An Evolutionary Framework for Rare-Event Prediction Problems with Machine Learning

DATA SCIENCE PORTUGAL

Mauro Filipe da Silva Pinto

mauropinto@dei.uc.pt

10/12/2019

Outline

Introduction

Epilepsy

EEG

Seizure Prediction

The Gap

Problems and Limitations

Interpretability

Evolutionary Algorithms

A search-algorithm

Application to Seizure Prediction

Hands on

Methodology

An Evolutionary Framework

Results

Training and Testing

Performance Evaluation

Discussion

Clinical Knowledge

A good Example

Limitations

Conclusion

Introduction

Epilepsy

- Epilepsy is a chronic disorder of the brain affecting people of all ages, **characterized by seizures.**
- In 30% of times, treatment is not effective, leading to Drug Resistant Patients.
- Why is prediction important?

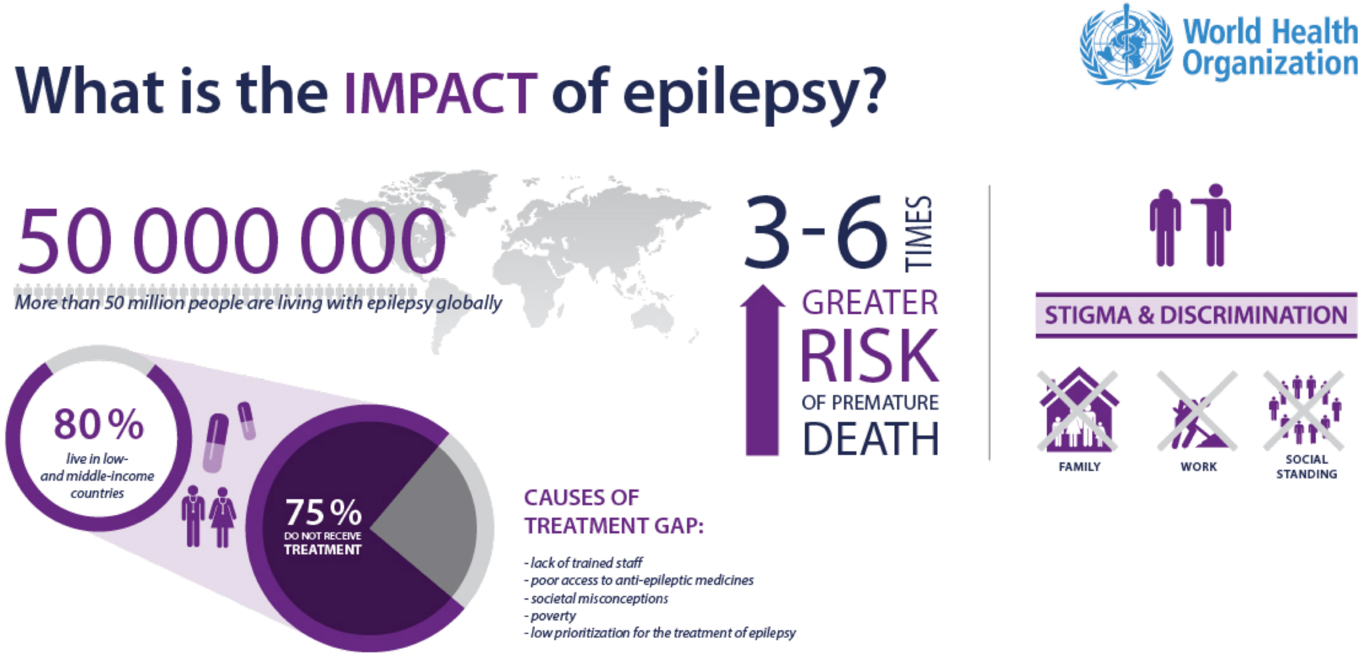


Figure 1:
http://apps.who.int/mediacentre/infographic/mental-health /epilepsy/en/index.html

Introduction

EEG

- The EEG captures the electrical activity in the brain in a time series.
- Main tool used by the physician.
- 4 periods for each seizure: the **Pre-Ictal and Inter-ictal** are the most relevant.

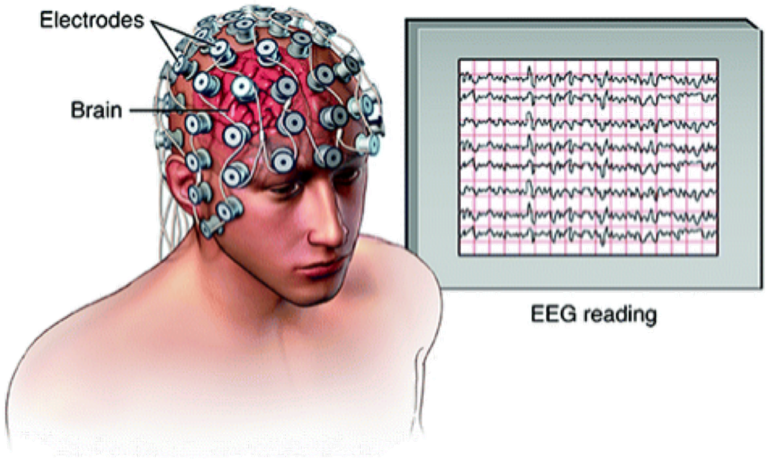


Figure 2: https://drmridha.com/services/eeg

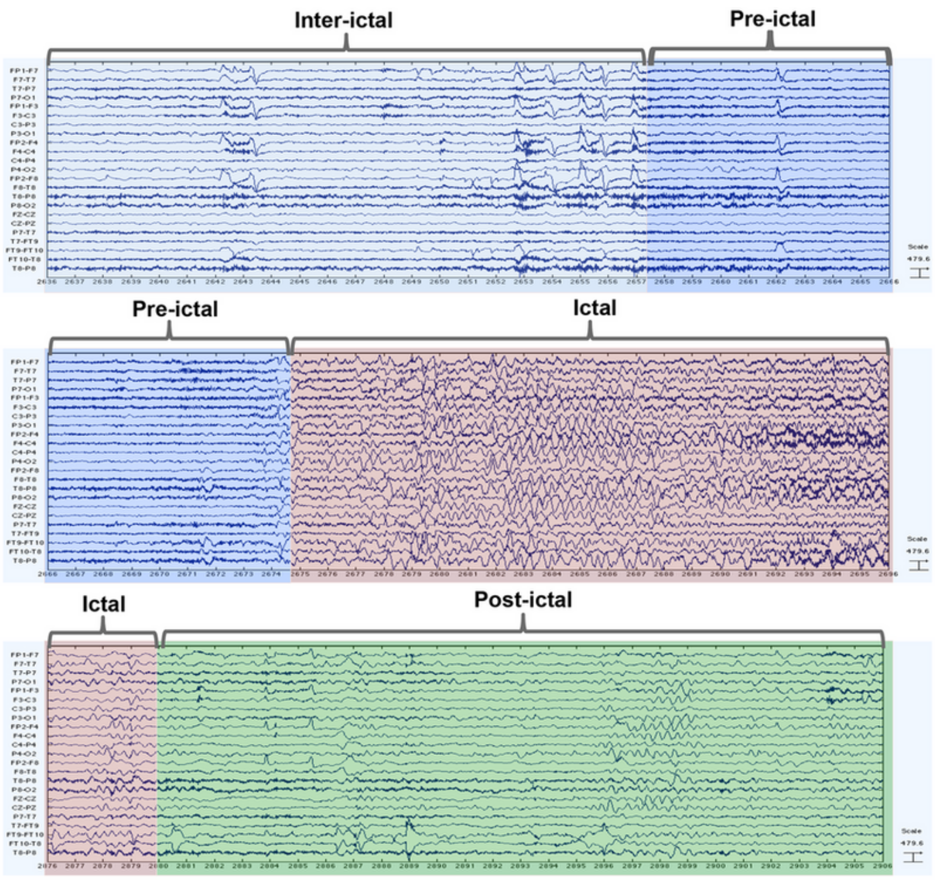


Figure 3: Moghim, N., & Corne, D. W. (2014). Predicting epileptic seizures in advance. PLoS ONE, 9(6). https://doi.org/10.1371/journal.pone.0099334

Introduction

Seizure Prediction

- This problem is a rare-event prediction task in a time-series.
- To receive online data and to timely anticipate anticipate an alarm.
- **Seizure Prediction Horizon (SPH)**
  - an intervention time
- **Seizure Occurrence Period (SOP)**
  - where the seizure will occur

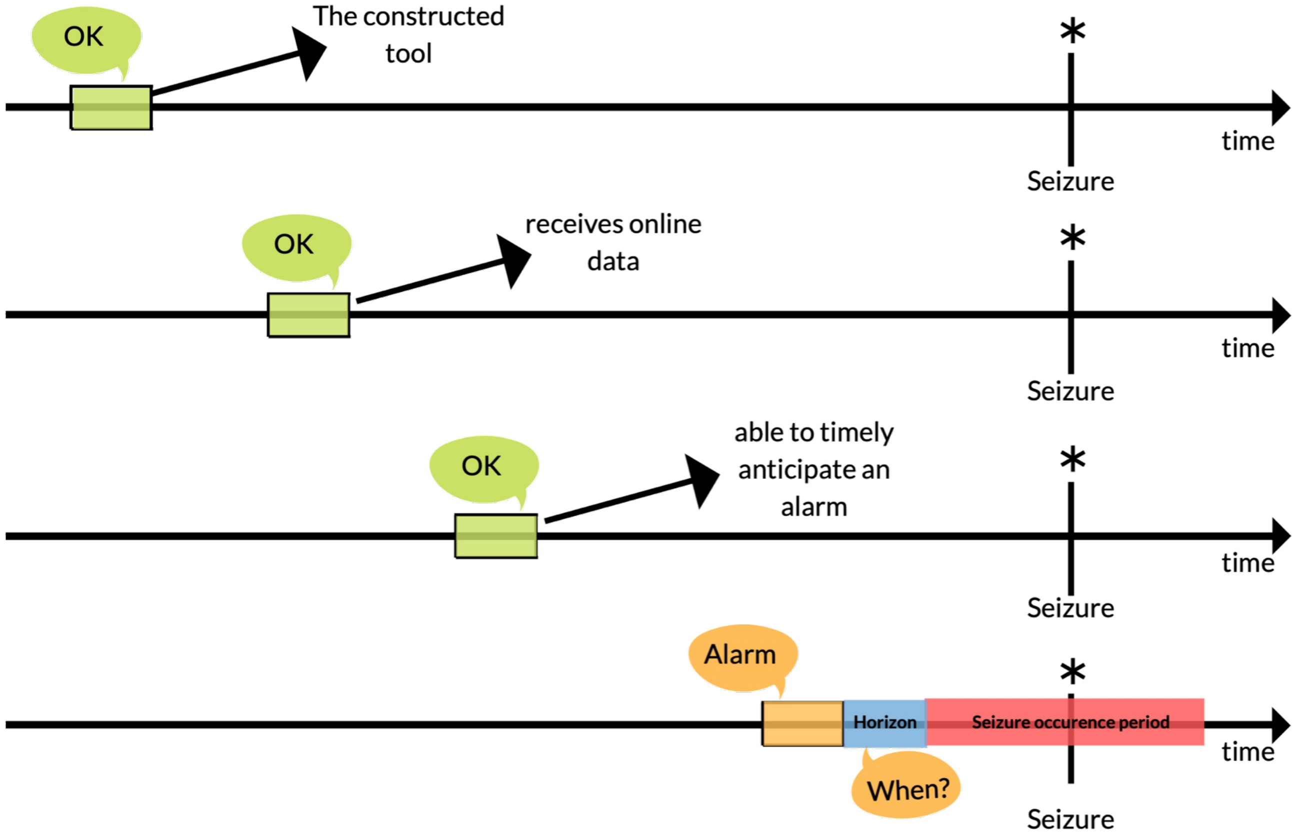
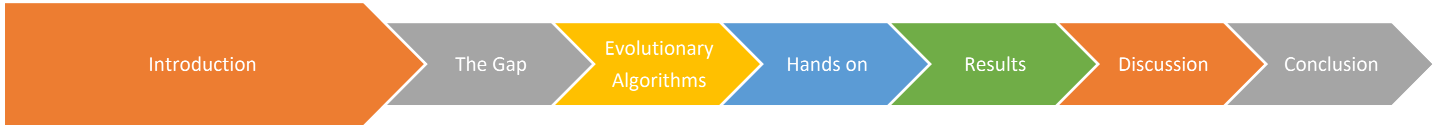


Figure 4: The objective of EEG seizure prediction

Introduction

Seizure Prediction



- For an alarm to be considered true: the seizure must occur within the SOP period.

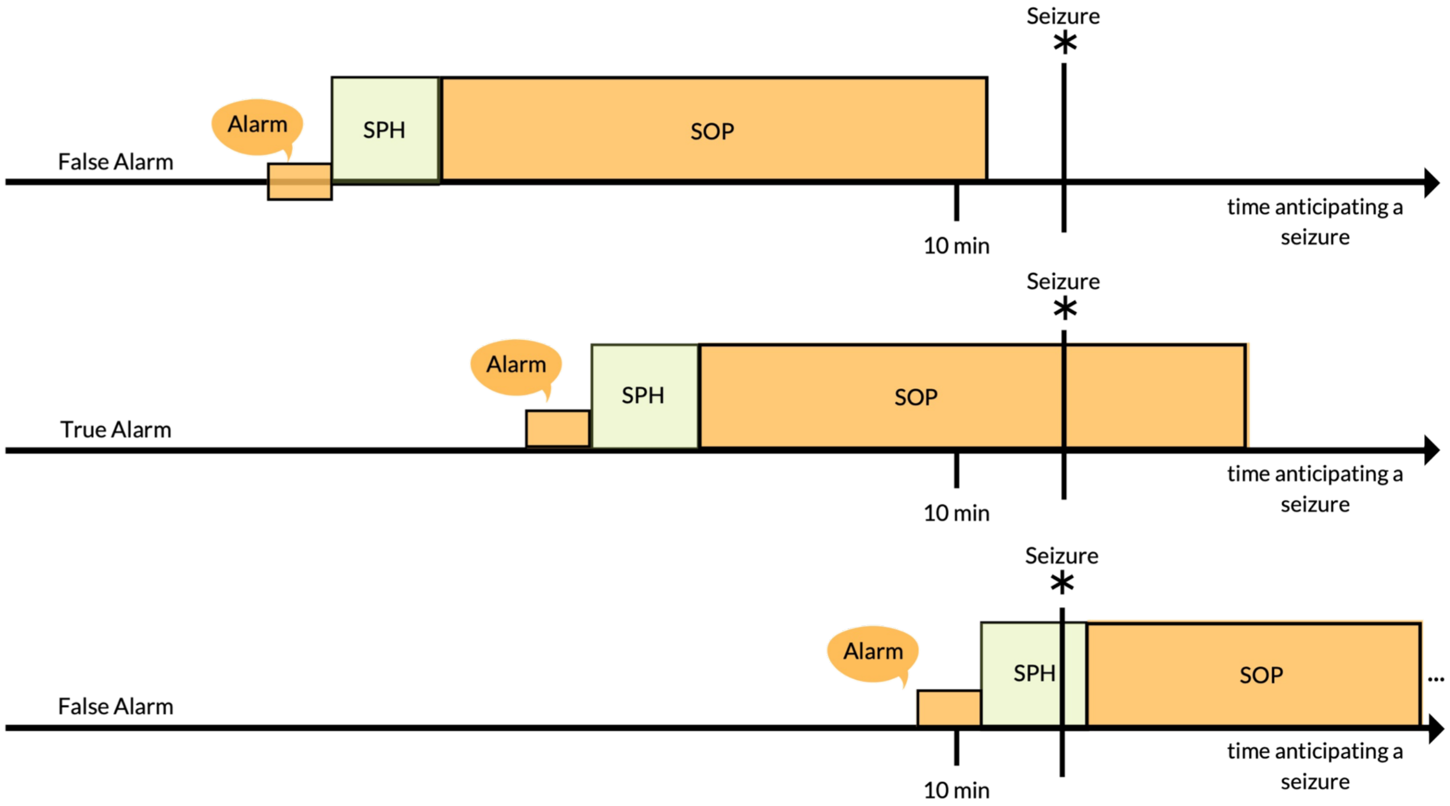
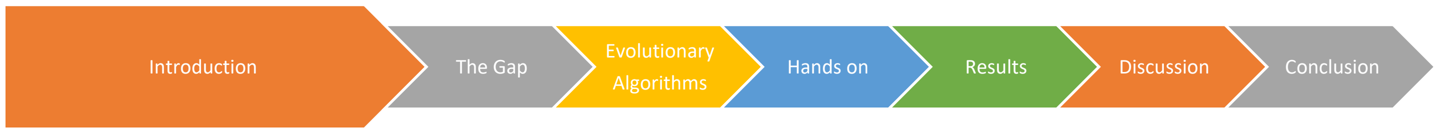


Figure 5: The practical definition of a True Alarm and False Alarm

Introduction

Seizure Prediction



- Gold Standard Metrics:
  - Sensibility: $\frac{Number\ of\ Predicted\ Seizures}{Number\ of\ Total\ Seizures}$
  - FPR/h= $\frac{Number\ of\ False\ Alarms}{Inter-ictal\ Period\ duration}$

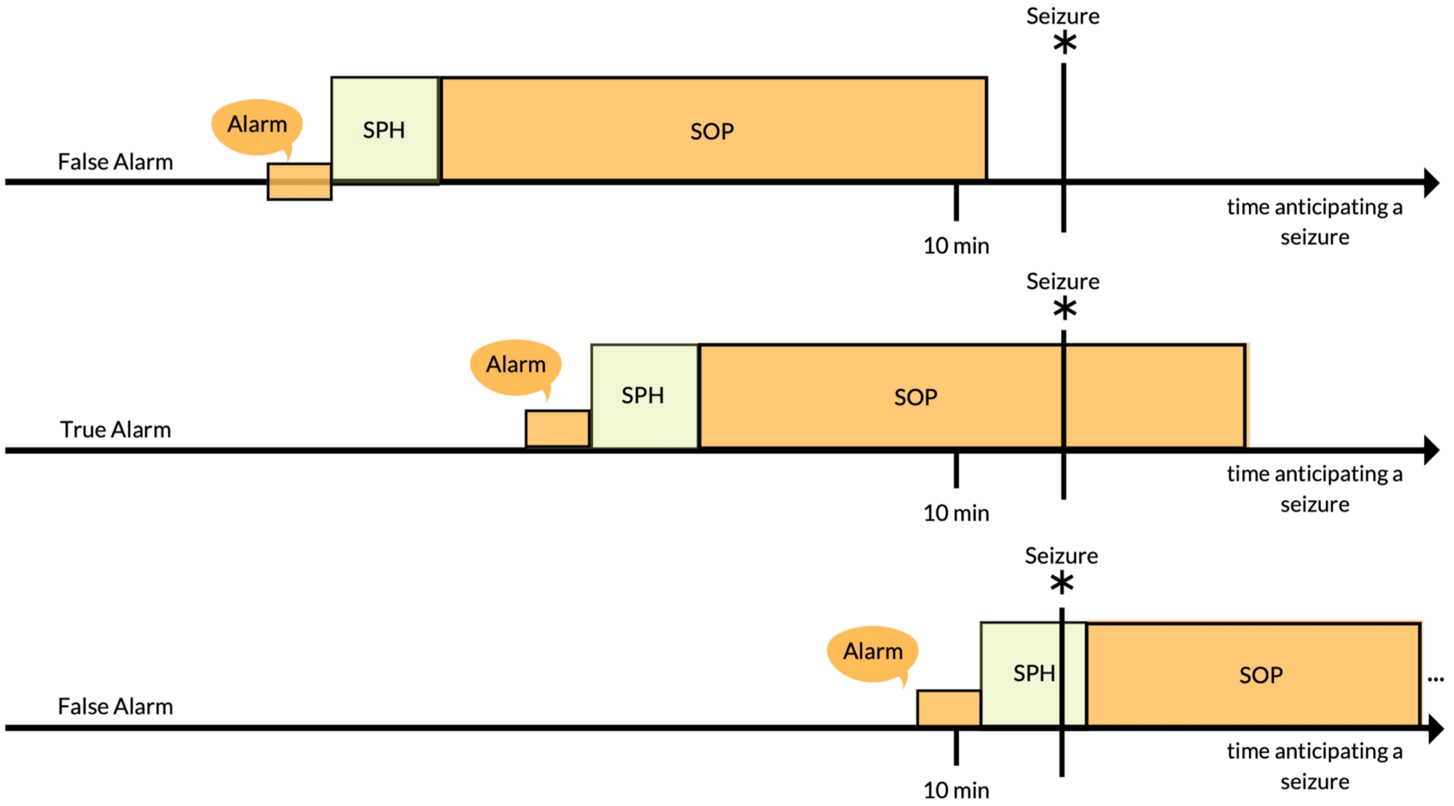


Figure 5: The practical definition of a True Alarm and False Alarm

Introduction

Seizure Prediction

- Statistical Validation
  - the probability of an unspecific predictor anticipating at least k of K seizures

$$P_{binom,d}(k,K,P) = 1 - \left[\sum_{1}^{j\leq k} \binom{K}{j} P^j (1-P)^{K-j}\right]^d$$

- Surrogate time-series validation may be more adequate as it handles different assumptions.

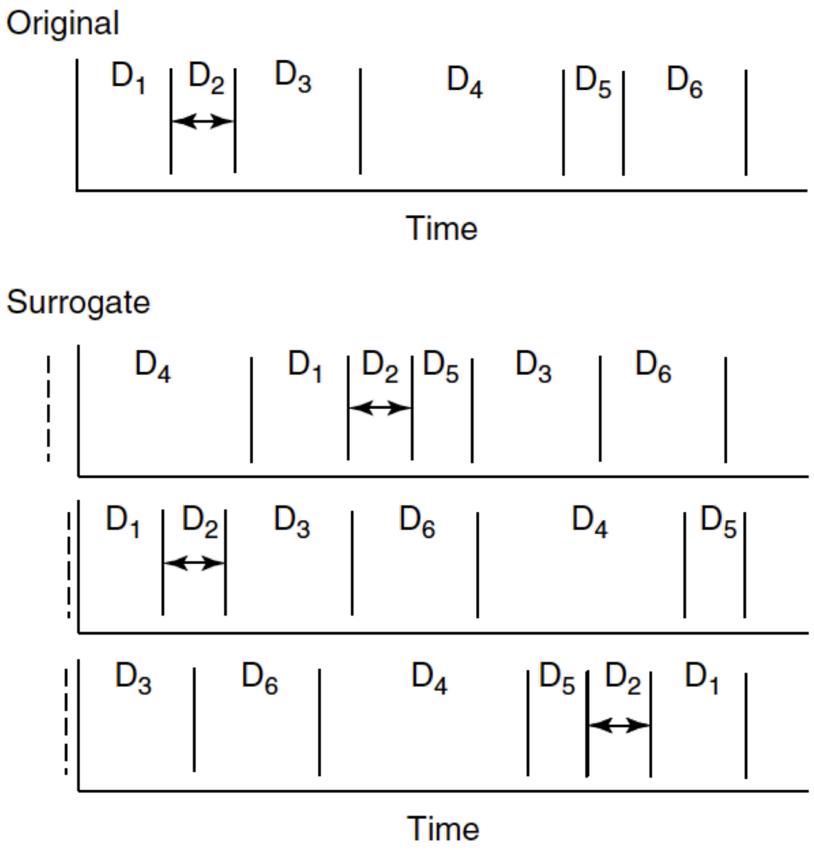


Figure 6: Original seizure times and the surrogate times bootstrapped from the inter-seizure intervals. The arbitrary onset times for the surrogates are obtained from a uniform distribution and are indicated by the dashed vertical lines.

Introduction

Extension to other rare-event prediction tasks in time-series

- A conversion for any rare-event prediction task in time-series is possible.

- SPH, SOP, Inter/Pre-ictal are just specific names for the seizure prediction problem.

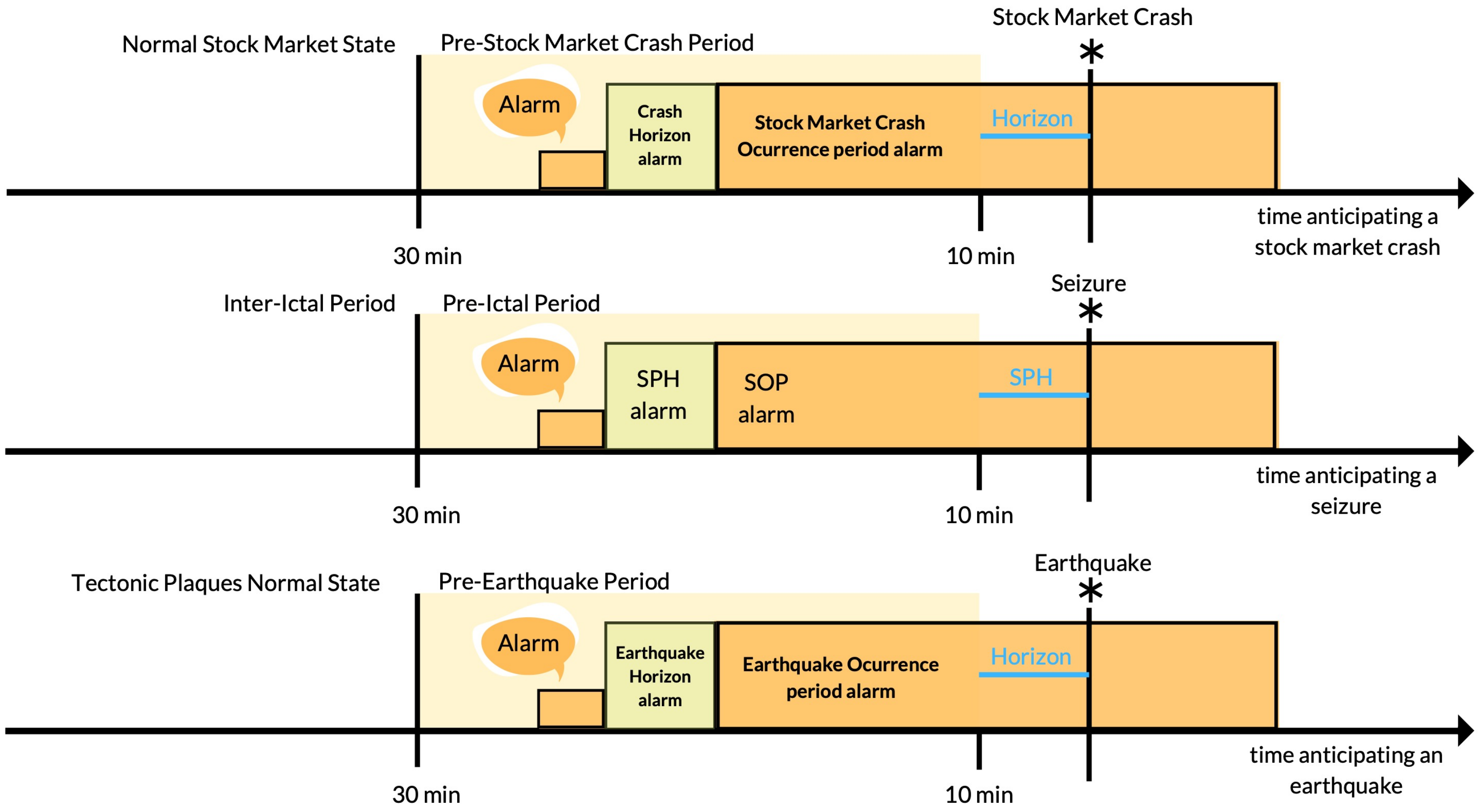


Figure 7 – An analogy between seizure prediction and earthquake and stock market crash prediction.

Introduction

Seizure Prediction

- Last, but not least: i**nterpretability.**
- Blackbox models cannot have a clinical translation.
- Natural Intelligence must be included.

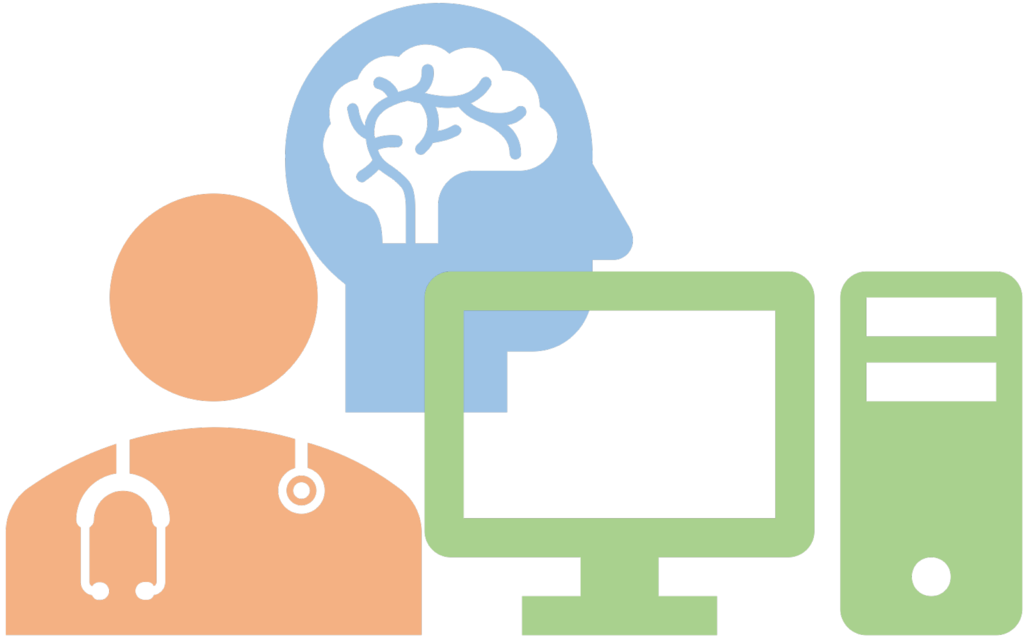


Figure 8: The importance of interpretability in seizure prediction systems.

The Gap

Pipeline

- Concept Drift
- Chronology of seizures
- Data imbalance

- **A rare-event prediction problem:**

  not a classic machine learning one.

- Complexity vs Interpretability.

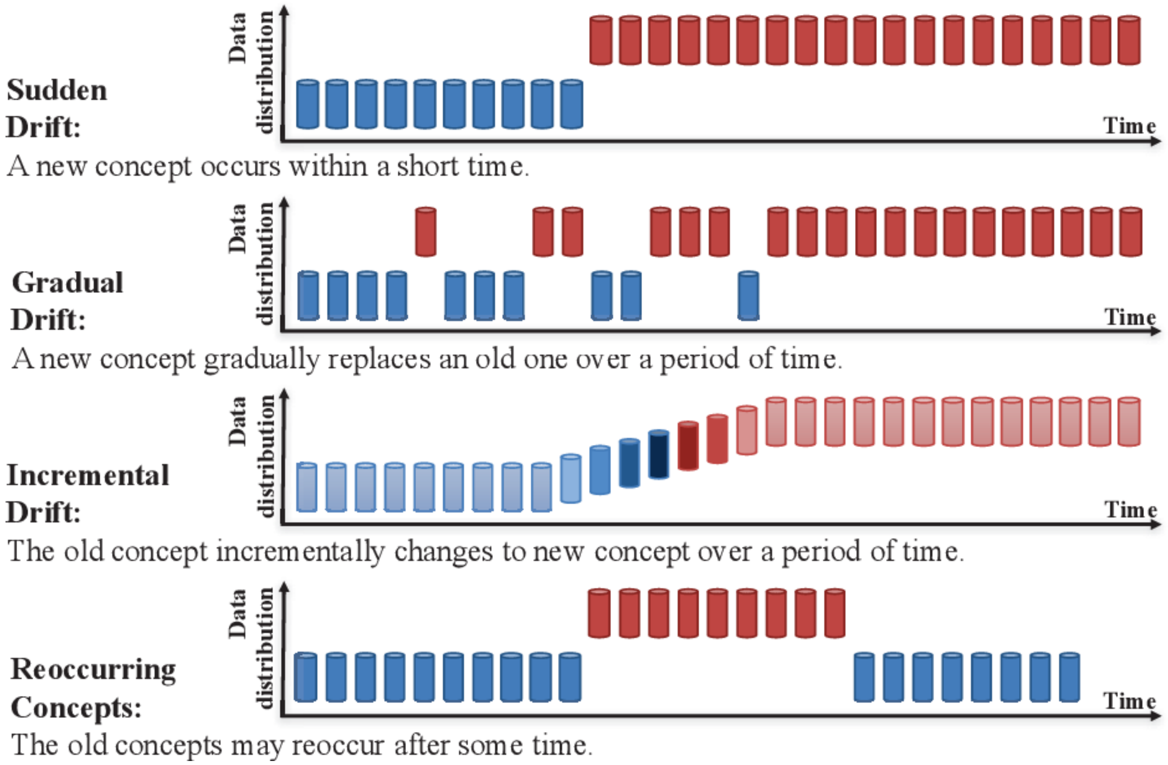


Figure 9: An example of concept drift.

The Gap

Pipeline - Limitations

- The existence of a standard framework that consists of:
  - Signal Pre-Processing;
  - Feature Extraction;
  - Feature Selection and Reduction;
  - Classification;
  - Regularization;
  - Performance Evaluation.
- Complexity vs Interpretability.

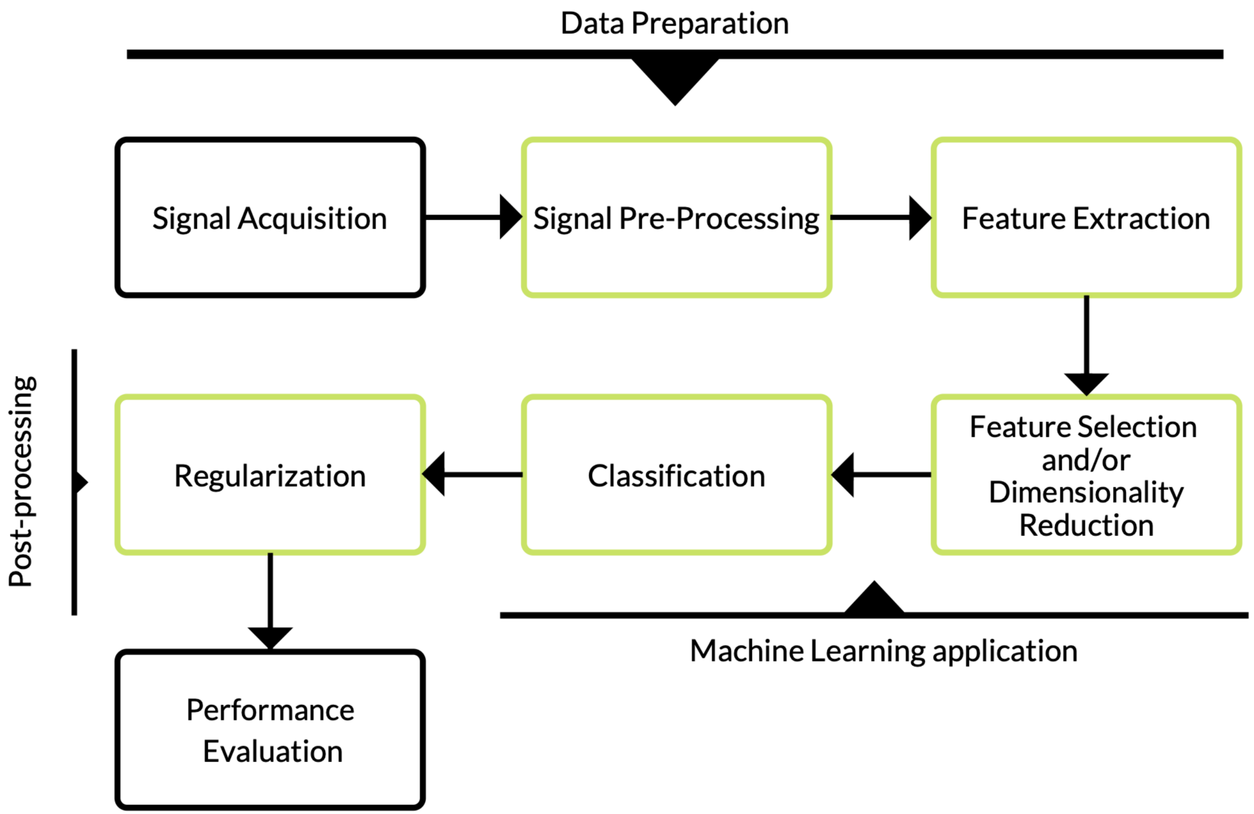


Figure 10 - A standard machine learning framework concerning EEG seizure prediction, adapted from Bou Assi, E., Nguyen, D. K., Rihana, S., & Sawan, M. (2017). Towards accurate prediction of epileptic seizures: A review. Biomedical Signal Processing and Control, 34, 144–157. https://doi.org/10.1016/j.bspc.2017.02.001

The Gap

Pipeline

- This pipeline can be different when using Deep Learning approaches.

- Deep Learning can be used to perform automatically signal processing, feature engineering and classification **(A).**
- Other authors provide as input EEG features instead of raw data **(B).**
- Deep Learning can also be used only for feature engineering purposes **(C)**.

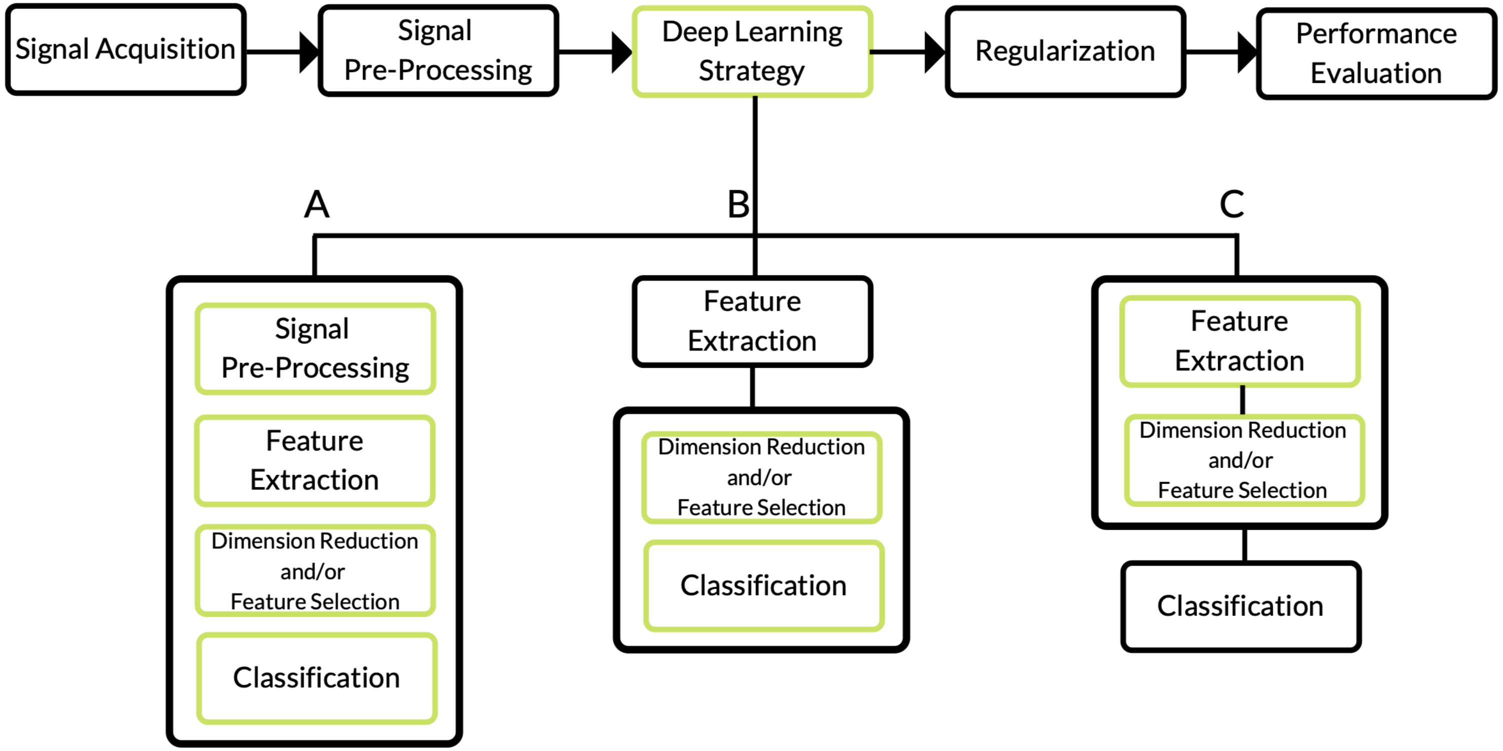
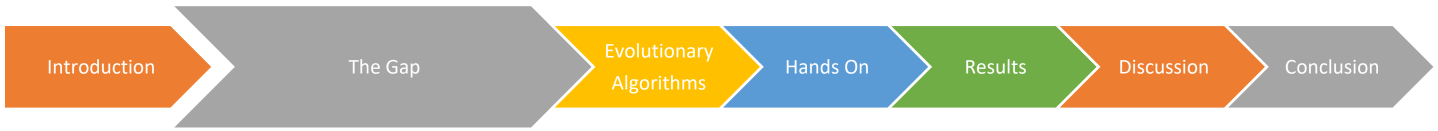


Figure 11 - Current Deep Learning approaches on EEG seizure prediction. While a Deep Learning algorithm is able to perform automatically signal processing, feature engineering and classification (alternative A), there are also authors that work with features as input (B) or even studies that use Deep Learning as a feature engineering method (C). Green steps are the ones performed with a deep learning model.

The Gap
Pipeline



- The used framework.
- According to a pre-ictal period threshold, features that best distinguish independent blocks of time are selected.
- Then, an ad-hoc regularization technique is trained to smooth the classification output.

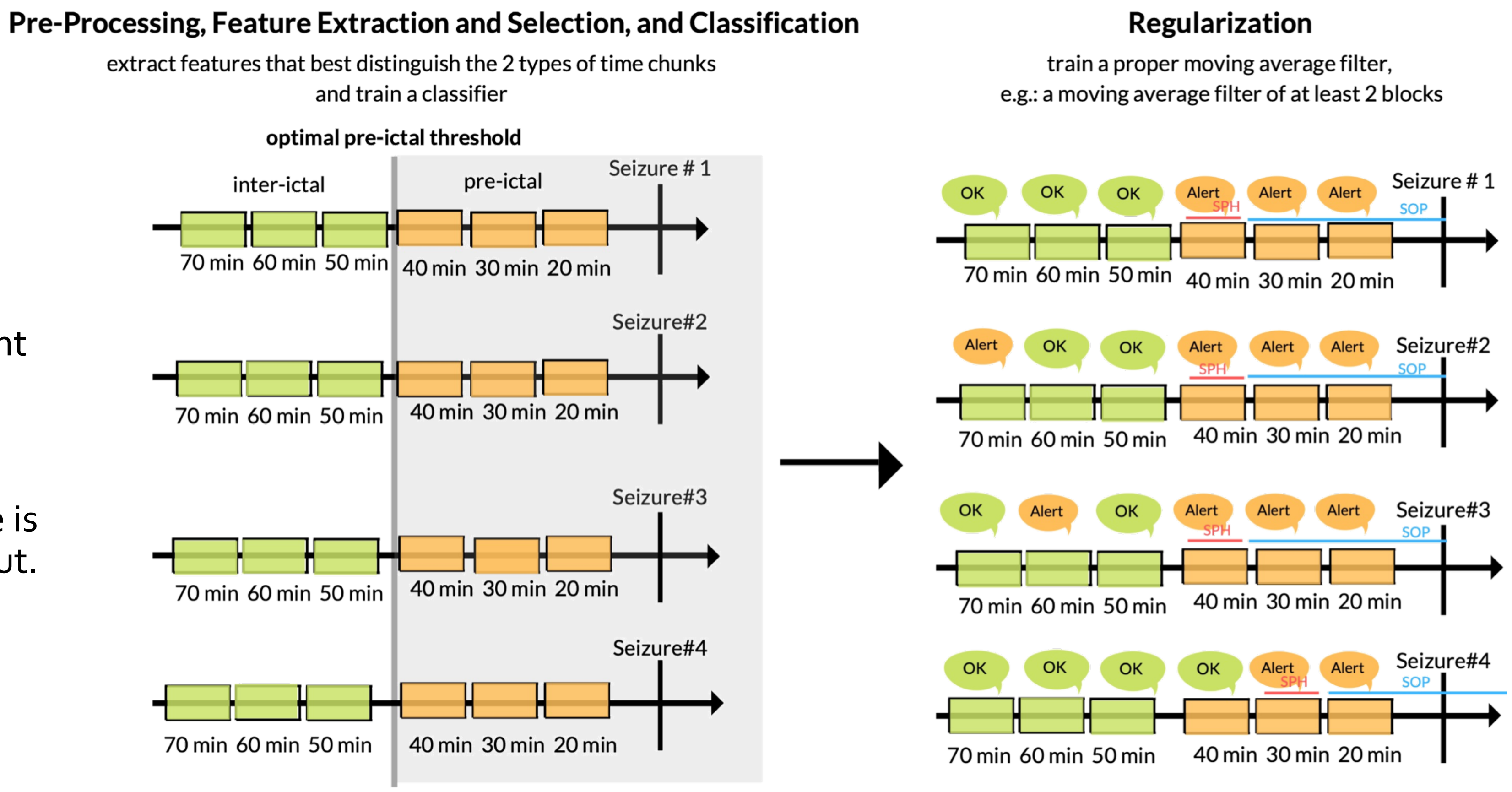


Figure 12 – Visual demonstration of the current pipeline methodology.

The Gap
Limitations found

Introduction | The Gap | Evolutionary Algorithms | Hands On | Results | Discussion | Conclusion

- **Signal Processing/Feature engineering must be envisioned as form of interpreting physiology.**

- A seizure is the consequence of a process comprised of a **chronological** brain activity.

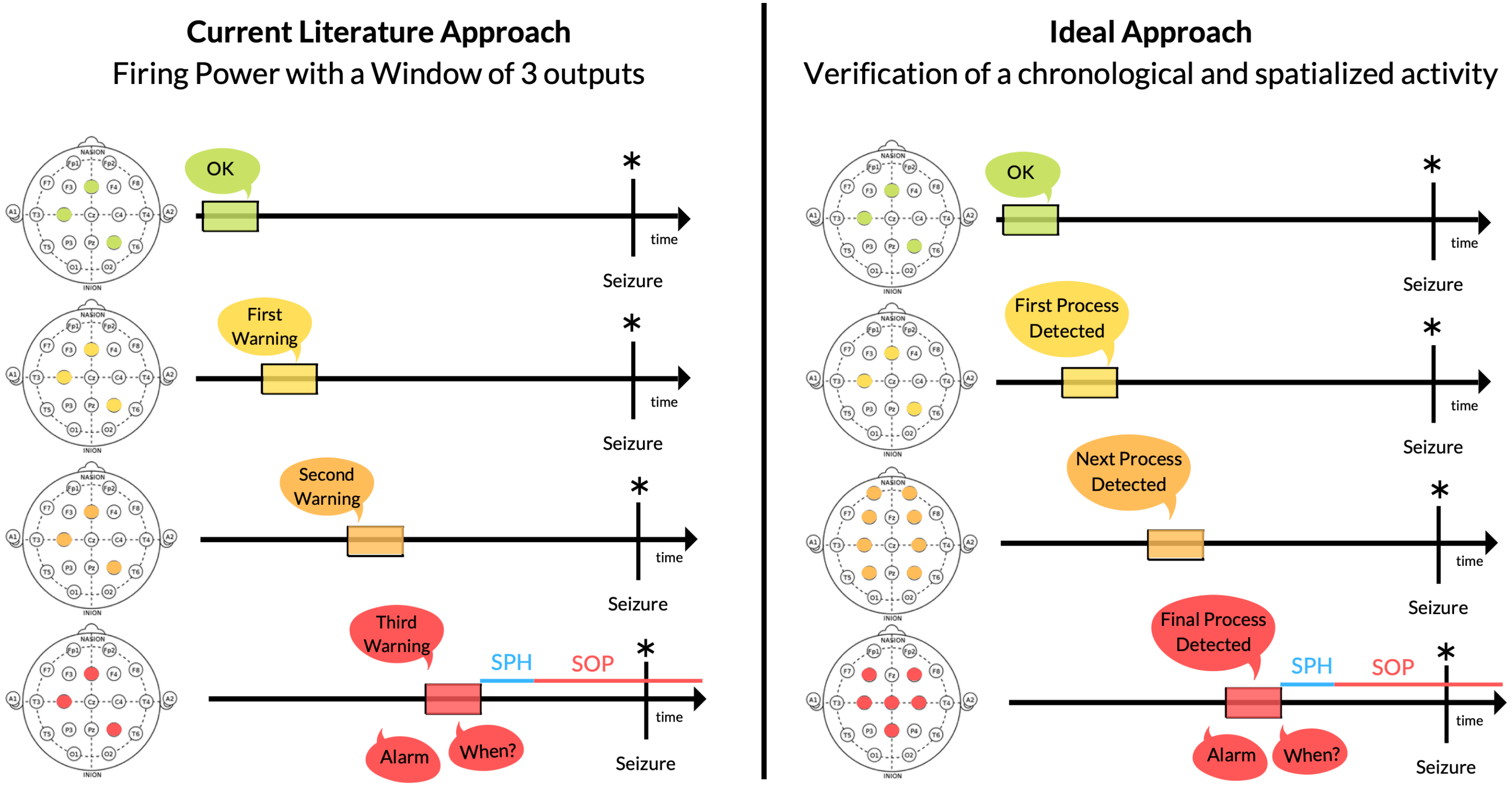


Figure 13 – The chronology concept.

The Gap

Limitations found

- **Signal Processing/Feature engineering must be envisioned as form of interpreting physiology.**

- Brain activity can have simultaneous spatial and temporal **multiscales**.

Figure 14 – The river example, spatial and temporal multiscales.

The Gap

Limitations found

- Deep Learning is able to do this, but how to interpret it?

- With a high number of features and possibilities, features are usually selected with filter methods.
- With this, interaction between features is not taken in account.

- Let's imagine:
  for 5 features, 8 different signal characteristics, 18 electrodes, 4 mathematical operators, 5 window lengths, 5 types of delays

$$\binom{5*8*18*4*5*5}{5} = \binom{72000}{5} = \frac{72000!}{5!(72000-5)!} \approx 1{,}61 * 10^{X}$$

possible combinations of sets of 5 features

The Gap

Limitations found

$\approx 1{,}61 * 10^{22}$ combinations

Evolutionary Algorithms

A search-algorithm

- **There is no need to evaluate all points in the search-space.**
- An Evolutionary Algorithm is a population-based metaheuristic algorithm, inspired by biological evolution.
- **Biological evolution is a search algorithm.**
- An example: bacteria vs antibiotics (simplified)

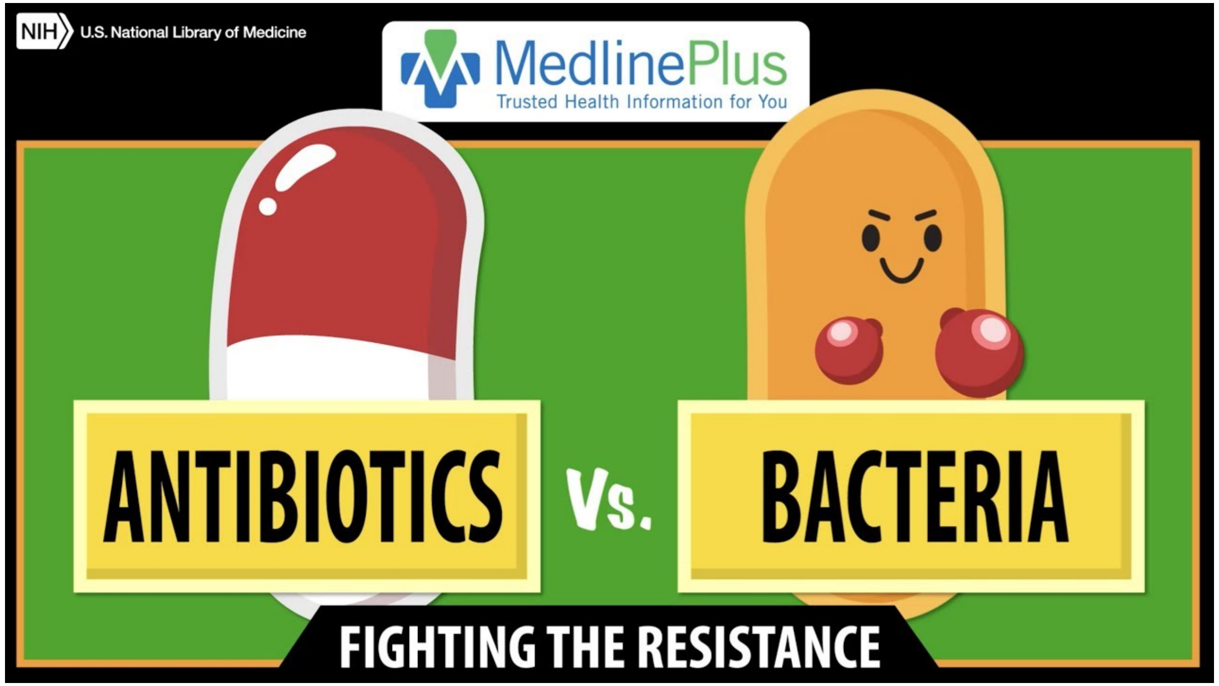


Figure 15 – U.S. National Library of Medicine. Antibiotics vs Bacteria

Evolutionary Algorithms

A search-algorithm

- Nature is searching for the bacteria that is antibiotic resistant

- Fitness is the individuals "ability to survive"

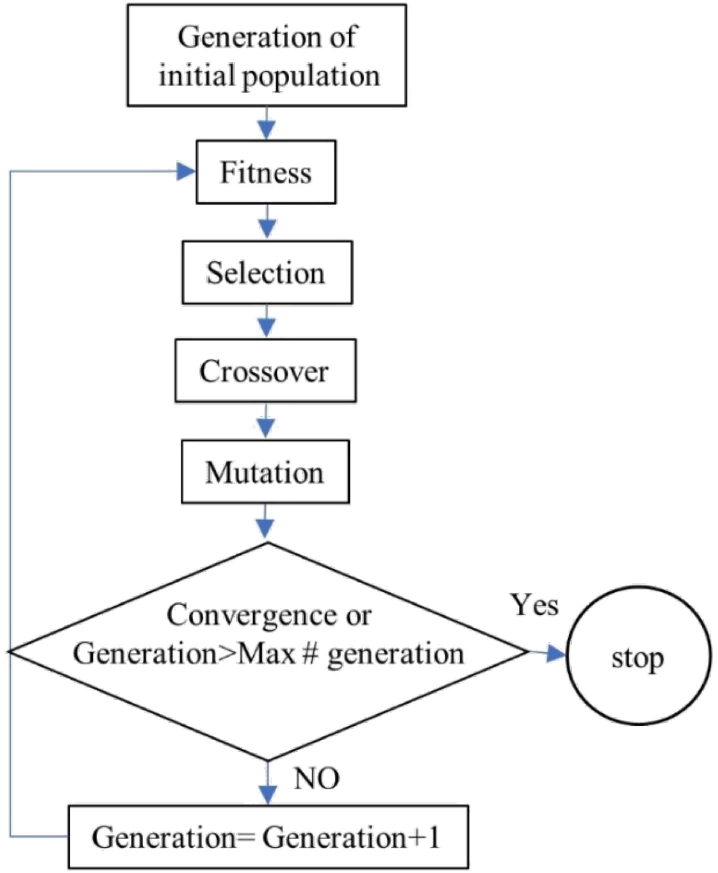


Figure 16 – Evolutionary Algorithm Flowchart, from Leal-Naranjo, José-Alfredo, et al. "Multi-objective optimization of a parallel manipulator for the design of a prosthetic arm using genetic algorithms." *Latin American Journal of Solids and Structures* 15.3 (2018).

Hands On

Seizure Prediction using an Evolutionary Algorithm

- My “bacteria”: a set of 5 features that will be used to predict seizures.
- Fitness: the performance on predicting seizures iteratively
- The best ones will disseminate their genetic material

Patient-Specific: Chronological Seizures

	Evolutionary Algorithm Training Seizures				Testing Seizures		
	Seizure #1	Seizure #2	Seizure #3	Seizure #4	Seizure #5	Seizure #6	Seizure #7
Test Seizure #2	Train	Test					
Test Seizure #3	Train	Train	Test				
Test Seizure #4	Train	Train	Train	Test			
Test Seizure #5	Train	Train	Train	Train	Test		
Test Seizure #6	Train	Train	Train	Train	Train	Test	
Test Seizure #7	Train	Train	Train	Train	Train	Train	Test
	Evaluate Training Performance and Choose Best Individuals According to Criteria				Evaluate Testing Performance		

Figure 17 – Fitness Function Scheme

Hands On

Genotype

- An individual is a set of F features:

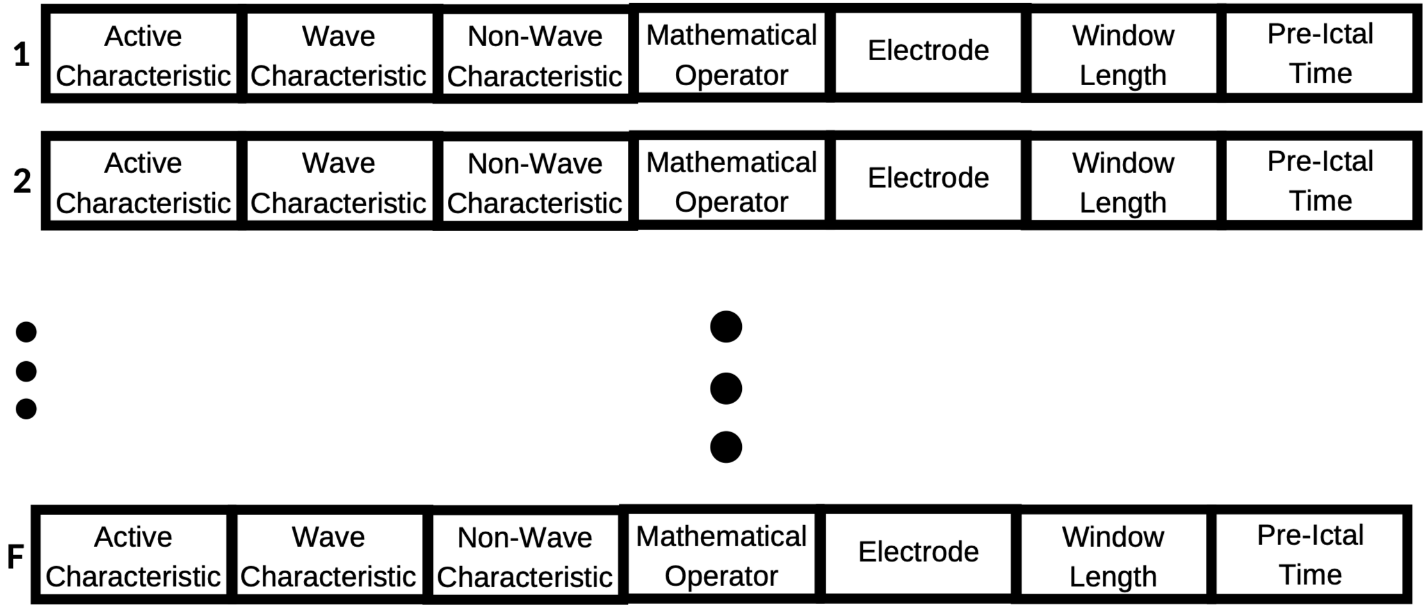


Figure 18 – The Genotype of an individual.

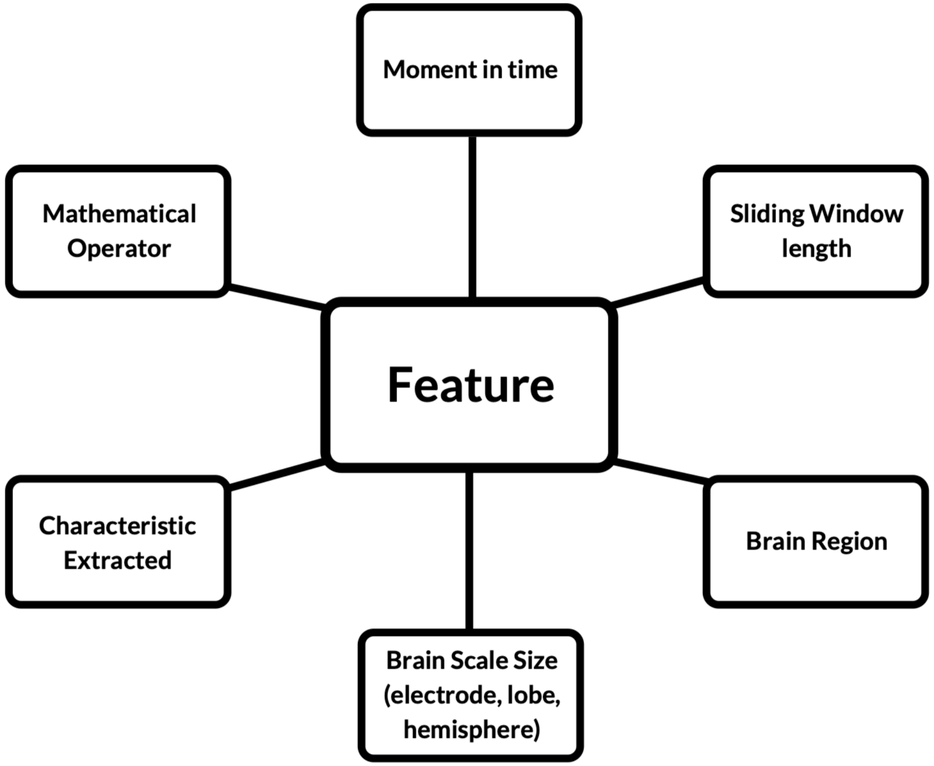
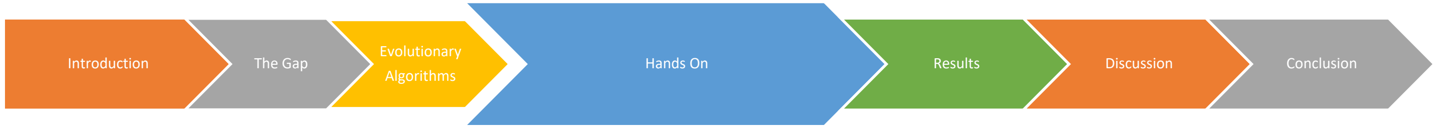


Figure 19 – A possible feature structure.

Hands On

Neighborhood concept



- The neighborhood of all genes

- The possible characteristics:
  - Mean Frequency
  - Band Power
  - Variance
  - Mean intensity
  - Relative Band Frequency Power

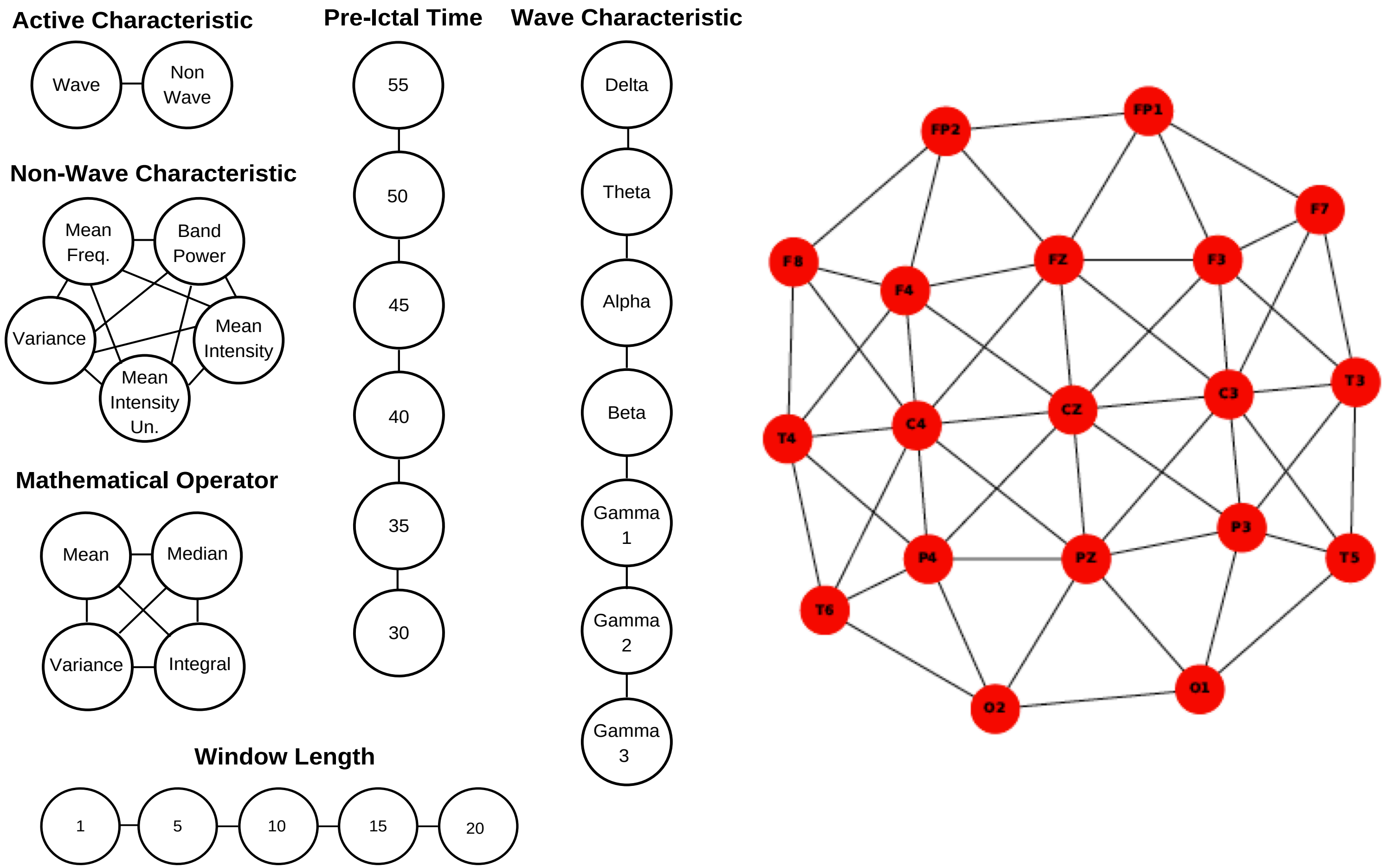


Figure 20 – The neighborhood concept for all genes.

Hands On

Genotype

A	Wave Origin	Alpha Wave	Band power	Variance	Cz	15 minutes	40 Minutes
B	Wave Origin	Beta Wave	Mean Frequency	Mean	O1	5 Minutes	20 Minutes
C	Non-Wave Origin	Beta Wave	Band Power	Integral	T6	1 Minute	30 Minutes

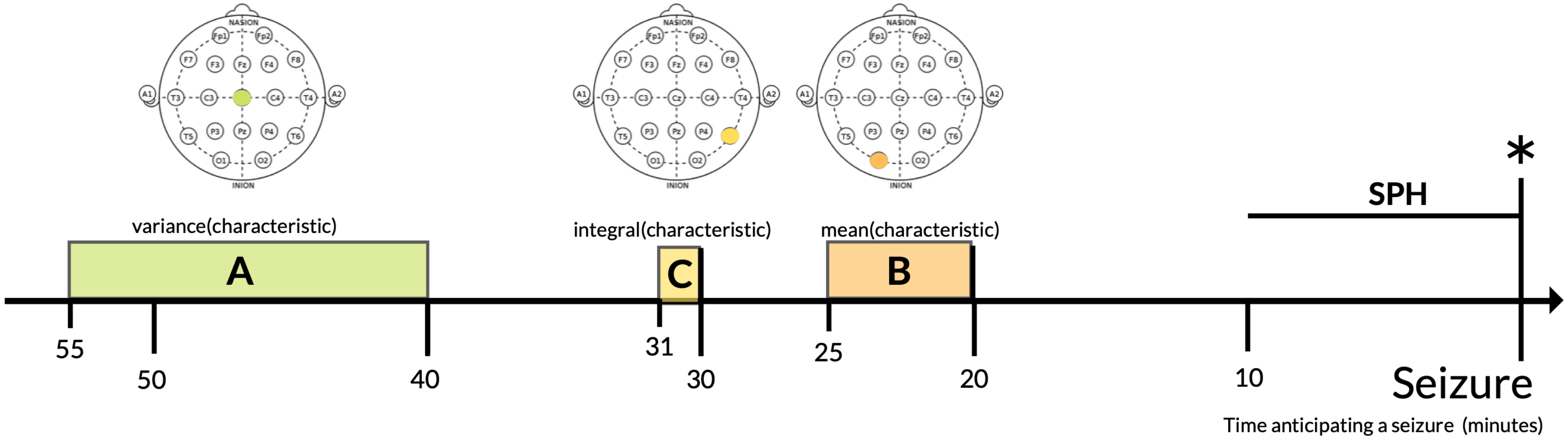


Figure 21 – An intuitive understanding of the genotype.

Hands On

Phenotype

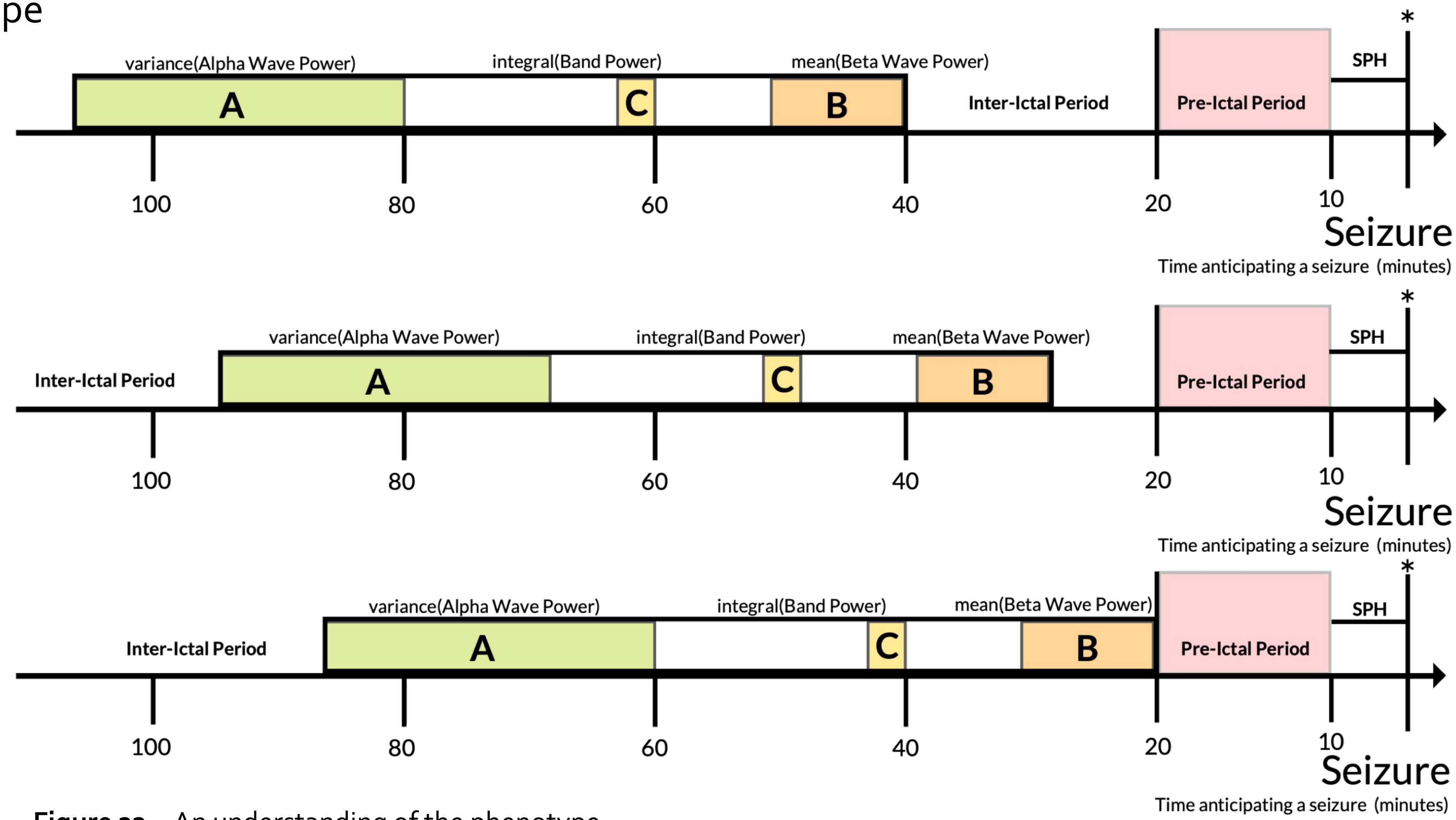


Figure 22 – An understanding of the phenotype.

Hands On

Fitness Function – Training Seizures

1) Feature and label extraction (with Training Seizures)

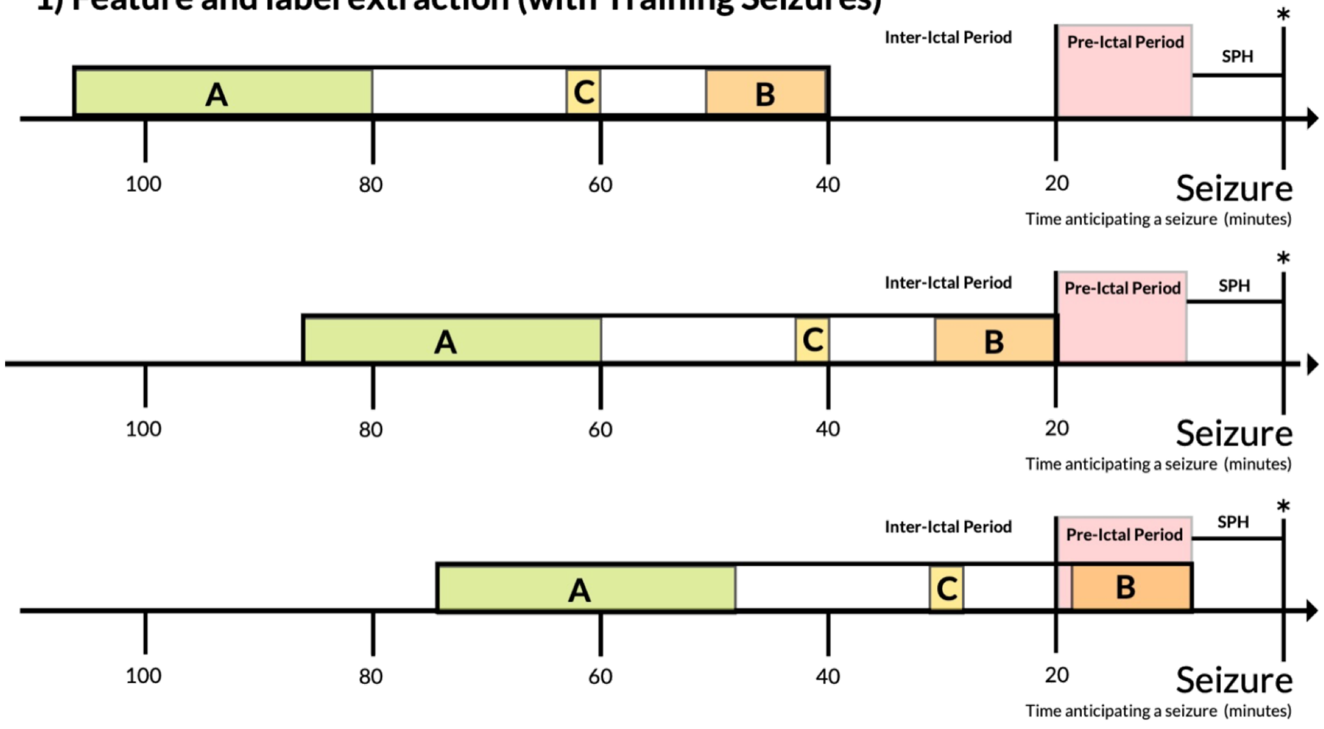


Figure 23 – Fitness function: training seizures

2) Machine Learning Binary Training

Feature 1	Feature 2	Feature 3	Label
A(t-1)	B(t-1)	C(t-1)	Inter-Ictal
A(t-2)	B(t-2)	C(t-2)	Inter-Ictal
A(t-3)	B(t-3)	C(t-3)	Inter-Ictal
A(t-4)	B(t-4)	C(t-4)	Inter-Ictal
A(t-5)	B(t-5)	C(t-5)	Inter-Ictal
A(t-6)	B(t-6)	C(t-6)	Inter-Ictal
A(t-7)	B(t-7)	C(t-7)	Inter-Ictal
A(t-8)	B(t-8)	C(t-8)	Inter-Ictal
A(t-9)	B(t-9)	C(t-9)	Inter-Ictal
A(t-10)	B(t-10)	C(t-10)	Inter-Ictal
...	...	...	...
A(t-n+1)	B(t-n+1)	C(t-n+1)	Pre-Ictal
A(t-n)	B(t-n)	C(t-n)	Pre-Ictal

Redundancy Removal correlation < 0.95

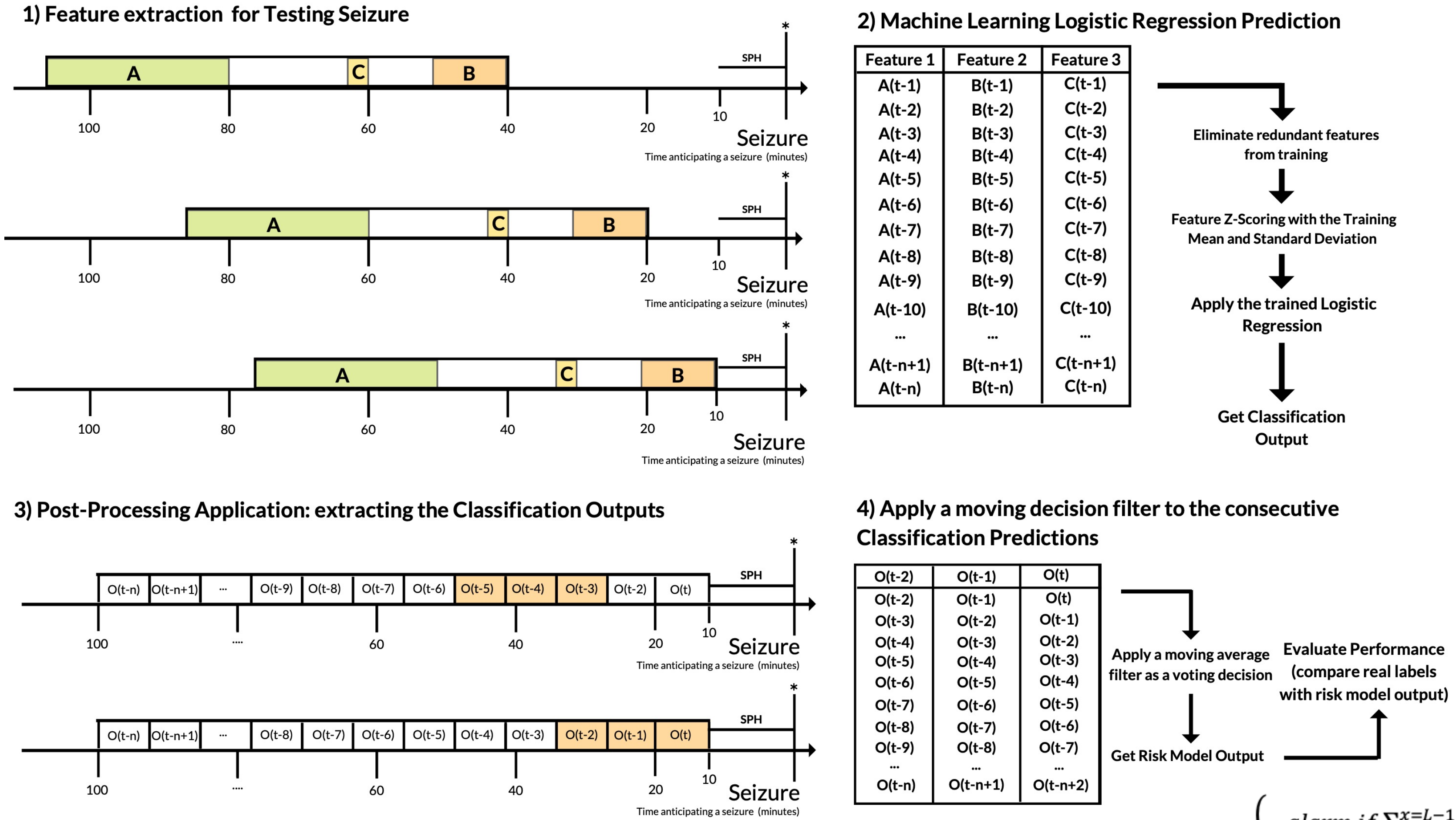
Feature Standardization (z-scoring)

Train a Logistic Regression Classifier

$$p(x) = \frac{1}{1+e^{-(\beta_0+\beta_1 x_1+\beta_2 x_2 \ldots+\beta_n x_n)}}$$

Hands On

Fitness Function – Testing Seizures



Feature 1	Feature 2	Feature 3
A(t-1)	B(t-1)	C(t-1)
A(t-2)	B(t-2)	C(t-2)
A(t-3)	B(t-3)	C(t-3)
A(t-4)	B(t-4)	C(t-4)
A(t-5)	B(t-5)	C(t-5)
A(t-6)	B(t-6)	C(t-6)
A(t-7)	B(t-7)	C(t-7)
A(t-8)	B(t-8)	C(t-8)
A(t-9)	B(t-9)	C(t-9)
A(t-10)	B(t-10)	C(t-10)
...	...	...
A(t-n+1)	B(t-n+1)	C(t-n+1)
A(t-n)	B(t-n)	C(t-n)

O(t-2)	O(t-1)	O(t)
O(t-2)	O(t-1)	O(t)
O(t-3)	O(t-2)	O(t-1)
O(t-4)	O(t-3)	O(t-2)
O(t-5)	O(t-4)	O(t-3)
O(t-6)	O(t-5)	O(t-4)
O(t-7)	O(t-6)	O(t-5)
O(t-8)	O(t-7)	O(t-6)
O(t-9)	O(t-8)	O(t-7)
...	...	...
O(t-n)	O(t-n+1)	O(t-n+2)

$$F.P.(t) = \begin{cases} alarm\ if\ \sum_0^{x=L-1} \frac{O(t-x)}{L} \geq t \\ no\ alarm\ if\ \sum_0^{x=L-1} \frac{O(t-x)}{L} < t \end{cases}$$

Figure 24 – Fitness function: testing seizures.

Hands On

Fitness Function – Evaluation Metrics

- *Fitness Seizure*

  $0.5 * (Seizure\ Sensitivty + Sample\ Sensitivity) - Penalty$

- Penalty

  FPR/h (1+Time Under False Alarm)

- The fitness value of one individual is the average fitness of all tested seizures.

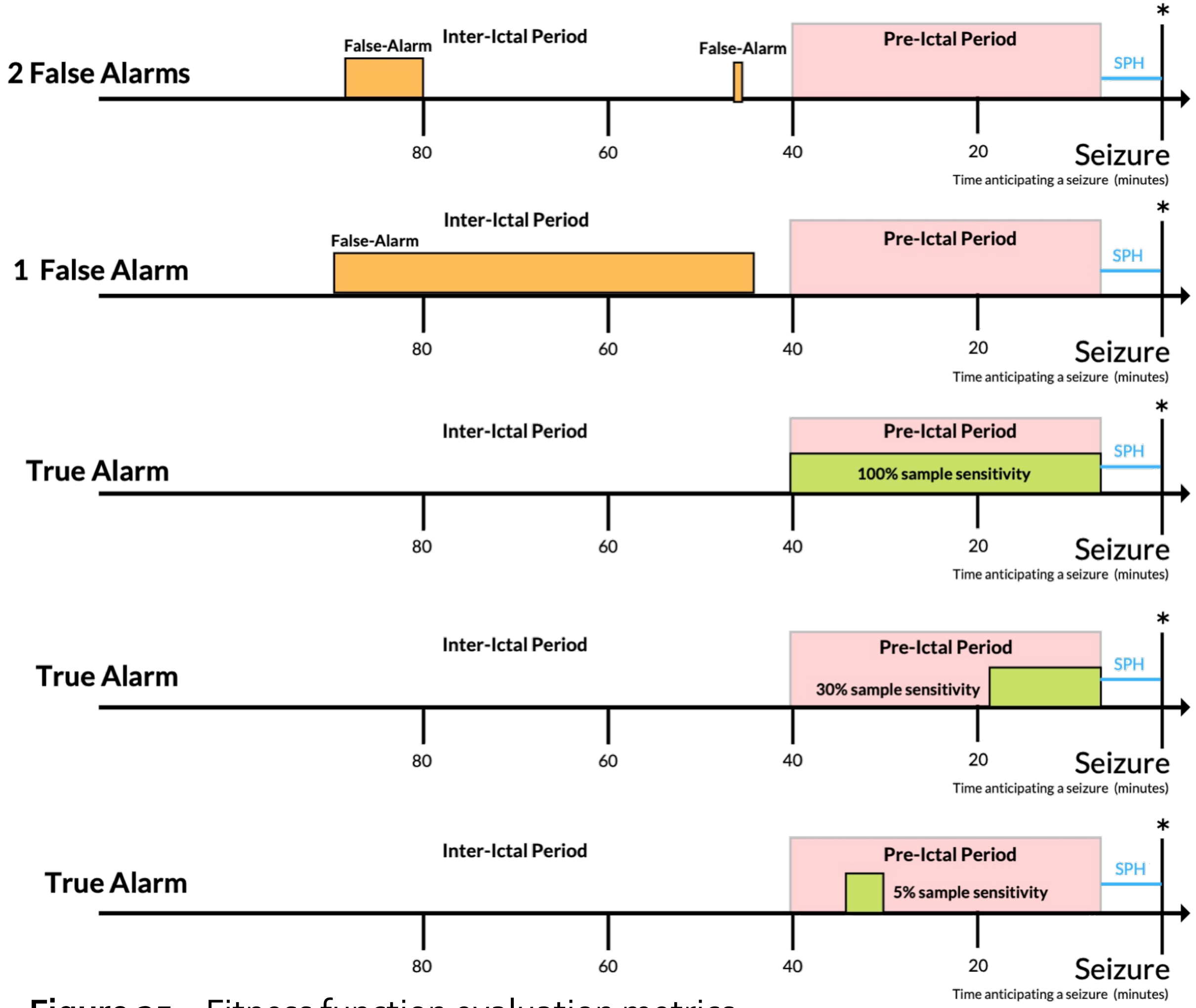


Figure 25 – Fitness function evaluation metrics

Results

Training

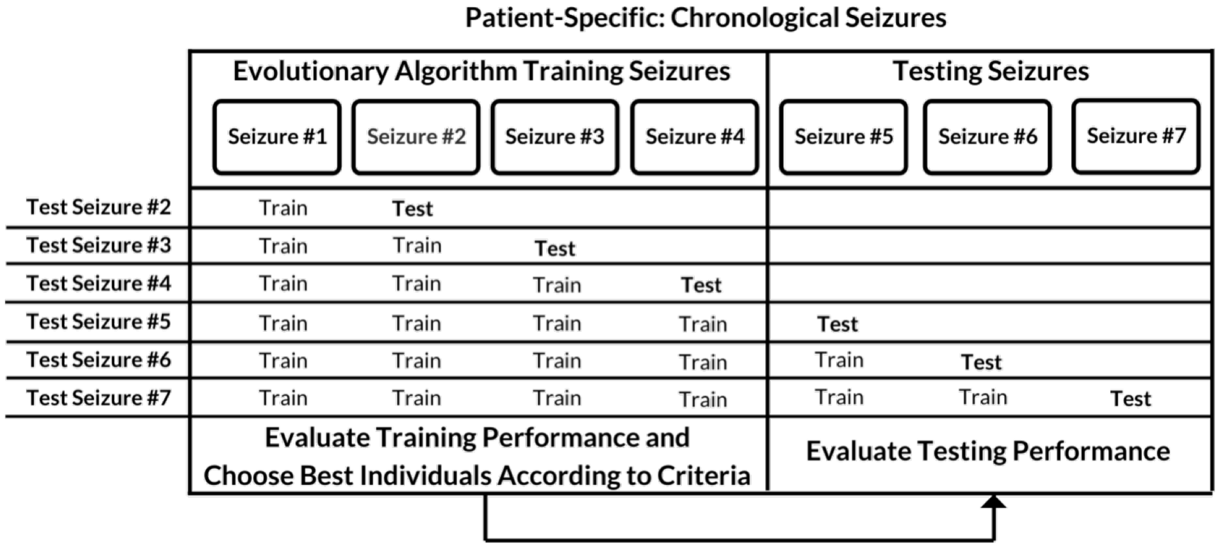


Figure 26 – Fitness Function Scheme

Table I – Training Results

Patient	Fitness	S_p	FPR/h	SOP	Seizures
1	1.00 ± 0.00	1.00 ± 0.00	0.00 ± 0.00	41.67 ± 4.71	2
2	0.83 ± 0.06	1.00 ± 0.00	0.05 ± 0.06	40.00 ± 0.00	3
3	0.73 ± 0.07	1.00 ± 0.00	0.09 ± 0.06	40.00 ± 0.00	3
4	$\mathbf{1.00 \pm 0.00}$	$\mathbf{1.00 \pm 0.00}$	$\mathbf{0.00 \pm 0.00}$	$\mathbf{40.00 \pm 0.00}$	2
5	0.81 ± 0.01	1.00 ± 0.00	0.10 ± 0.00	50.00 ± 0.00	4
6	0.64 ± 0.07	0.92 ± 0.12	0.13 ± 0.05	50.00 ± 0.00	4
7	0.98 ± 0.02	1.00 ± 0.00	0.00 ± 0.00	45.00 ± 0.00	3
8	0.98 ± 0.02	1.00 ± 0.00	0.00 ± 0.00	43.33 ± 2.36	2
9	0.87 ± 0.03	1.00 ± 0.00	0.00 ± 0.00	38.33 ± 2.36	3
10	0.83 ± 0.06	1.00 ± 0.00	0.04 ± 0.06	38.33 ± 4.71	3
11	0.89 ± 0.14	1.00 ± 0.00	0.06 ± 0.09	36.67 ± 6.24	2
12	0.79 ± 0.00	1.00 ± 0.00	0.12 ± 0.00	40.0 ± 0.00	3
13	0.76 ± 0.02	1.00 ± 0.00	0.12 ± 0.00	40.0 ± 0.00	3
14	0.75 ± 0.09	0.89 ± 0.16	0.05 ± 0.06	53.33 ± 0.06	3
15	0.77 ± 0.16	0.89 ± 0.16	0.00 ± 0.00	46.67 ± 2.36	3
16	0.97 ± 0.01	1.00 ± 0.00	0.00 ± 0.00	33.33 ± 2.36	2
17	0.98 ± 0.01	1.00 ± 0.00	0.00 ± 0.00	40.0 ± 0.00	2
18	1.00 ± 0.00	1.00 ± 0.00	0.00 ± 0.00	51.67 ± 2.36	2
19	0.79 ± 0.03	1.00 ± 0.00	0.12 ± 0.00	33.33 ± 2.36	3
Mean	$\mathbf{0.86 \pm 0.01}$	$\mathbf{0.98 \pm 0.01}$	$\mathbf{0.05 \pm 0.01}$	$\mathbf{42.19 \pm 0.69}$	Sum= 52

Results

Training

Patient	Fitness	S_p	FPR/h	SOP	Seizures
Mean	$\mathbf{0.86 \pm 0.01}$	$\mathbf{0.98 \pm 0.01}$	$\mathbf{0.05 \pm 0.01}$	$\mathbf{42.19 \pm 0.69}$	Sum= 52

Results

Testing

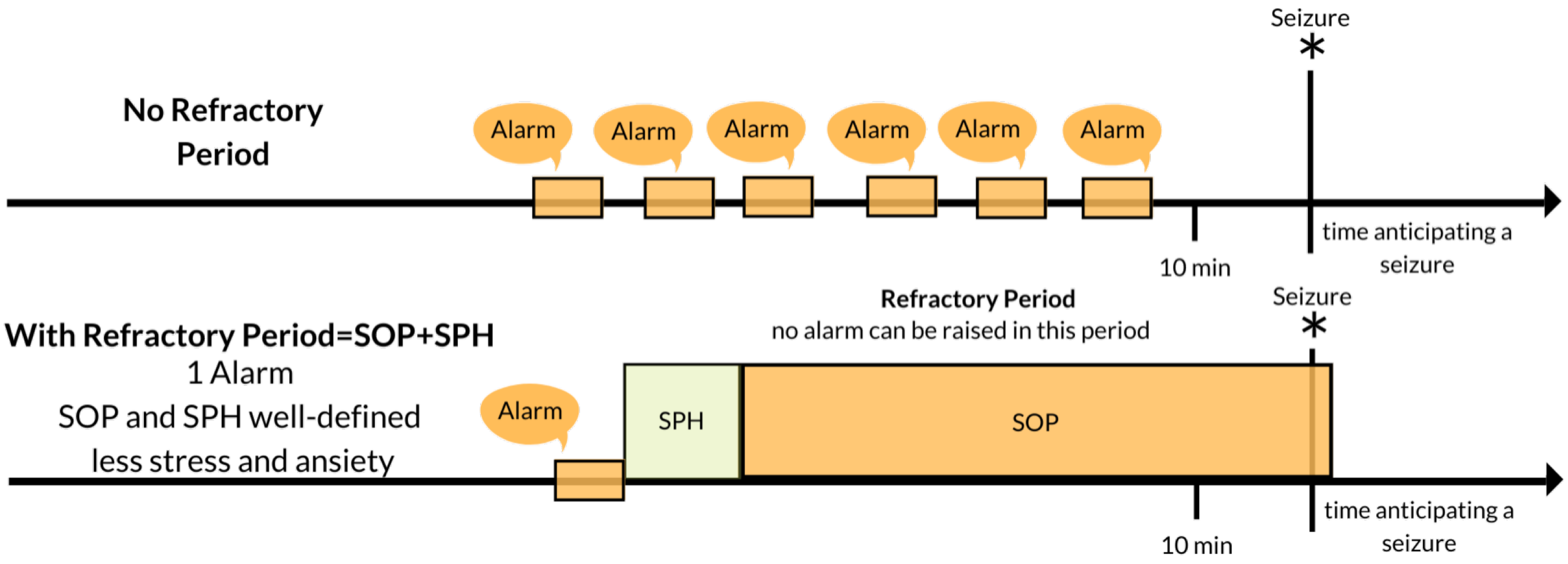


Figure 27 – Application of a refractory behavior

Results

Testing

Patient	S_p	FPR/h	Tested Seizures	Outperforms Random Pred.?	Outperforms Surrogate Method?
Mean	$\mathbf{0.47 \pm 0.06}$	$\mathbf{0.21 \pm 0.03}$	Sum= 46	**8=42%**	**12=63%**

Table II – Testing Results

Patient	S_p	FPR/h	Tested Seizures	Outperforms Random Pred.?	Outperforms Surrogate Method?
1	0.03 ± 0.12	0.16 ± 0.12	2	No	No
2	0.30 ± 0.28	0.13 ± 0.14	2	No	No
3	0.49 ± 0.22	0.46 ± 0.12	3	No	No
4	$\mathbf{0.95 \pm 0.15}$	$\mathbf{0.12 \pm 0.14}$	2	Yes	**Yes**
5	0.69 ± 0.28	0.25 ± 0.14	3	Yes	Yes
6	0.20 ± 0.24	0.15 ± 0.15	3	No	Yes
7	0.50 ± 0.37	0.10 ± 0.14	2	Yes	Yes
8	0.25 ± 0.25	0.23 ± 0.12	2	No	No
9	0.80 ± 0.24	0.28 ± 0.19	2	Yes	Yes
10	0.48 ± 0.27	0.33 ± 0.13	3	No	No
11	0.55 ± 0.33	0.16 ± 0.19	2	Yes	Yes
12	0.50 ± 0.13	0.12 ± 0.16	2	Yes	Yes
13	0.07 ± 0.13	0.18 ± 0.07	3	No	No
14	0.47 ± 0.22	0.22 ± 0.13	3	No	Yes
15	0.44 ± 0.31	0.30 ± 0.16	3	No	No
16	0.40 ± 0.27	0.26 ± 0.21	2	No	Yes
17	0.63 ± 0.22	0.12 ± 0.17	2	Yes	Yes
18	0.78 ± 0.31	0.21 ± 0.18	2	Yes	Yes
19	0.36 ± 0.20	0.20 ± 0.15	3	No	Yes
Mean	$\mathbf{0.47 \pm 0.06}$	$\mathbf{0.21 \pm 0.03}$	Sum= 46	8=42%	**12=63%**

Discussion

Are the results good?

- With the last study from the same lab, similar methodology, same database origin and same validation methods:

    38,47% sensitivity and FPR/h= 0,20

    11% validated patients

    10 second SPH

    Direito, Bruno, et al. "A realistic seizure prediction study based on multiclass SVM." *International journal of neural systems* 27.03 (2017): 1750006.

- This study:

    46% sensitivity and FPR/h=0,21

    42% validated patients

    10 minutes SPH

- **However: 19 patients were used due to computational cost. This constitutes only a proof-of-concept.**

Discussion

Stratification

- Stratification based on clinical aspects improves results

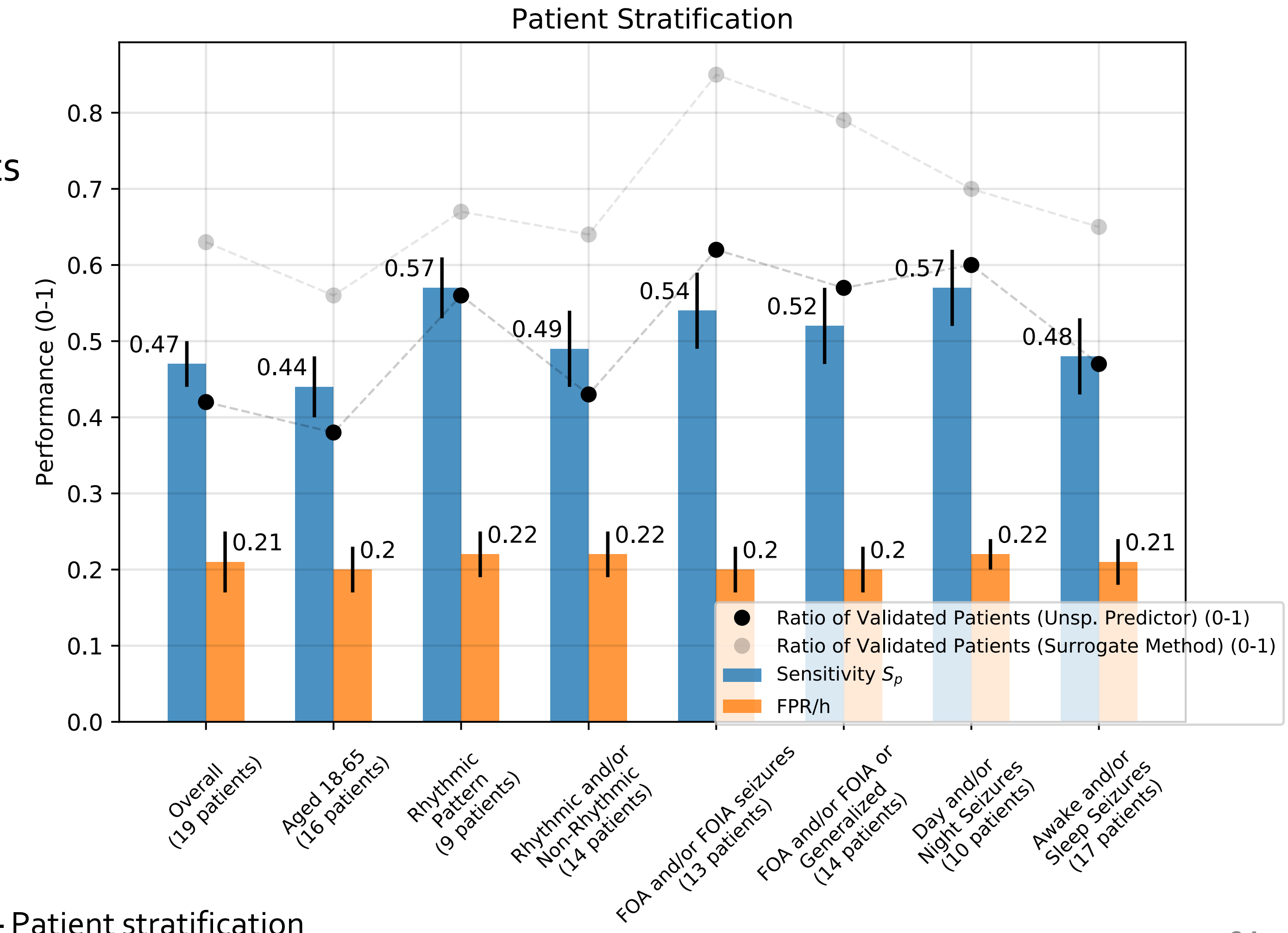


Figure 28 – Patient stratification

Discussion

A Good Example: Patient #4

- The results with 30 executions:
  - In training: $S_p = 1.00 \pm 0.00, \mathrm{FPR/h} = 0.00 \pm 0{,}00$
  - In testing: $S_p = 0.95 \pm 0.15, \mathrm{FPR/h} = 0.12 \pm 0{,}14$
- **We can trust in the models made for patient #4. Now what?**

Discussion

A Good Example: Patient #4

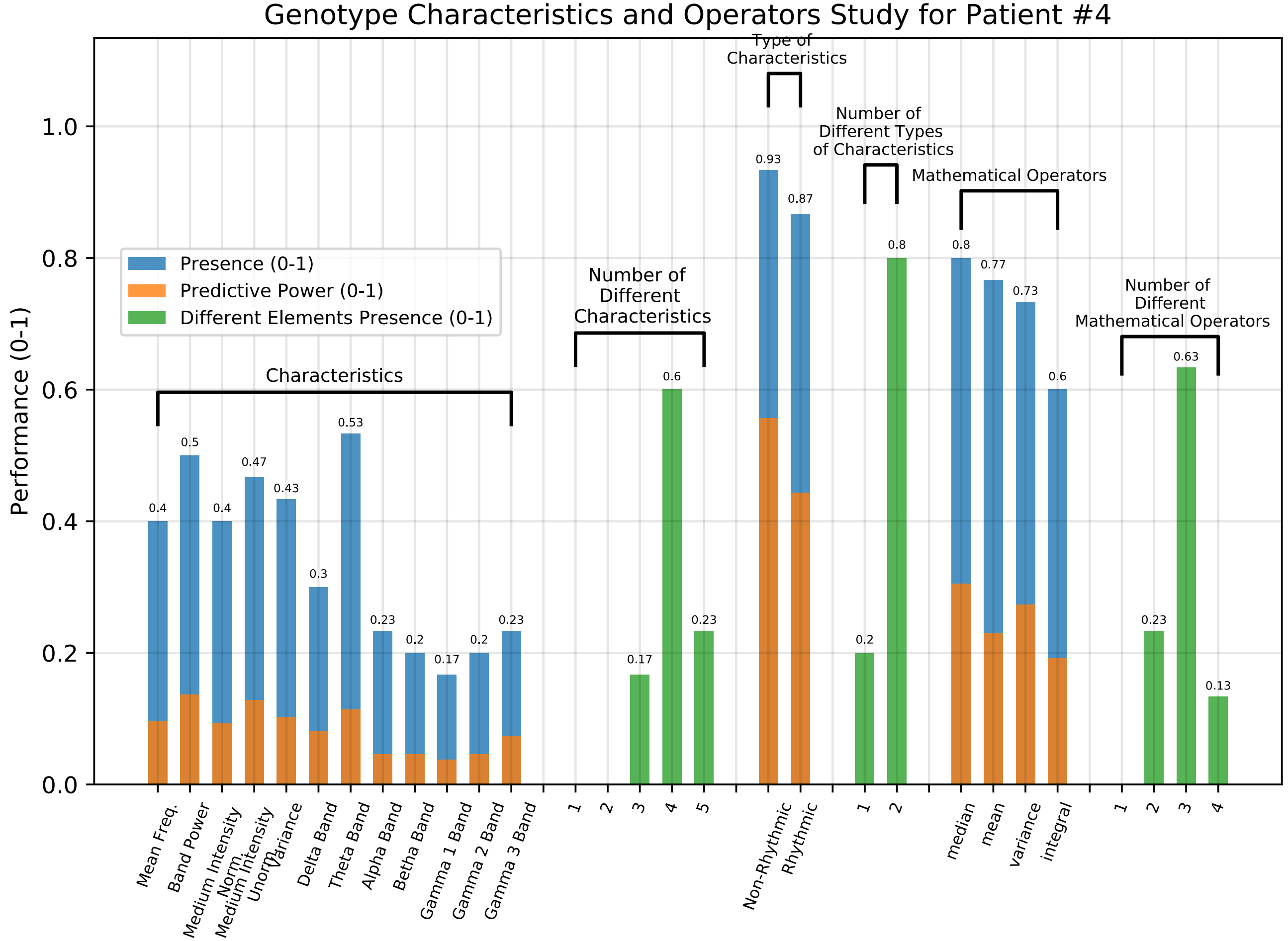


Figure 29 – Genotype characteristics and operators study

Discussion

A Good Example: Patient #4

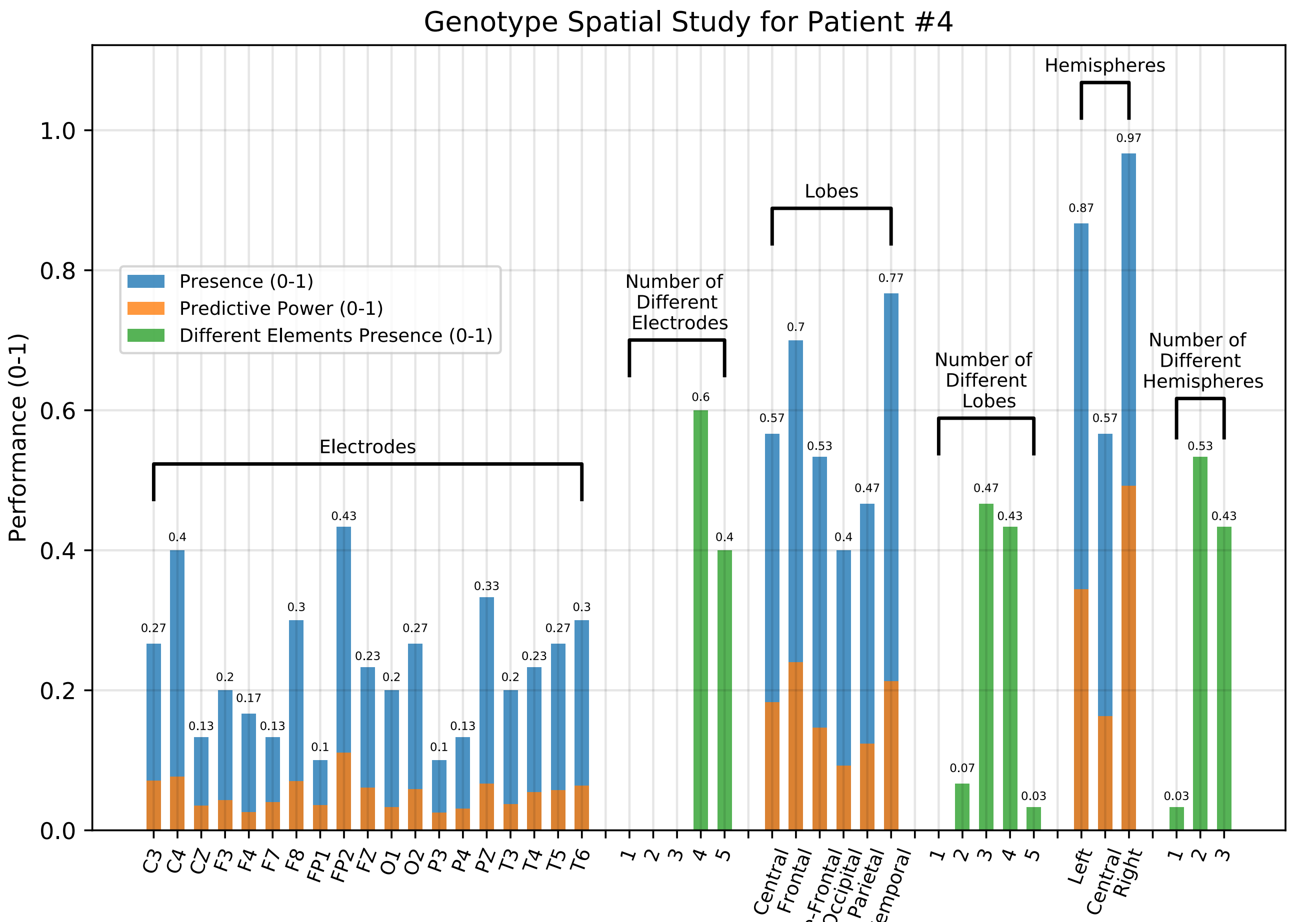


Figure 30 – Spatial study

Discussion

A Good Example: Patient #4

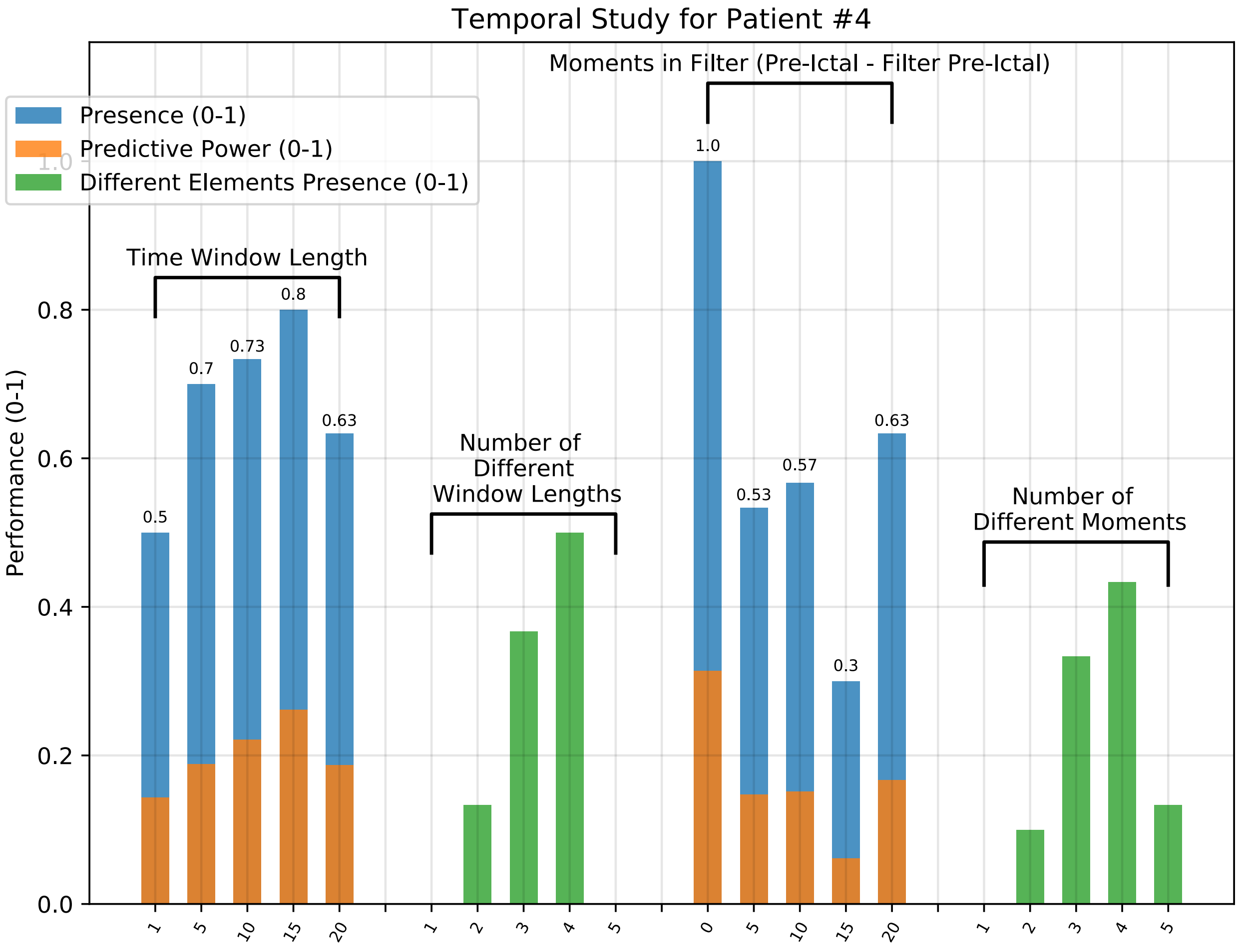


Figure 31 – Temporal study.

Discussion

A Good Example: Patient #4

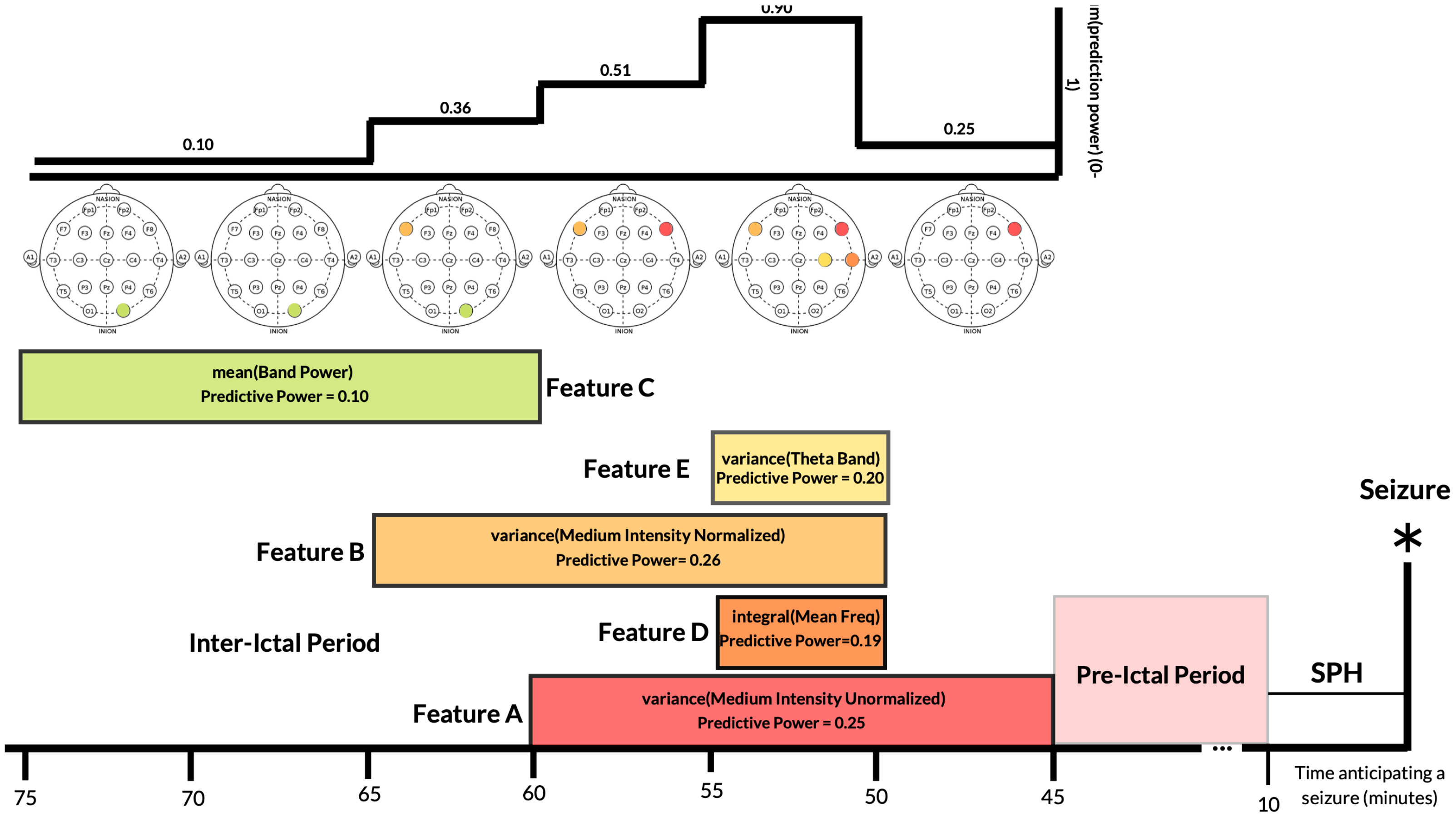


Figure 32 – A good example.

Discussion

A Good Example: Patient #4

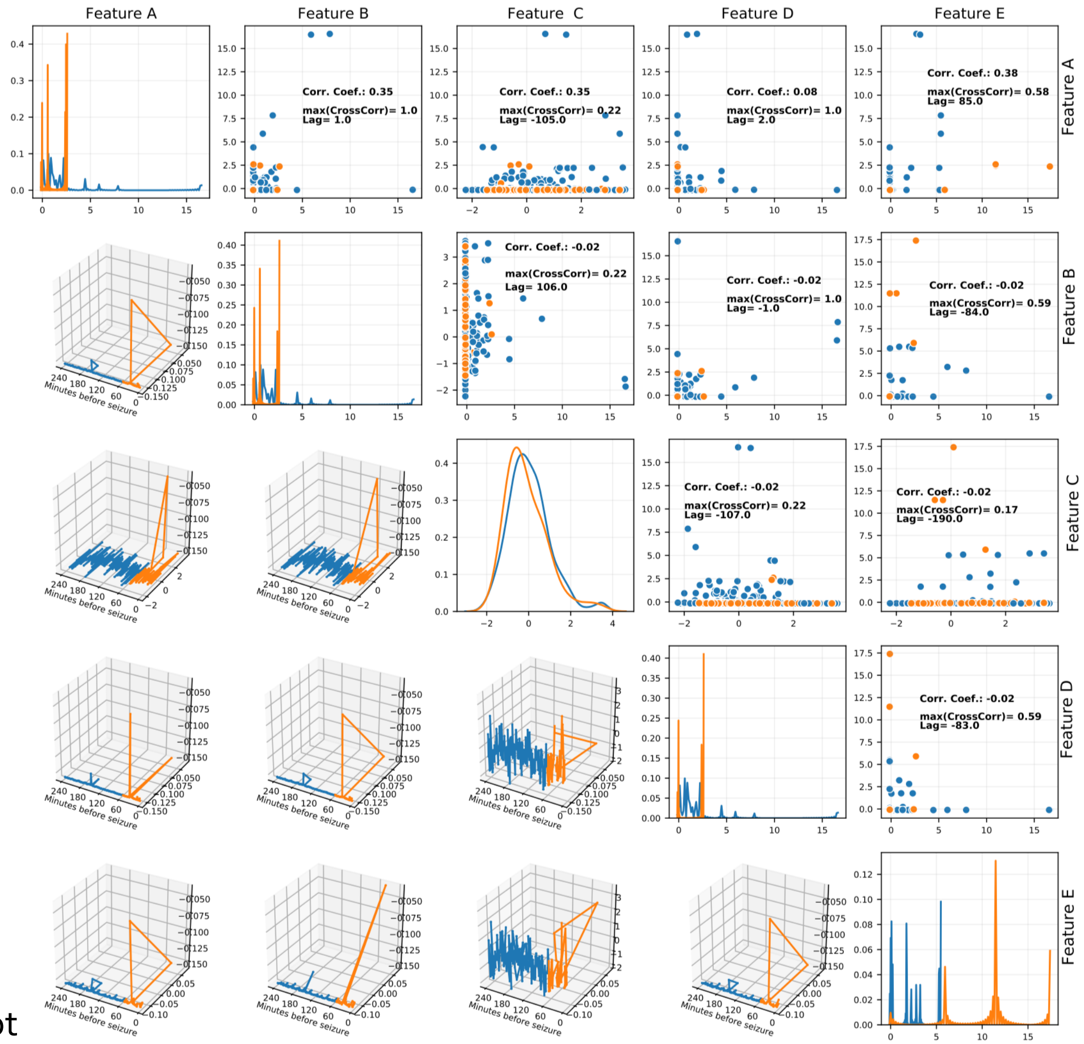


Figure 33 – A good example. Feature Plot

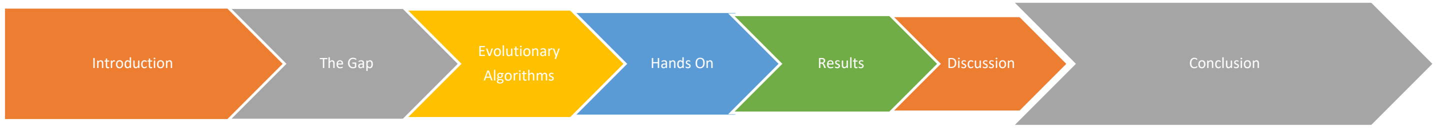
Conclusions

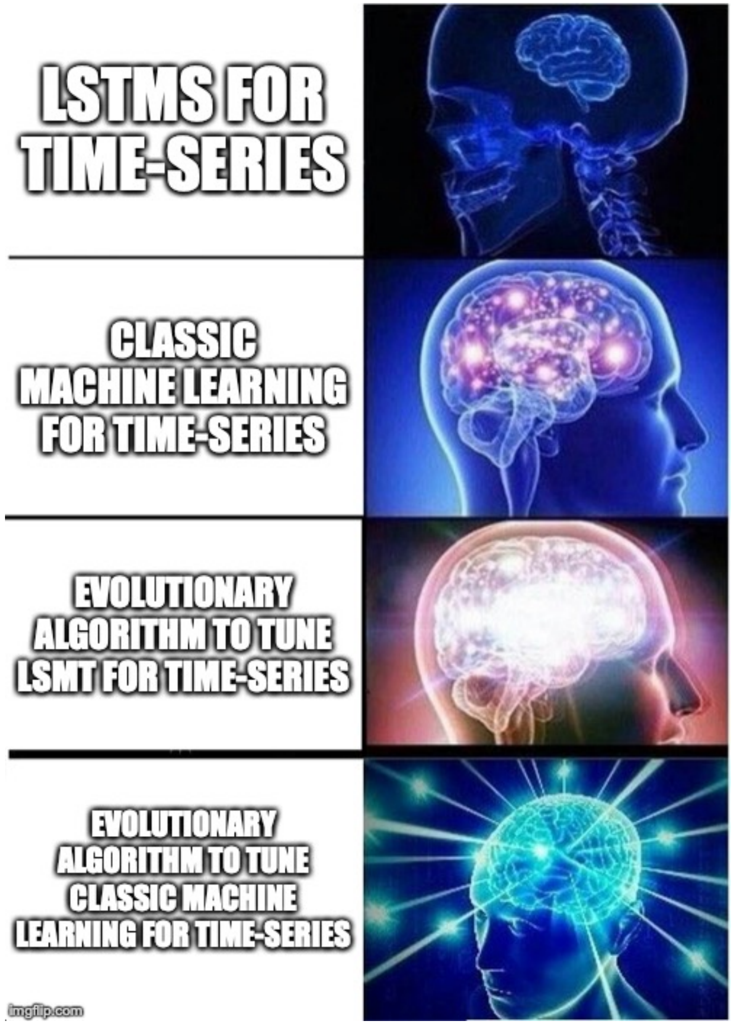
- If you want to use Machine Learning in a rare-event prediction problem, why not search for the optimal setting in a rare-event situation?

- Interpretability and simplicity are the true motivation of this approach.

- **This is just a proof-of-concept.**

Conclusions





Thank You