perceptron (MLP) classifier to do the classification. We keep their main framework in one of our approaches. However, they only test a few possible orders for each series and they force the series of the same class to be generated from models with the same order, while we explore all the possible orders below a maximum order to fit the best model for each individual series. In another paper, Abdullah et al. [5] deployed Stationary Wavelet Transform (SWT) to extract features from EEG signals and then do the classification through Hidden Markov Model (HMM). In our work, we also try to classify with HMM. But we use a different data preprocessing approach.

In recent years, there has been a great success of using deep learning to solve challenging machine learning problems, including seizure detection task. In particular, since the innate structure of Recurrent Neural Network (RNN) makes it work well in sequence learning task, recent work on seizure detection tends to use either GRU or LSTM to address the problem. Sachin Talathi [6] proposed a deep learning framework via the use of RNNs with Gated Recurrent Unit (GRU) hidden units to classify single-channel EEG time series data. We also test the power of GRU in terms of seizure detection. Moreover, in our work, we incorporate an ensemble mechanism to further improve the model performance.

## 3 Problem Definition and algorithms

### 3.1 Task

Our major task is to distinguish among EEG data of healthy people, that of interictal people and that of ictal people. Therefore, our problem is defined to be a supervised, 3-class classification problem. The data is sampled as time series, so we are expected to capture the differences among groups through methods suitable for time-related characteristics.

### 3.2 Algorithm

#### 3.2.1 Gaussian Process Probabilistic Amplitude Demodulation (GP-PAD)

Due to the highly ill-posed nature of EEG data, we perform Gaussian Process Probabilistic Amplitude Demodulation [7] on the data before we build our classification models. This algorithm decomposes the data into the multiplication of two components: A slowing varying and positive envelope component ($m_t$) and a quickly varying carrier component ($c_t$). See Eq.(1).

$$y_t = m_t c_t \tag{1}$$

This demodulation process is based on a forward model specifying the parametrized joint probability of the signal, carrier and modulator. The model consists of the likelihood $p(y_{1:T}|c_{1:T}, m_{1:T}, \theta)$, the prior distribution over the carrier $p(c_{1:T}|\theta)$ and that over the modulator (envelope) $p(m_{1:T}|\theta)$, see Eq.(2).

$$p(y_{1:T}, c_{1:T}, m_{1:T}|\theta) = p(y_{1:T}|c_{1:T}, m_{1:T}, \theta)p(c_{1:T}|\theta)p(m_{1:T}|\theta) \tag{2}$$

The likelihood is defined to be a Gaussian-distributed random variable with a mean equals to the product of the envelope and carrier, and a time-dependent variance denoted by $\sigma_{y,t}^2$ in Eq.(3). As shown in Eq.(4), the prior distribution for the carrier is simply assumed to be the same as Gaussian white noise, which has no correlation with time.

$$p(y_{1:T}|c_{1:T}, m_{1:T}, \theta) = \mathcal{N}(y_t; m_t c_t, \sigma_{y,t}^2) \tag{3}$$

$$p(c_{1:T}|\theta) = \mathcal{N}(c_t; 0, 1) \tag{4}$$

The prior distribution for the modulator has to be assumed carefully, as it is restricted to be positive and could not be considered as Gaussian. In this model, it is generated by the application of a point-wise non-linear function to the transformed-modulator ($x_t$), where $x_t$ is assumed to be drawn from a stationary Gaussian process (Eq.(5)). $m_t$ is the result of $x_t$ through a 'soft threshold linear' function, see Eq.(6).

$$p(x_{1:T}|\mu, \Gamma) = \mathcal{N}(x_{1:T}; \mu_{1:T}, \Gamma_{1:T,1:T}) \tag{5}$$

$$m_t = m(x_t) = \sigma_m \log(1 + \exp(x_t)) \tag{6}$$

Later on, gradient-based method is chose for the inference approximation as full distributional inference here is intractable. When learning for the model, Bayesian methods are applied to automatically learn parameters from data.

#### 3.2.2 Autoregressive (AR) Model

It is a nature thinking to use an autoregressive model for time series process. An AR model of order $p$ indicates that a current sample $x_t$ is determined by a linear combination of sample from previous $p$ time steps plus a Gaussian white noise $w_t$:

$$x_t = \Sigma_{i=1}^{p} \phi_i x_{t-i} + w_t \tag{7}$$

Here, we use AR model to approximate the EEG data and assume that the AR order as well as the corresponding parameter sets $\phi$ could reveal the underlying structures of the time series and therefore could be then passed through a general classification model, like the multi-layer perceptron we used here, to help distinguish the status of patients.

#### 3.2.3 Hidden Markov Model (HMM)

HMM is a Markov process with hidden states. We build HMM for each kind of seizure state and assume that each model have several hidden states. In each model, the hidden state $z_t$ might change or stay in the current one with probability controlled by a transition matrix at every time step. The observation $x_t$, which represents the current EEG point, is sampled from the corresponding observation model of current hidden state. Figure 1 illustrates a general Hidden Markov process.

#### 3.2.4 Recurrent Neural Network (RNN)

Recurrent Neural Network is especially suitable for time series data as its connections between nodes form a directed graph along a sequence. Here we use a variation of RNN, GRU, which adds shortcuts to RNN through update gates and reset gates in order to achieve better performance. An illustration for GRU is shown in Figure 2. At each time step, we feed current point into the network, and the hidden state is updated accordingly. The last hidden state is considered to have obtained information from the whole time series and therefore be decoded to get make classification.

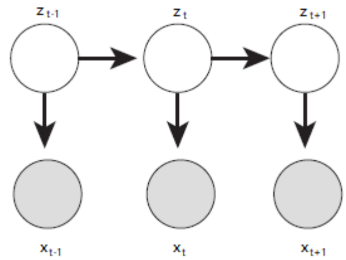



Figure 1: Illustration for HMM

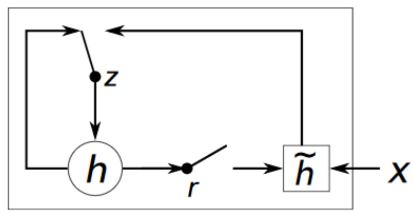



Figure 2: Illustration for GRU, adapted from [8]

# 4 Experimental evaluation

## 4.1 Data

The EEG dataset is from the publicly available database on the website of Bonn University [9]. EEG signals from three different groups are analyzed: group Healthy (Set A & B), group Interictal (Set C & D), and group Ictal (set E).

The EEG segments in set A and B are from surface EEG recordings of five healthy volunteers with eyes open and closed respectively. The EEG segments in set C an D are from EEG recordings of five epileptic patients, during seizure free intervals and from the hippocampal formation of the opposite brain hemisphere. Finally, EEG segments in E contain seizure activity.

Each set contains 100 single channel EEG segments of 23.6 second duration, and each segment is sampled at 173.61 Hz (4097 data points). Therefore, the total number of data points per set is 409700.

## 4.2 Methodology

We split the data into training and test data. For each class, we randomly select 30% of instances to be included in the test set to evaluate our model, and use the rest of the data for training. We conducted stratified 5-fold cross validation for hyperparameter tuning and model selection. Since our labels are relatively balanced, we use accuracy in the validation set to evaluate our models.

### 4.2.1 Preprocessing

To begin with, we normalize all the EEG series. In order to get rid of the power supply noise, we first apply a 50 Hz notch filter and then a 40 Hz low-pass filter to each EEG series for further processing. Furthermore, since most algorithms requires stationary series while most EEG data are ill-posed, we also performed GP-PAD to each series. Only carrier decomposed from each series are used for modeling.‘

### 4.2.2 AR

We first try to extract features through AR modeling. To be more specific, we fit an AR model to each EEG series. In order to find the optimal order of AR model, we set a maximum order (40) and

Table 1: Model comparison

| Model | Validation Accuracy |
|---|---|
| AR model | 0.8750 |
| HMM model | 0.7325 |
| RNN model | 0.9375 |

build models of all the orders below this maximum order for each individual series. Then the best fitted model is selected through BIC criteria, which can often achieve sparser model than AIC criteria. The coefficients of the auto-regressive terms are used as features, where we insert 0 as coefficients of orders between the optimal order and the maximum order. The features are then fed into a multi-layer perceptron to do classification.

#### 4.2.3 HMM

For HMM modeling, we first separate training data set into data sets for each class. In each class, we treat individual EEG series as independent. Using training data, we train one HMM per class. The number of hidden state we used for each model is 4. Then for each series in the test set, we calculate log likelihood of the series under each model. The labels of the series are set to be the label of the most likely models.

#### 4.2.4 RNN

Since the learning of RNN requires enormous amount of data, and too long series will raise some computational issue. We cut each long series into 50 short ones. Then we feed each short series into a single layer GRU with hidden size of 32. The last hidden state of the GRU is then fed into a fully connected layer with softmax output. The model is trained with Adam optimizer with learning rate of 0.001. The gradient is computed using a minibatch of 64 EEG series. After we train the model, we can give a predicted label to each short series. The labels of the original series are determined through majority vote.

### 4.3 Results

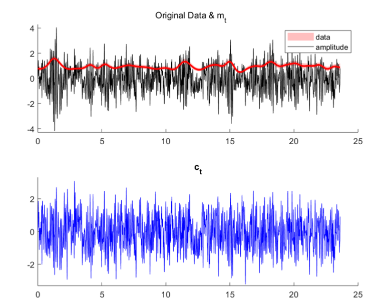
For each model in question, we obtain its accuracy on the crossed validation sets and calculate the mean of 5 accuracy. The results are shown in Table 1. From the results, we can see that RNN performs the best while HMM has relatively poor performance. This may due to the reason that RNN has very complicated structure, which render the algorithm strong flexibility in dealing with complex data such as EEG signals. Though generally speaking HMM is more complicated than AR model, in our setting we make an assumption that all training samples are generated by a single HMM model, while actually they may come from much noisier processes. This leads to the lack of fit of our model, making the HMM have the worst result.

We further evaluate our best algorithm on the test set. It turns out our RNN algorithm can achieve an accuracy of 98% in the test set. Specifically, our model can correctly classify all the healthy and ictal cases, and 92.5% of the interictal cases. This result is reasonable since interictal data usually has similar behaviors with health or ictal data according to the severity of patient, making it hard to classify.

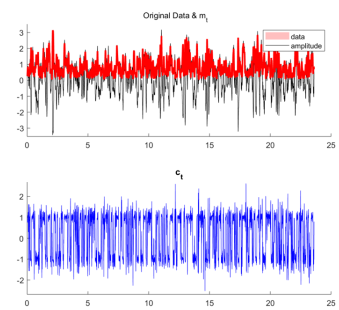
### 4.4 Discussion

We mainly make two hypotheses during our modeling. First, we assume after data preprocessing, our EEG data is stationary and suitable for modeling with the methods in question. Second, we assume EEG data of different types have different patterns which can be regarded as coming from disparate generative process. Here we validate these two hypotheses through visualization.

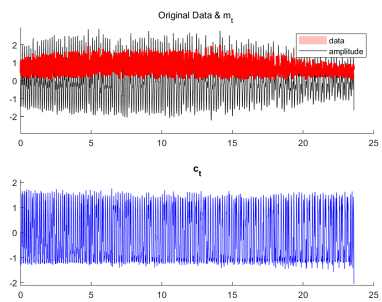
We first draw time plots of several sample EEG series drawn from different classes before and after data processing (Fig. 3). From the original plots, we can already tell that there is significant difference among EEG data of disparate types. Comparing with healthy EEG data, ictal EEG data changes more frequently and drastically while interictal data has patterns in between. We also draw several




(a) Healthy

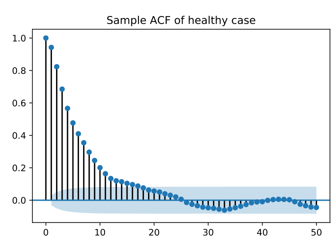



(b) Interictal

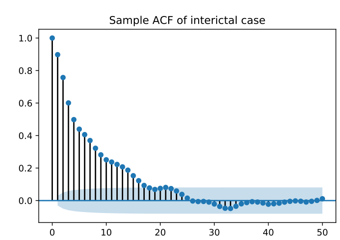



(c) Ictal

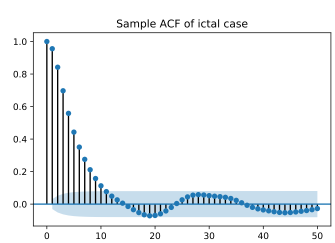
Figure 3: Samle Time Plot




(a) Sample ACF of healthy series



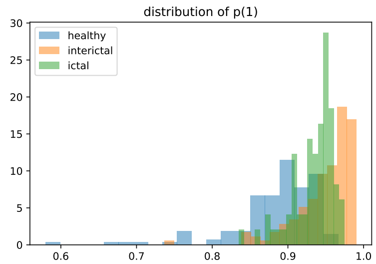

(b) Sample ACF of interictal series



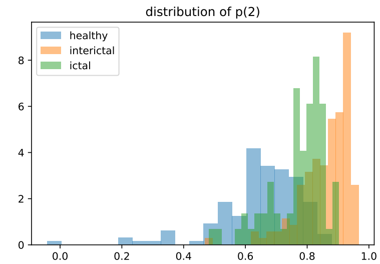

(c) Sample ACF of ictal series

Figure 4: Samle ACF

ACF plots of instances sampled from different classes (Figure 4). We can see that the ACF decay exponentially so that after data preprocessing, the data seem to be stationary. Actually, most of them pass Augmented Dickey Fuller (ADF) test. For the healthy data, the envelope changes really slowly, and its value is close to 1, which means the the carrier is very similar to the original data. On the contrary, the envelope of the ictal data varys very fast, and the carrier is very different from the




(a) Distribution of lag 1 ACF




(b) Distribution of lag 2 ACF

Figure 5: ACF distribution of different classes

original data, which is more orderly. Therefore, our data preprocessing can transform the original data into stationary time series while keeping the different patterns of each classes.

We also draw the distribution of lag 1 and lag 2 ACF of different classes after preprocessing (Figure 5). From the plot we can see that distributions of lag 1 and lag 2 ACF are different among three groups of patients. Interictal EEG has strong positive auto correlation while that of healthy EEG is relatively weak. This further lends credence to our assumption that EEG of different classes have significantly different properties, which lays the foundation of our classification methods.

We can also explore the differences among classes from a modeling's perspective. Though according to our results, RNN performs the best, it is a black box which lack interpretability. On the other hand, AR model also achieves satisfying results. Since AR model is relatively simple and has clear statistical explanation, it is more suitable for interpreting results. For a stationary AR model, its coefficients of the auto regressive term can help determine most of the important statistics of the time series. Difference in terms of auto regressive terms is a good reflection of the difference with respect to the internal pattern. Here we draw the boxplot of the optimal order of AR models (Figure 6). From the results we can see that the three groups are somewhat separable based on the optimal order and healthy data have in general higher optimal orders. Actually, there is work using only optimal order to do the classification[10]. However, we believe including information of the AR coefficients would give us higher accuracy.

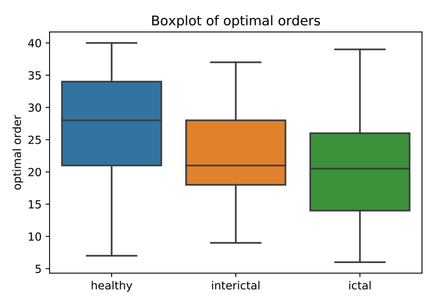



Figure 6: Boxplot of optimal orders

## 5 Conclusions

In this project, we build several time series models to conduct seizure detection task. Specifically, we compare the AR model, HMM model and RNN model in terms of their power to recognize the pattern of EEG signals from different classes, and we find that RNN has the best performance. Finally we obtain a RNN model which can achieves 0.98 accuracy on the test set.

Our work also has several limitations. First, after conducting GP-PAD, we only keep carriers for further analysis. However, there may be valuable information in the envelope that can be used by other models. Second, our data set is not large enough to build a robust model. Third, we only use single channel EEG data. In the future, we may need to explore more methods, such as Gaussian process and deep Kalman Filter. And we shall also make comparison between cross-patients modeling and patient-by-patient modeling.

## References


[1] SR Mousavi, M Niknazar, and B Vosoughi Vahdat. Epileptic seizure detection using ar model on eeg signals. In *Biomedical engineering conference, 2008. CIBEC 2008. Cairo International*, pages 1–4. IEEE, 2008.

[2] Srinivasan Vairavan, Eswaran Chikkannan, and Sriraam Natarajan. Approximate entropy-based epileptic eeg detection using artificial neural networks. *IEEE Transactions on Information Technology in Biomedicine A Publication of the IEEE Engineering in Medicine Biology Society*, 11(3):288, 2007.

[3] Hojjat Adeli, Samanwoy Ghosh-Dastidar, and Nahid Dadmehr. A wavelet-chaos methodology for analysis of eegs and eeg subbands to detect seizure and epilepsy. *IEEE Transactions on Biomedical Engineering*, 54(2):205–211, 2007.

[4] Stephen Faul, Gregor Gregorcic, Geraldine Boylan, William Marnane, Gordon Lightbody, and Sean Connolly. Gaussian process modeling of eeg for the detection of neonatal seizures. *IEEE Transactions on Biomedical Engineering*, 54(12):2151–2162, 2007.

[5] Mohd Hafidz Abdullah, Jafri Malin Abdullah, and Mohd Zaid Abdullah. Seizure detection by means of hidden markov model and stationary wavelet transform of electroencephalograph signals. In *Biomedical and Health Informatics (BHI), 2012 IEEE-EMBS International Conference on*, pages 62–65. IEEE, 2012.

[6] Sachin S Talathi. Deep recurrent neural networks for seizure detection and early seizure detection systems. *arXiv preprint arXiv:1706.03283*, 2017.

[7] Richard E Turner and Maneesh Sahani. Demodulation as probabilistic inference. *IEEE Transactions on Audio, Speech, and Language Processing*, 19(8):2398–2411, 2011.

[8] Kyunghyun Cho, Bart Van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. Learning phrase representations using rnn encoder-decoder for statistical machine translation. *arXiv preprint arXiv:1406.1078*, 2014.

[9] 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*, 64(6):061907, 2001.

[10] JLH Carvalho, AF Rocha, Icaro Dos Santos, C Itiki, LF Junqueira, and FAO Nascimento. Study on the optimal order for the auto-regressive time-frequency analysis of heart rate variability. In *Engineering in Medicine and Biology Society, 2003. Proceedings of the 25th Annual International Conference of the IEEE*, volume 3, pages 2621–2624. IEEE, 2003.

[11] Meysam Golmohammadi, Saeedeh Ziyabari, Vinit Shah, Eva Von Weltin, Christopher Campbell, Iyad Obeid, and Joseph Picone. Gated recurrent networks for seizure detection. In *Signal Processing in Medicine and Biology Symposium (SPMB), 2017 IEEE*, pages 1–5. IEEE, 2017.

[12] Ramy Hussein, Hamid Palangi, Rabab Ward, and Z Jane Wang. Epileptic seizure detection: A deep learning approach. *arXiv preprint arXiv:1803.09848*, 2018.

[13] Pierre Thodoroff, Joelle Pineau, and Andrew Lim. Learning robust features using deep learning for automatic seizure detection. In *Machine learning for healthcare conference*, pages 178–190, 2016.

[14] Samanwoy Ghosh-Dastidar, Hojjat Adeli, and Nahid Dadmehr. Principal component analysis-enhanced cosine radial basis function neural network for robust epilepsy and seizure detection. *IEEE Transactions on Biomedical Engineering*, 55(2):512–518, 2008.

## A Contribution statement

- Shizhan Gong: data collection, implement RNN, write up
- Yakun Wang: data preprocessing, implement AR, write up
- Liangzhi Li: literature review, implement HMM, write up

## B Github link

https://github.com/670973787/Seizure-Detection