

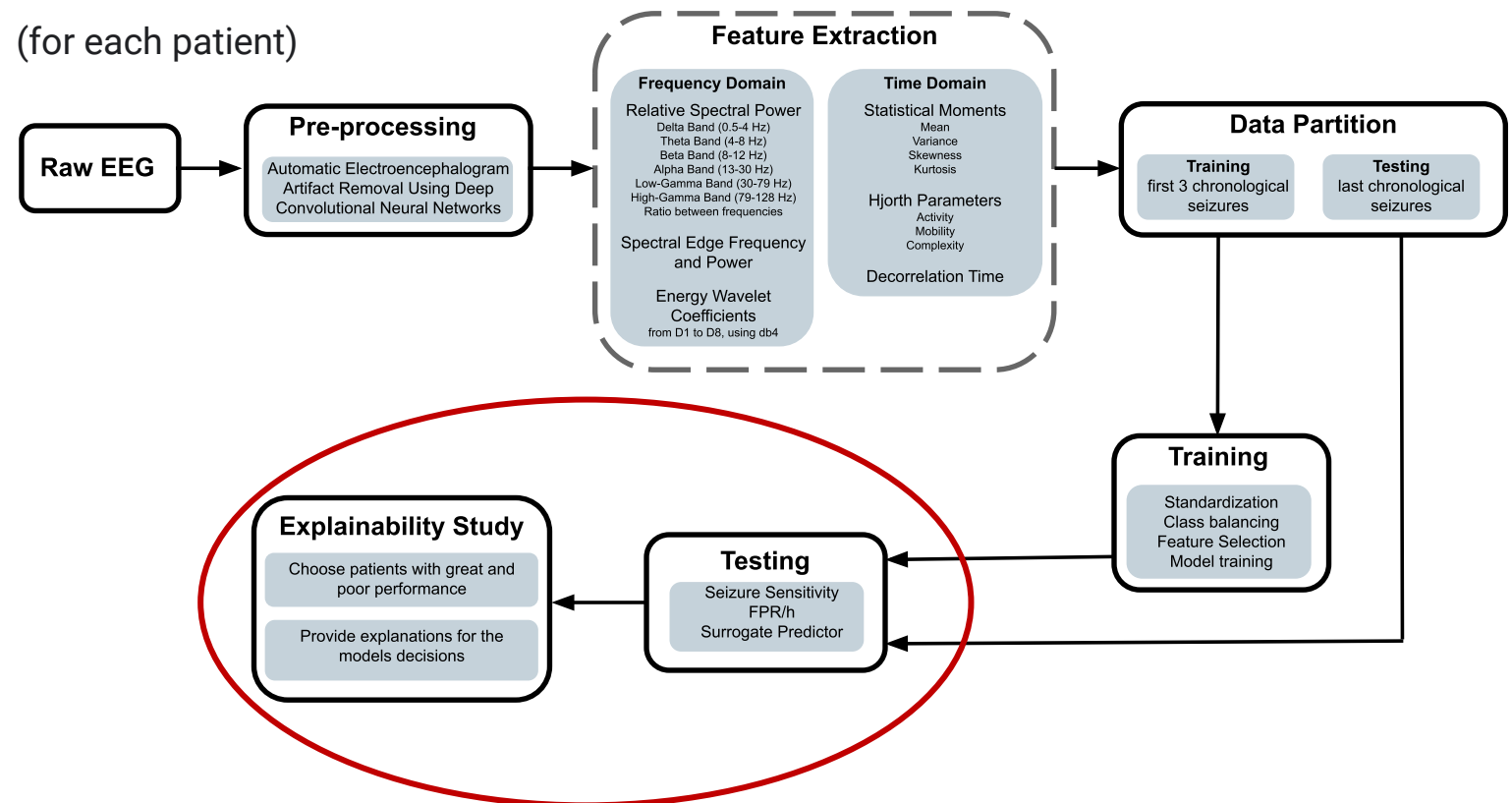
# Towards the explanation of Machine Learning methods in EEG seizure prediction

# Motivation

- Current state-of-the-art methods for electroencephalogram seizure prediction use Machine Learning (ML). However, as they deal with multidimensionality, they may create black boxes where clinicians may not have sufficient trust to use them for high-risk decisions.
- This work studies which explanations may be provided about the models to increase trust in their decisions.
- First stage: data scientists. **Second stage: clinicians**

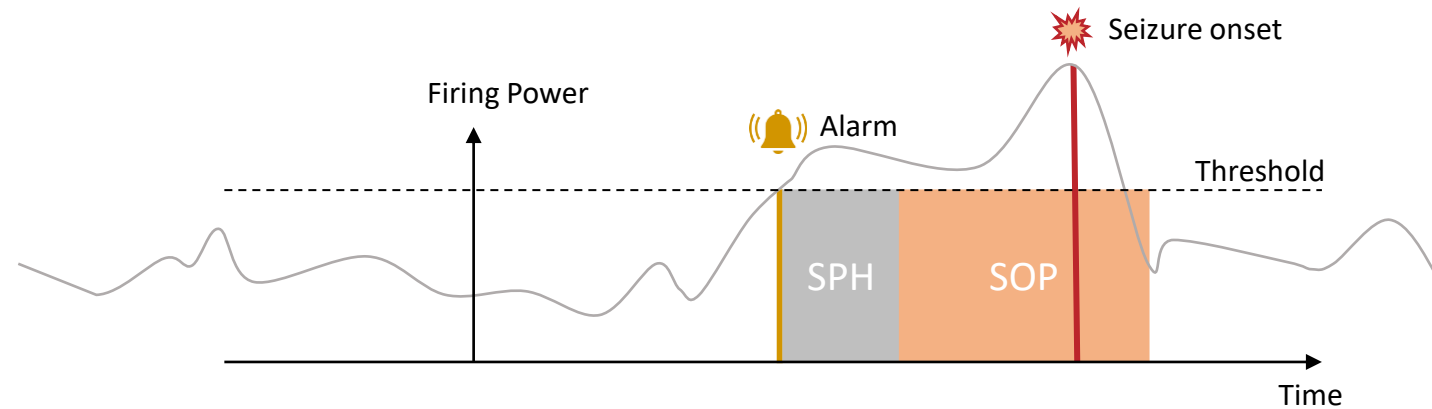
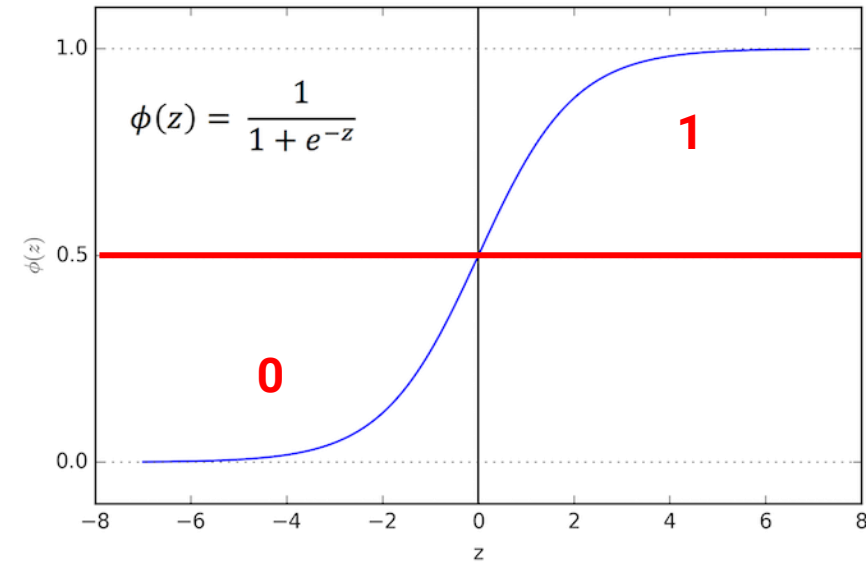
# General Methodology

- 40 Patients from EPILEPSIAE in pre-surgical monitoring
  - 104 tested seizures and 2055 hours of data



# Logistic Regression

- 1) With the first 3 seizures:
  - Find the optimal pre-ictal period within a range of values: 30-60 minutes
  - Find the optimal number of features
- 2) Label windows of signal of 5 seconds as either interictal or preictal
- 3) Feed this examples to a Machine Learning Algorithm, and obtain a model:
  - Logistic Regression
- 4) Apply the model to new data and smooth the output over time:
  - Evaluate Performance



# Patient 8902 – Performance

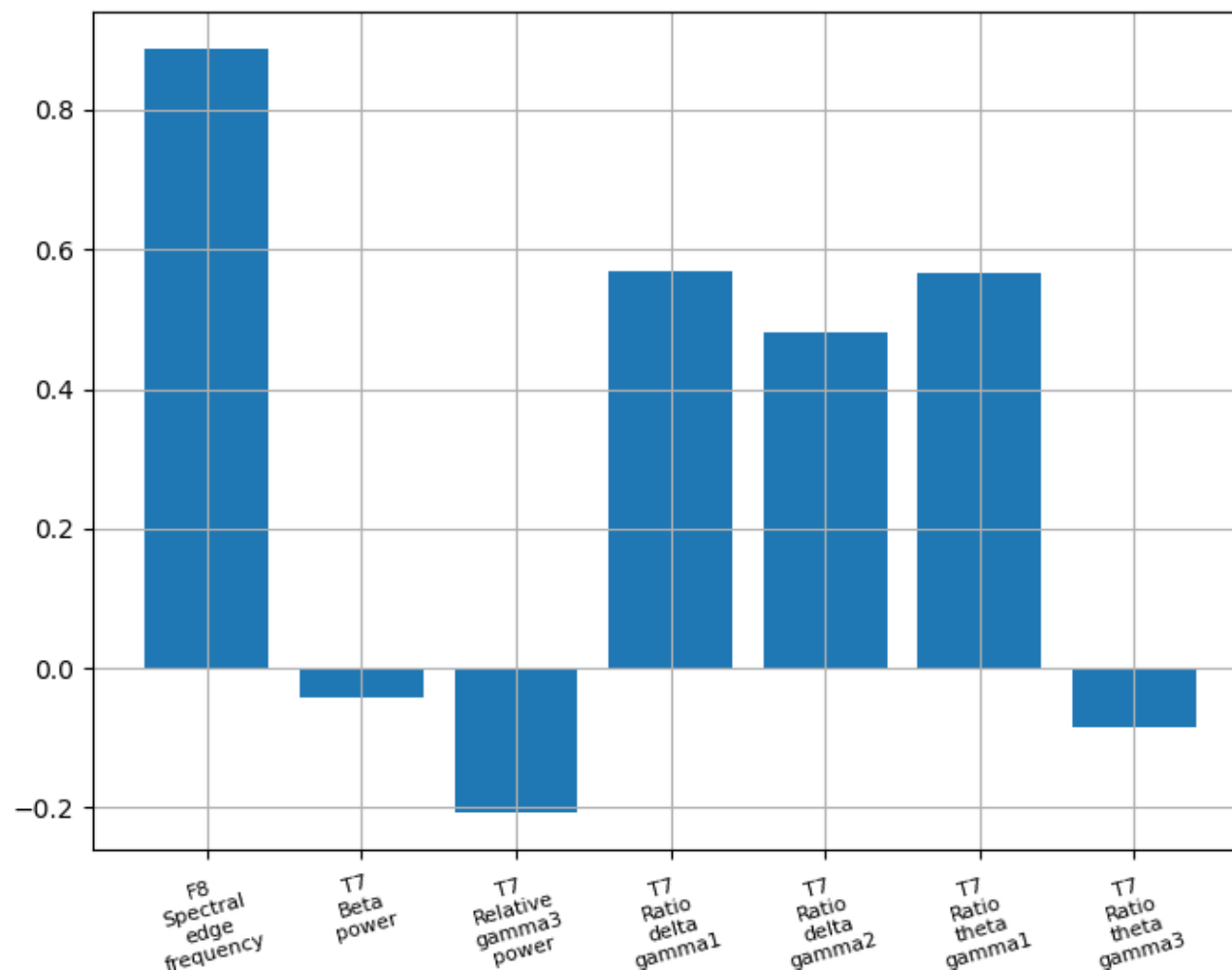
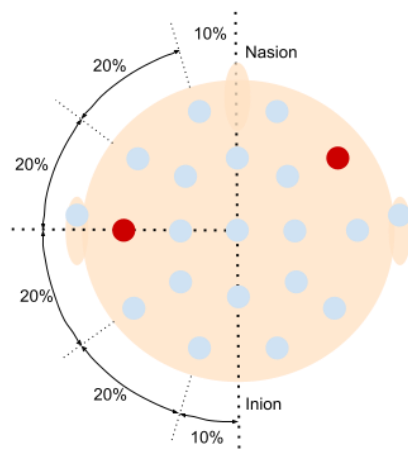
- SS:100%, FPR/h=0.11
- Logistic Regression with 7 features
- All seizures during the night. Sleep related?
- Surgery decision: not offered
- Focus localization=tpl  
Temporal polar left

	Onset	Type	Pattern	Vigilance state at onset
Training Seizures	Day 0 00:51:14	Unclassified	Rhythmic alpha waves	Awake
	Day 1 00:03:23	FOIA	Rhythmic beta waves	Awake
	Day 1 06:37:05	FOIA	Rhythmic alpha waves	Awake
Testing Seizures	Day 2 01:35:56	FOIA	Amplitude depression	Awake
	Day 2 06:10:26	FOIA	Rhythmic alpha waves	Awake

	Onset	Predicted	#False Alarms	Hours of analysed data
Testing Seizures	Seizure 4 Day 2 01:35:56	Yes	1	≈17
	Seizure 5 Day 2 06:10:26	Yes	1	≈4

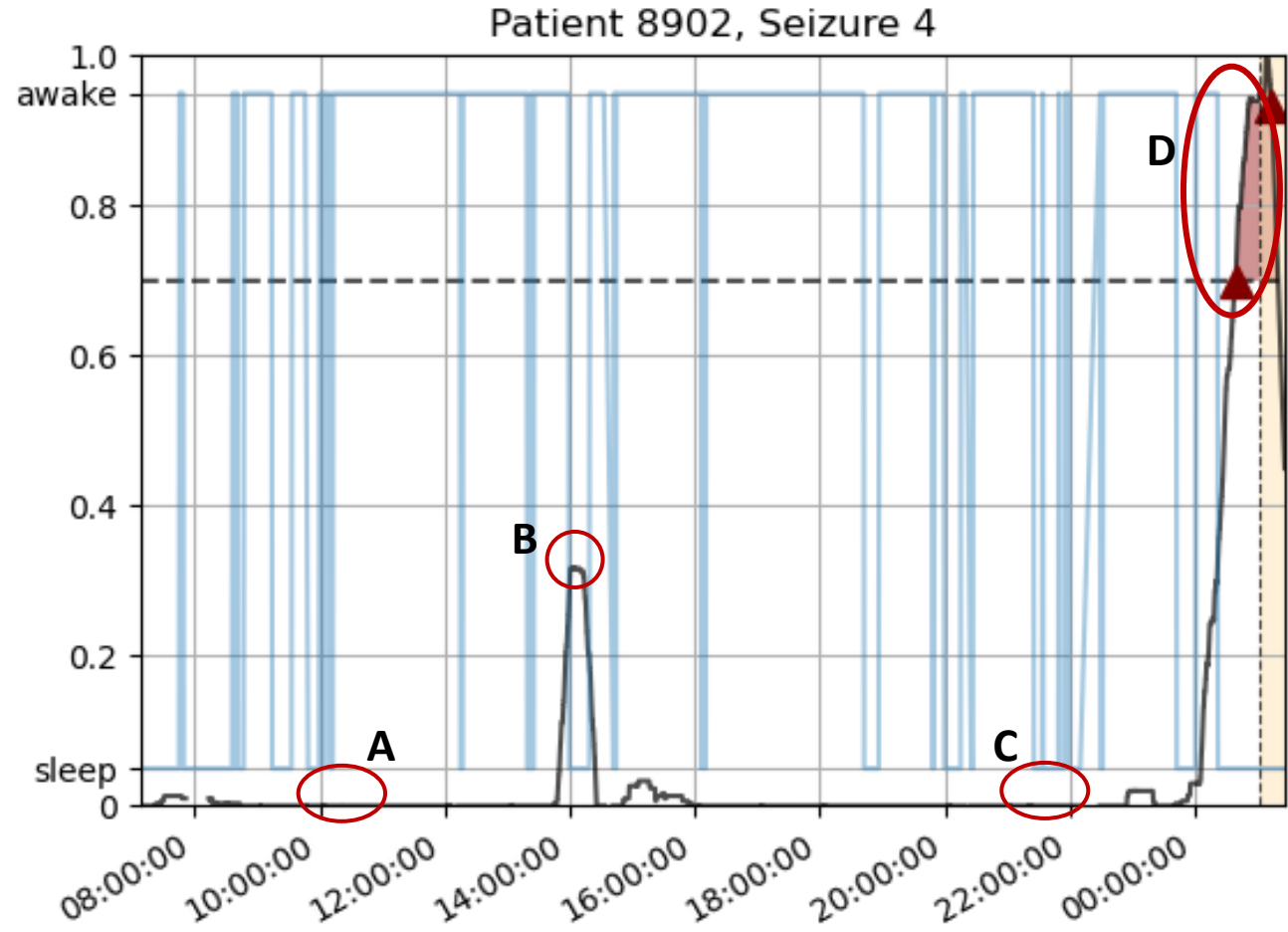
# Patient 8902 –Regression coefficients

- What can we say about these features?
- Ratios between gamma bands
- Strong influence of the gamma bands
- Electrodes: F8 and T7
- Surgery decision: not offered
- Focus localization= tpl  
Temporal polar left



# Patient 8902 – Plot over time

- Point A (10h00-10h30):
  - FP: 0
  - Morning control
- Point B (13h50-14h15):
  - Small peak
  - Related to lunch, sleep after lunch?
- Point C (21h30-22h00):
  - FP: 0
  - But interestingly, the Firing Power curve is already going down
- Point D (23h30-01h00):
  - FP started to rise at 00h00.
  - Capture the before 00h00 and the ungoing



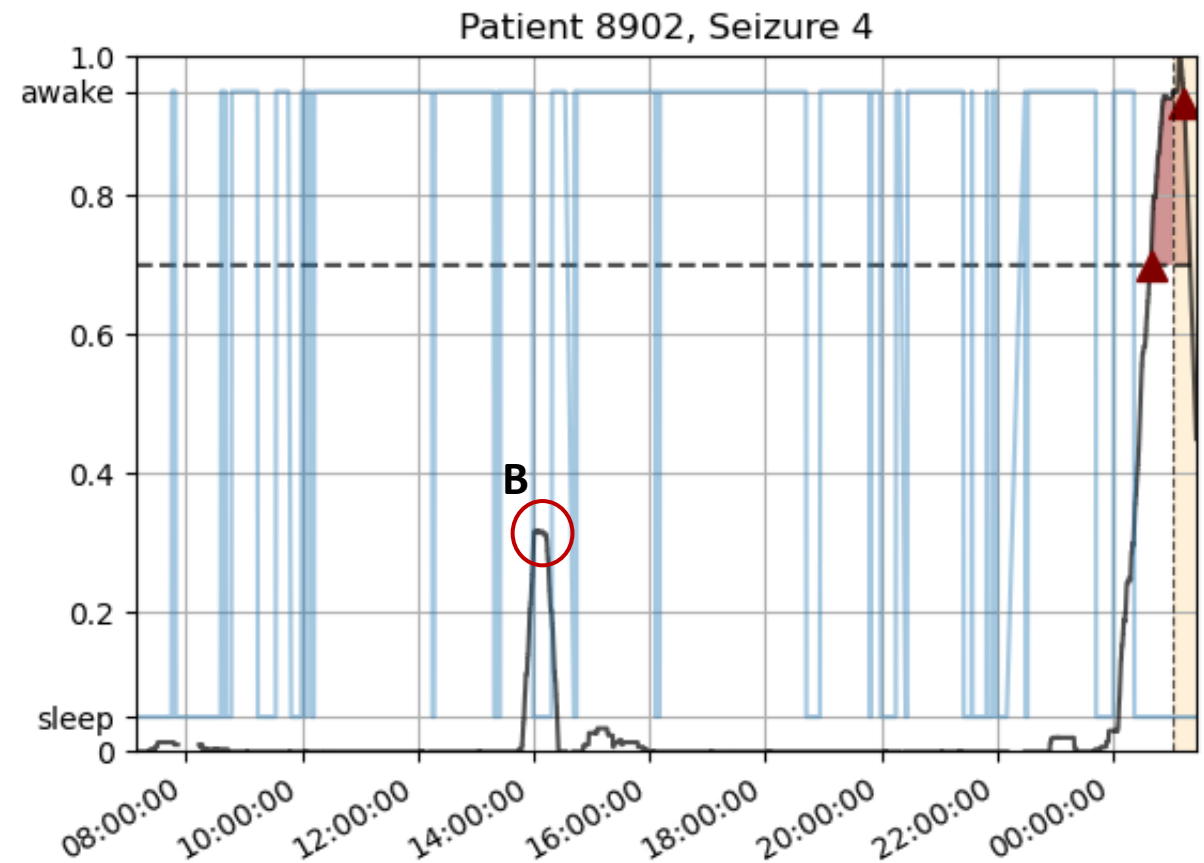
# Patient 8902 – Counterfactual explanations

- For the prediction to be different (no peak), we would have to have the following features (top 3 by order of importance):

- T7\_Beta\_power:  $\uparrow$
- T7\_Ratio\_theta\_gamma3:  $\downarrow$
- T7\_Relative\_gamma3\_power:  $\downarrow$

- Or, there was a peak because:

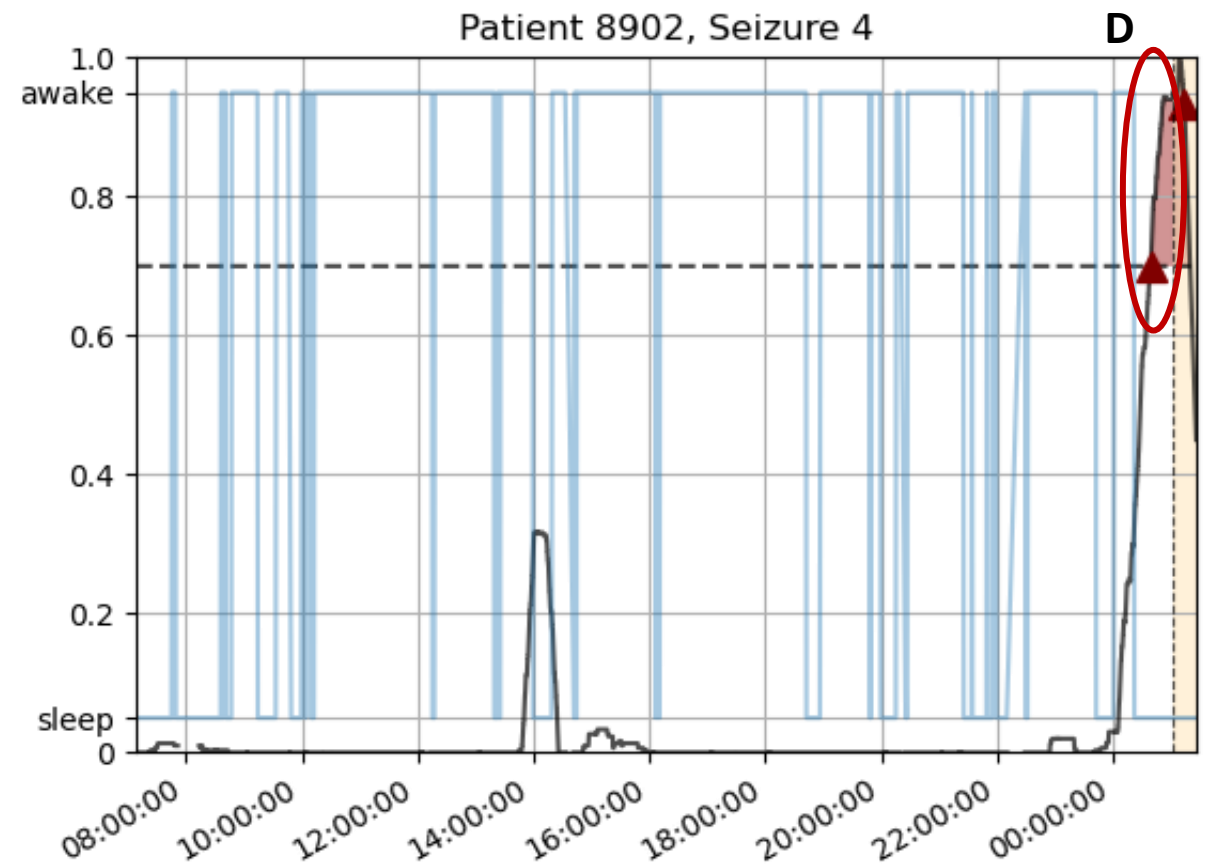
- T7\_Beta\_power was low
- T7\_Ratio\_theta\_gamma3 were high
- T7\_Relative\_gamma3\_power were high





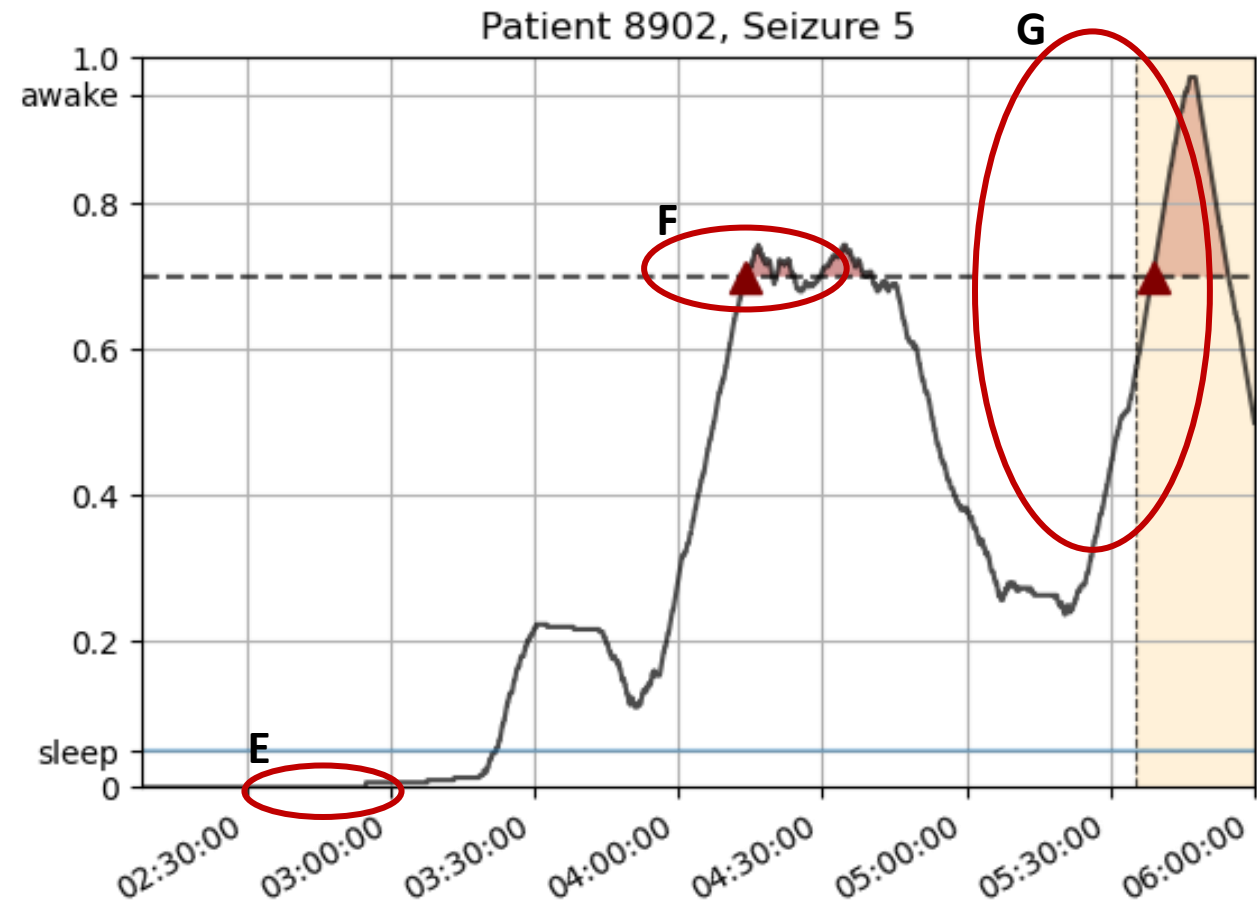
# Patient 8902 – Counterfactual explanations

- For the prediction to be different (no alarms), we would have to have the following features (top 3 by order of importance):
  - T7\_Relative\_gamma3\_power:  $\uparrow$
  - T7\_Beta\_power:  $\uparrow$
  - T7\_Ratio\_theta\_gamma3:  $\uparrow$
- Or, there was a peak because:
  - T7\_Relative\_gamma3\_power were low
  - T7\_Beta\_power were low
  - T7\_Ratio\_theta\_gamma3 were low



# Patient 8902 – Over time

- Point E (02h30-03h00):
  - FP=0
  - Control
- Point F (04h00-04h40):
  - A false alarm
- Point G (05h00-05h45):
  - A control part
  - And then FP rises and an alarm is correctly triggered



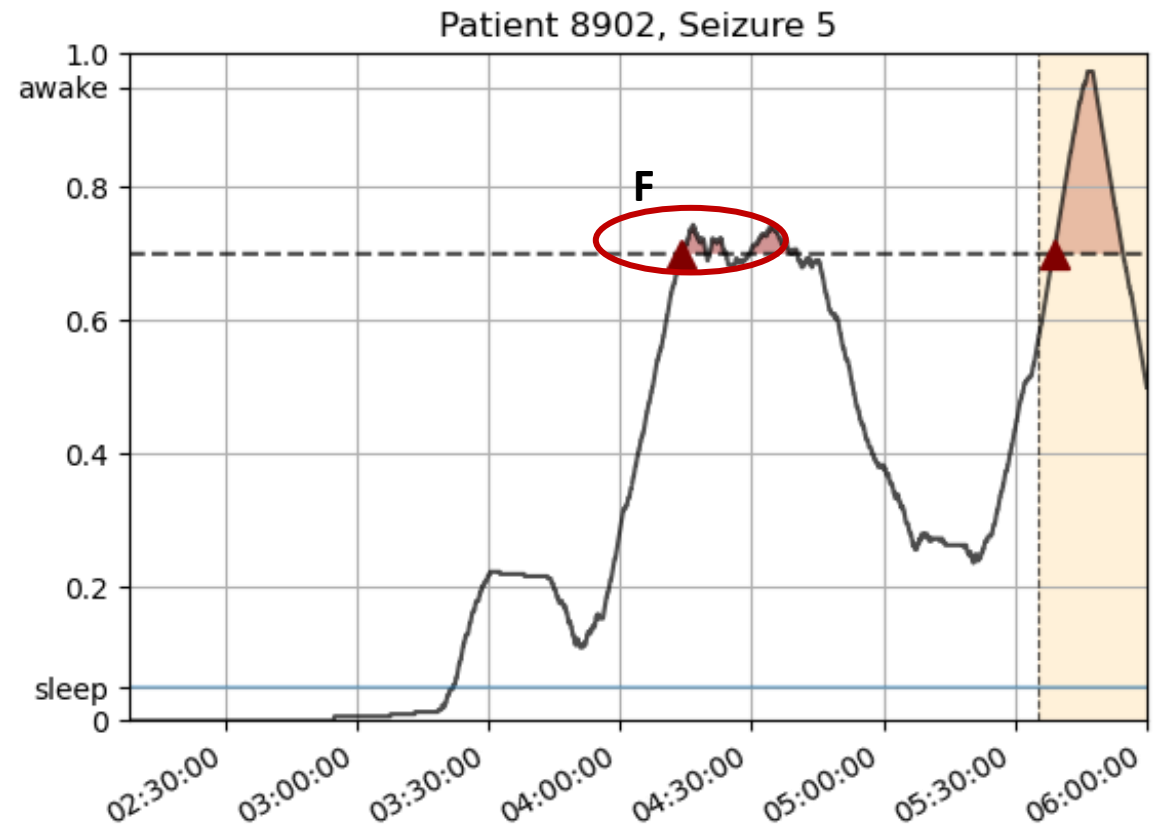
# Patient 8902 – Counterfactual explanations

- For the prediction to be different (no peak), we would have to have the following features (top 3 by order of importance):

- F8\_Spectral\_edge\_frequency: ↓
- T7\_Ratio\_delta\_gamma1: ↓
- T7\_Ratio\_theta\_gamma1: ↓

- Or, there was a peak because:

- F8\_Spectral\_edge\_frequency were high
- T7\_Ratio\_delta\_gamma1 were high
- T7\_Ratio\_theta\_gamma1 were high



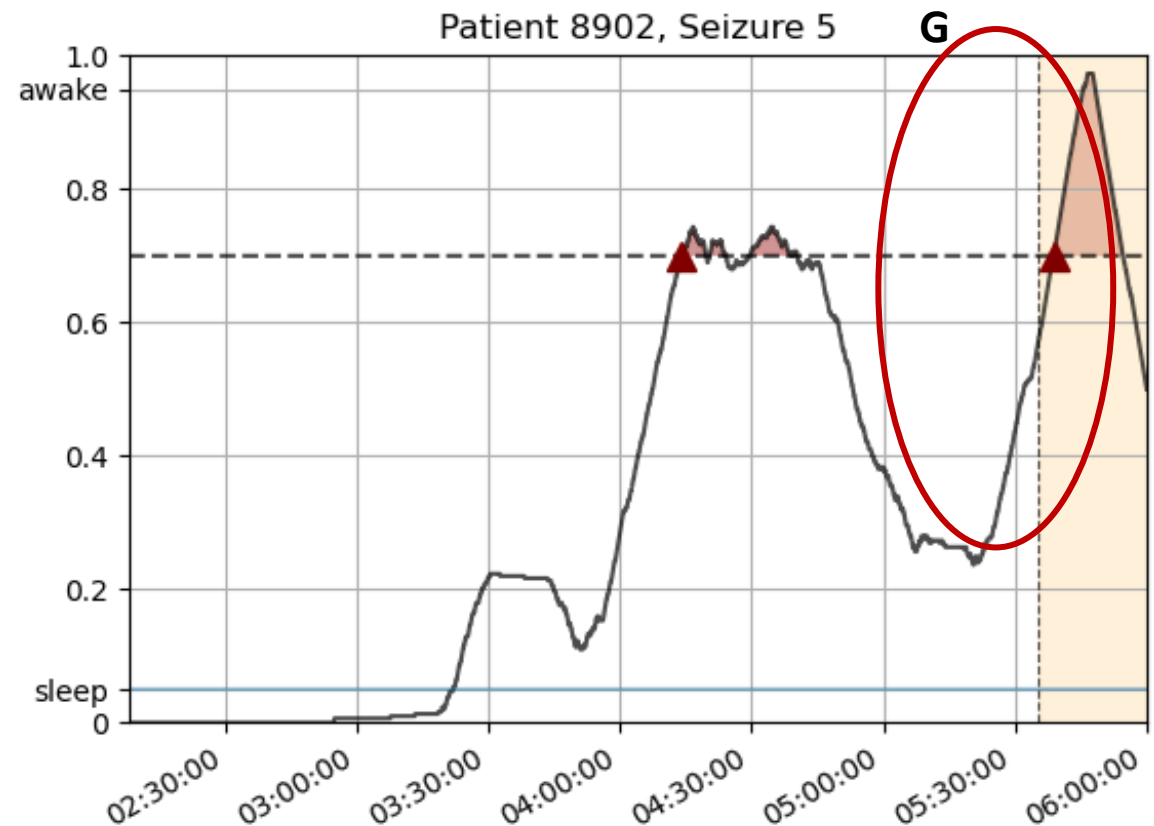
# Patient 8902 – Counterfactual explanations

- For the prediction to be different (no peak), we would have to change the following features (top 3 by order of importance):

- T7\_Ratio\_theta\_gamma3: ↓
- T7\_Relative\_gamma3\_power: ↑
- F8\_Spectral\_edge\_frequency: ↓

- Or, there was a peak because:

- T7\_Ratio\_theta\_gamma3 were high
- T7\_Relative\_gamma3\_power were low
- F8\_Spectral\_edge\_frequency were high



# Patient 93402 – Overall performance

- SS:100%, FPR/h=0.50
- Above chance but FPR/h too high
- Logistic Regression with 3 features
- Surgery decision: not offered
- Focus localization= tpl  
Temporal polar left

	Onset	Type	Pattern	Vigilance state at onset
Training Seizures	Day 0 22:17:50	FBTC	Rhythmic theta waves	Non-REM II
	Day 1 10:21:34	FOIA	Rhythmic theta waves	Non-REM II
	Day 1 23:20:24	FOIA	Rhythmic theta waves	Non-REM II
Testing Seizures	Day 4 00:59:09	Unclassified	Rhythmic theta waves	Non-REM II
	Day 4 06:26:26	Unclassified	Rhythmic theta waves	Non-REM II

	Onset	Predicted	#False Alarms	Hours of analysed data
Testing Seizures	Seizure 4 Day 4 00:59:09	Yes	15	≈49
	Seizure 5 Day 4 06:26:26	Yes	6	≈5

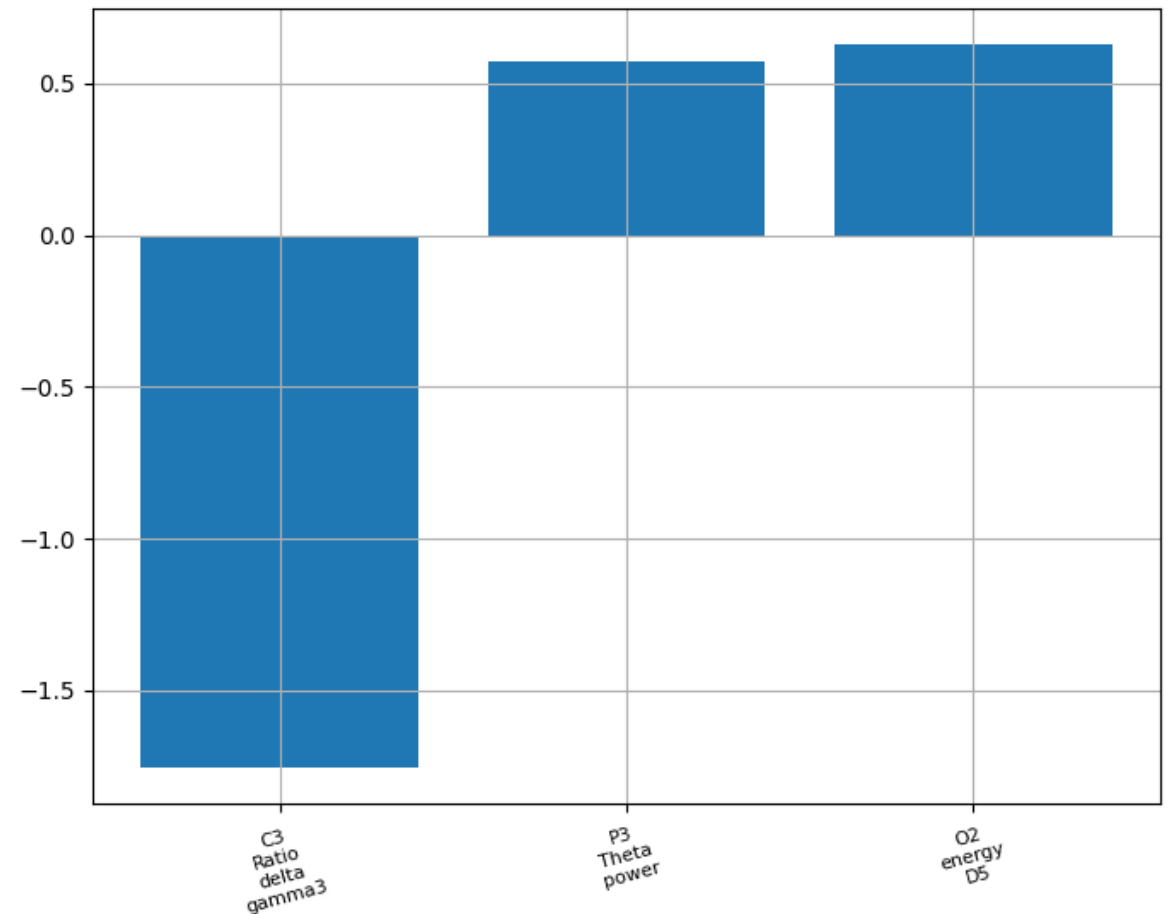
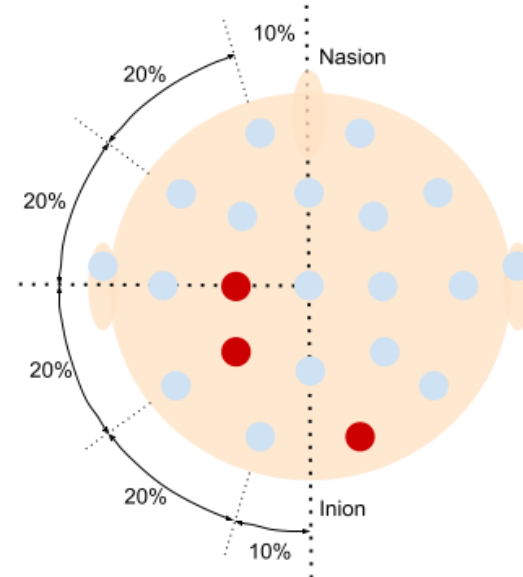
# Patient 93402 – Analysing regression coefficients

- What can we say about these features?

- Theta band power, which makes a lot of sense

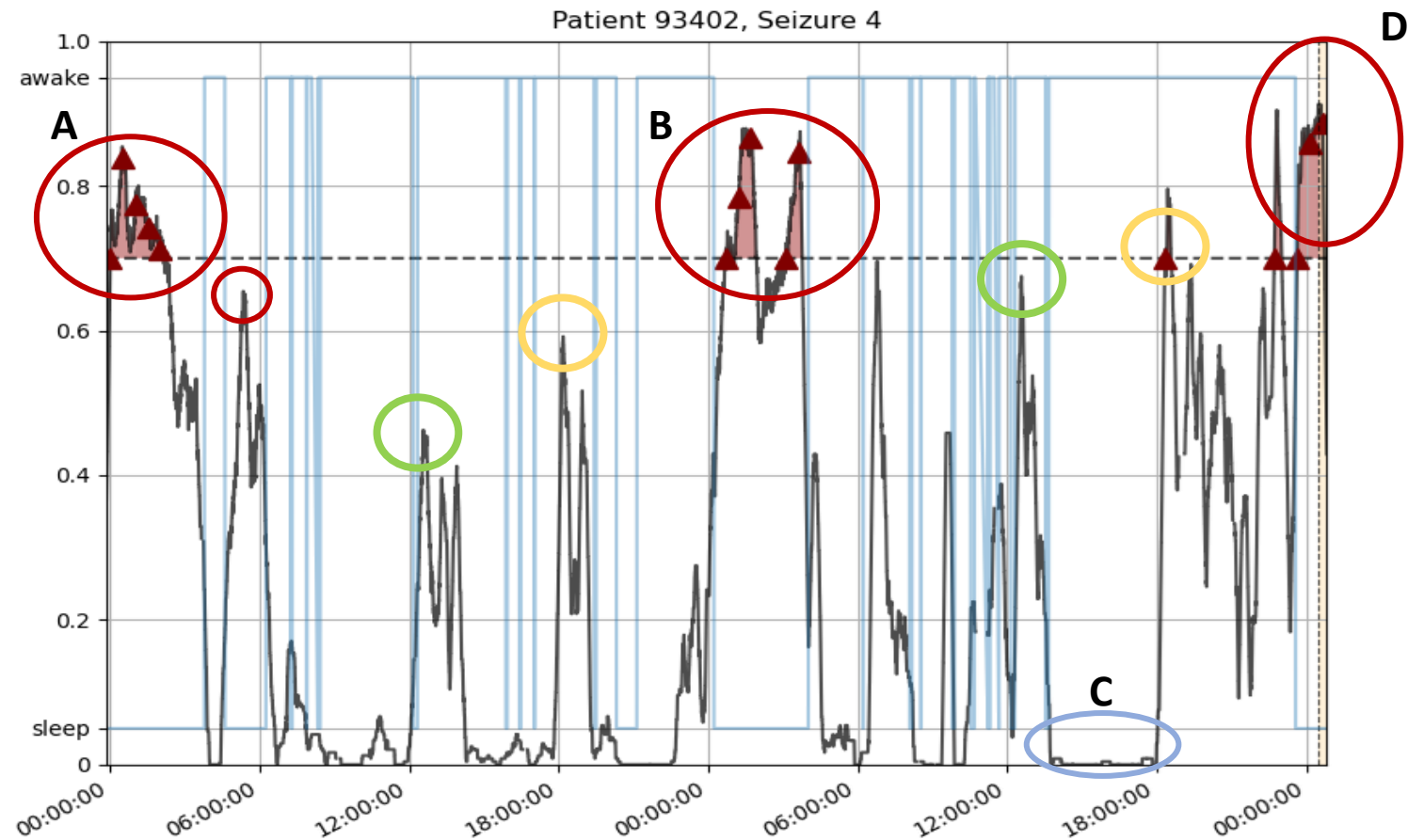
- Influence of the gamma bands

- Electrodes: C3 , P3, and O2.
  - C3 and P3 make sense. Near focus



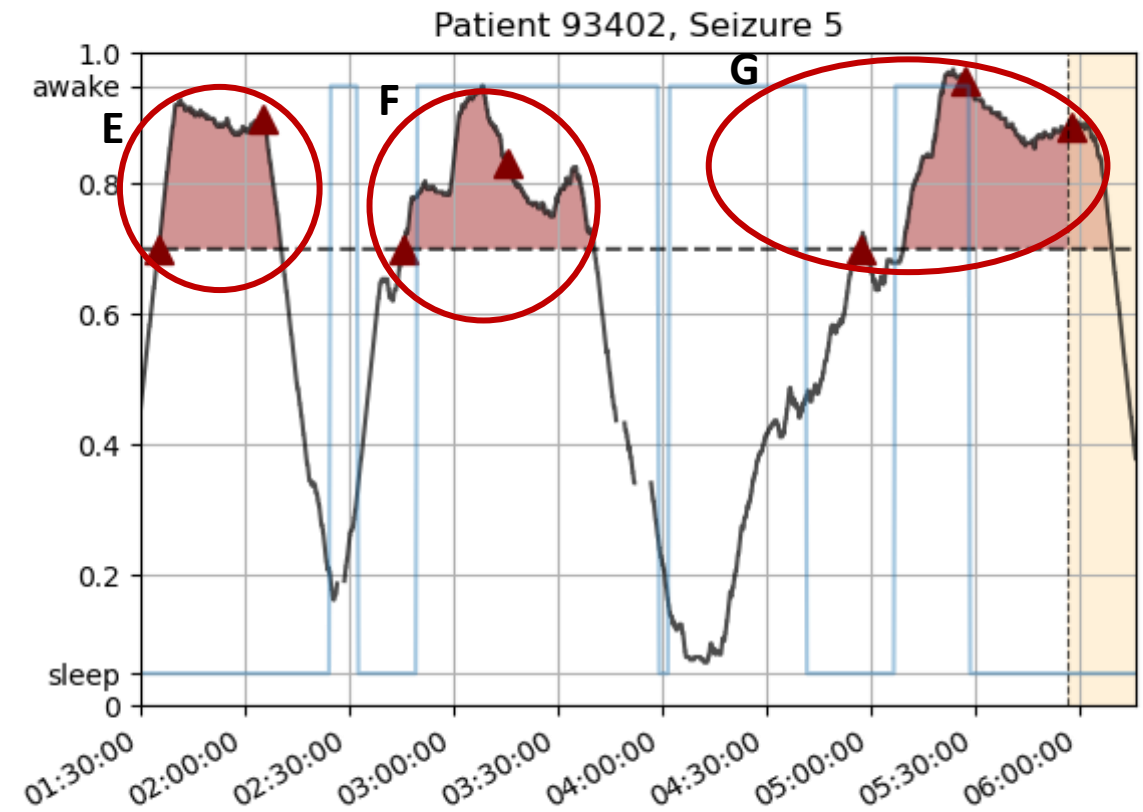
# Patient 93402 – Over time

- Cluster A (00h00-04h00, Day 0):
  - A series of false alarms
- Cluster B (00h00-06h00, Day 1):
  - A series of false alarms
- Point D (23h20-00h35, Day 2):
  - False alarms but seizure predicted
- Green and Yellow circles
- Point C (14h00-18h00, Day 2):
  - FP=0
  - Control method
- Circadian cycles influence? Sleep stages?
- Seizure susceptibility?



# Patient 93402 – Over time

- Point E (01h30-02h00):
  - Two false alarms
- Point F (02h30-03h30):
  - Two false alarms
- Point G (4h00-06h00):
  - From a decreasing and low FP
  - To a rising FP, false alarms, and predicted Seizure





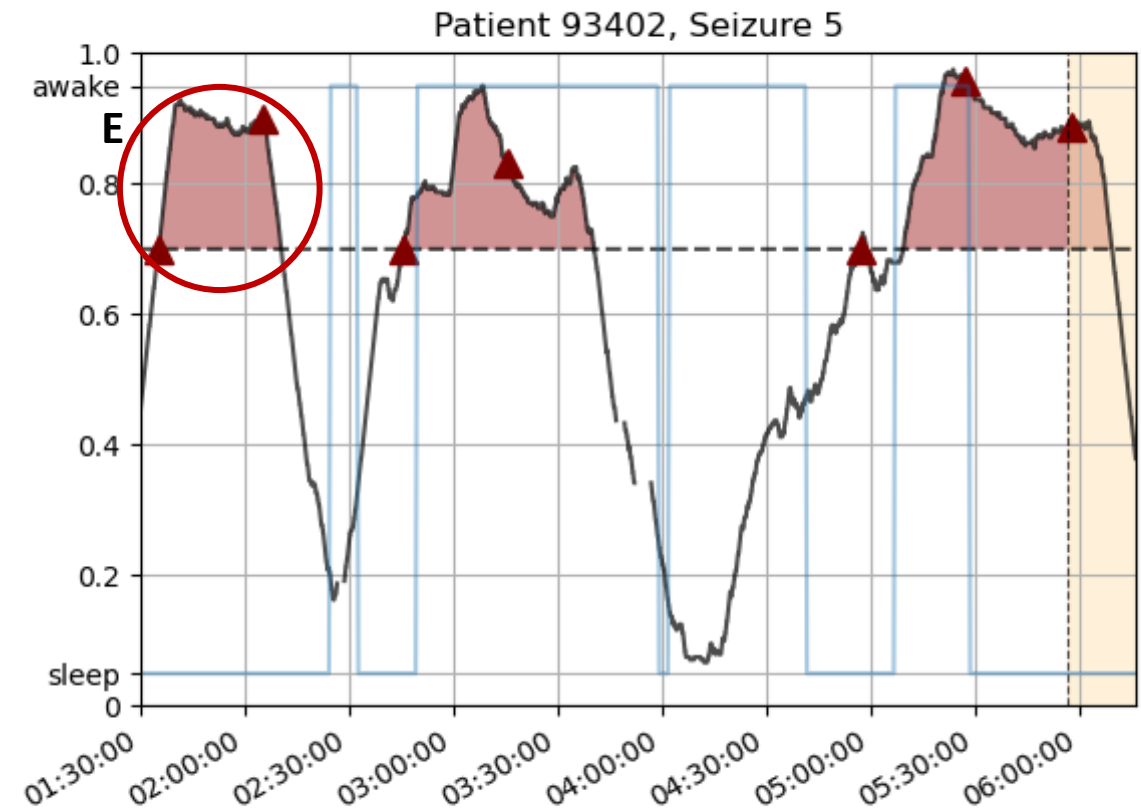
# Patient 93402 – Counterfactual explanations

- For the prediction to be different (no peak), we would have to have the following features (top 3 by order of importance):

- C3\_Ratio\_delta\_gamma3:  $\uparrow$
- O2\_energy\_D5:  $\downarrow$
- P3\_Theta\_power  $\downarrow$

- Or, there was a peak because:

- C3\_Ratio\_delta\_gamma3 were low
- O2\_energy\_D5 were high
- P3\_Theta\_power were high



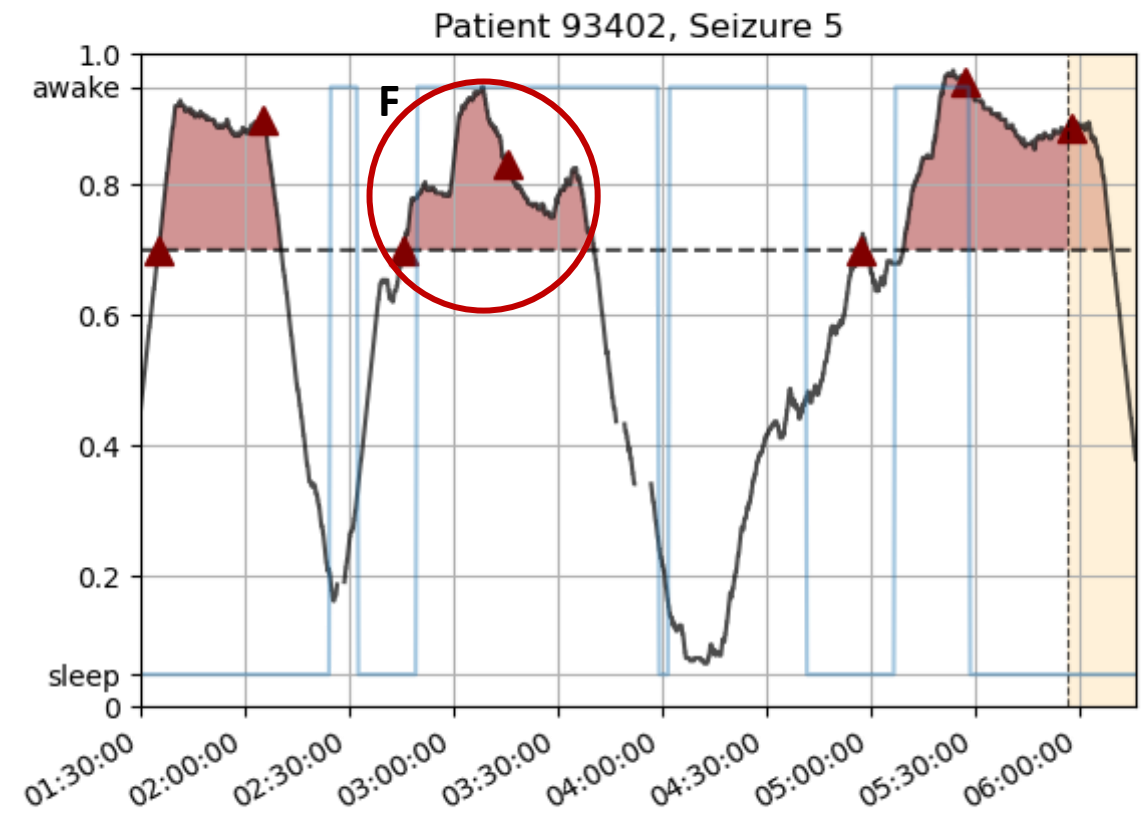
# Patient 93402 – Counterfactual explanations

- For the prediction to be different (no peak), we would have to have the following features (top 3 by order of importance):

- C3\_Ratio\_delta\_gamma3:  $\uparrow$
- O2\_energy\_D5:  $\downarrow$
- P3\_Theta\_power:  $\downarrow$

- Or, there was a peak because:

- C3\_Ratio\_delta\_gamma3 was low
- O2\_energy\_D5: was high
- P3\_Theta\_power was high



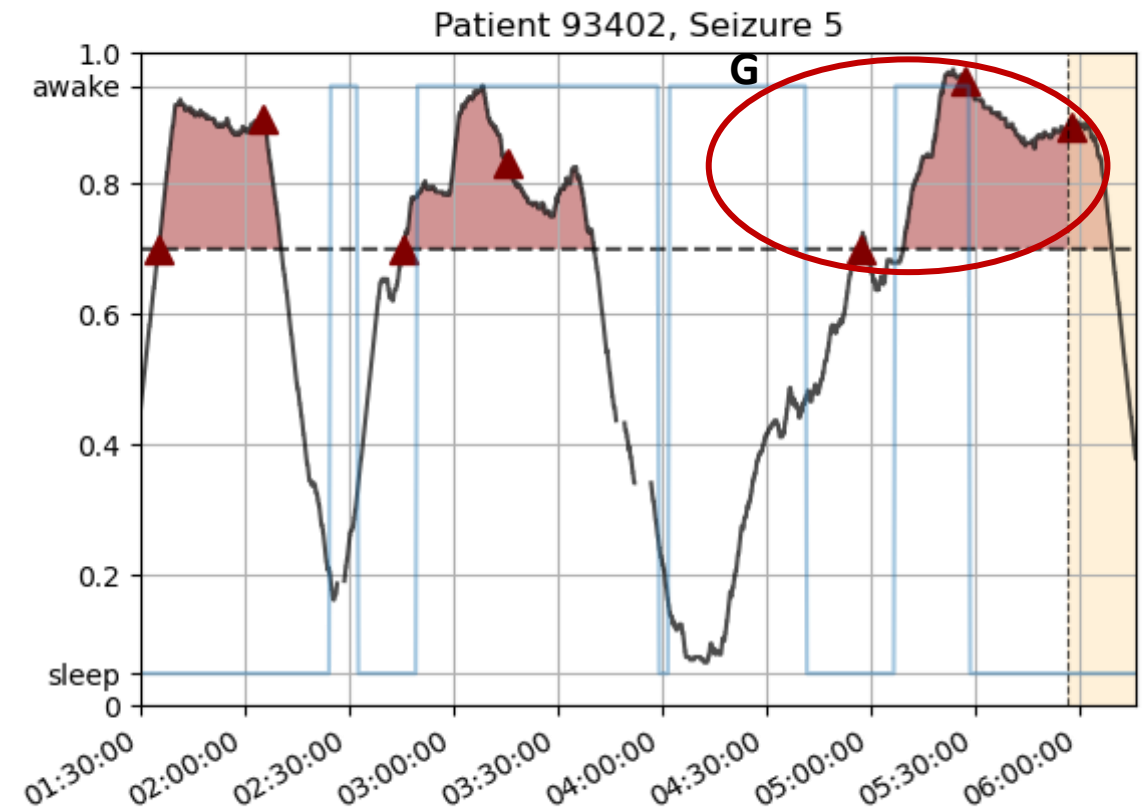
# Patient 93402 – Counterfactual explanations

- For the prediction to be different (no peak), we would have to have the following features (top 3 by order of importance):

- C3\_Ratio\_delta\_gamma3:  $\uparrow$
- O2\_energy\_D5:  $\downarrow$
- P3\_Theta\_power:  $\downarrow$

- Or, there was a peak because:

- C3\_Ratio\_delta\_gamma3 was low
- O2\_energy\_D5 was high
- P3\_Theta\_power was high



# Patient 402 – Overall performance

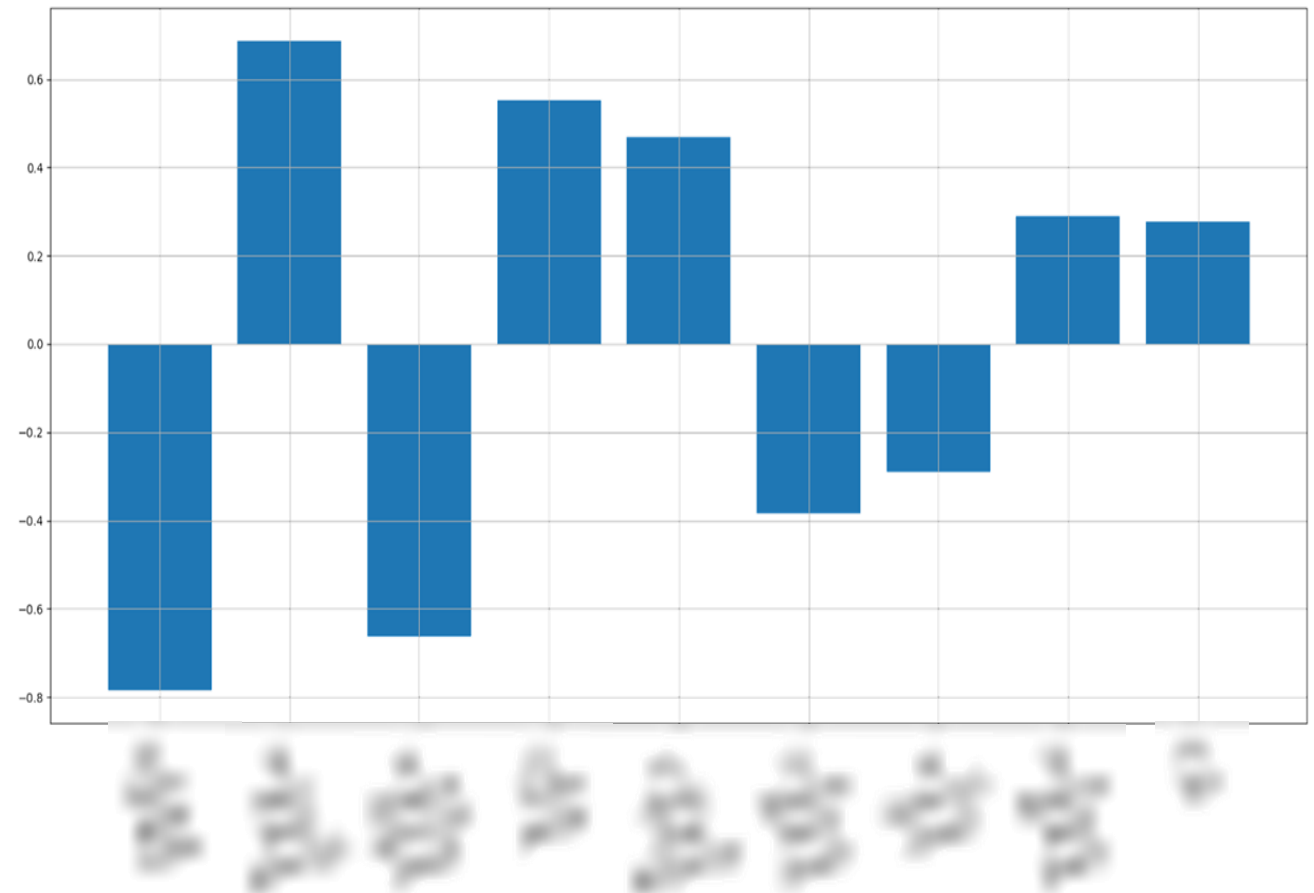
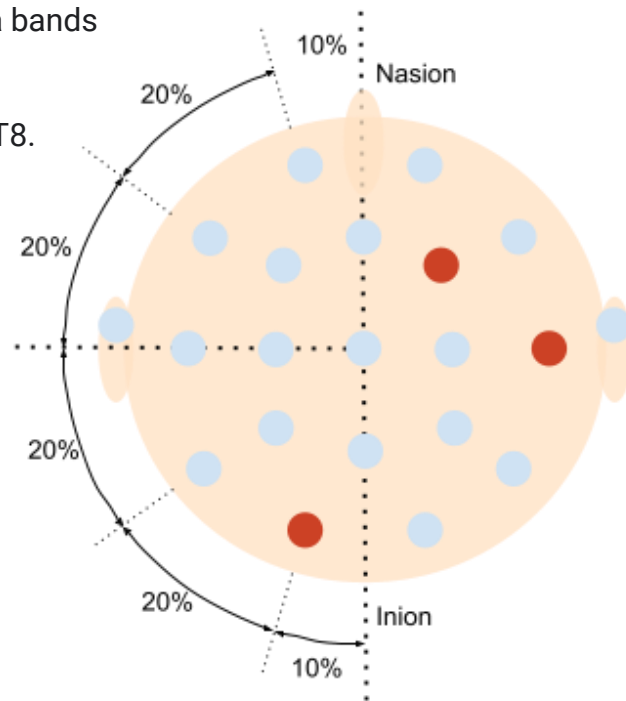
- SS:0%, FPR/h=0.00
- No single alarm was raised
- Logistic Regression with 20 features
- Surgery decision: invasive
- Focus localization= t-r, t-l  
Temporal right, temporal left

	Onset	Type	Pattern	Vigilance state at onset
Training Seizures	Day 0 23:45:26	FOIA	Rhythmic theta waves	Awake
	Day 2 22:27:34	FBTC	Rhythmic theta waves	Awake
	Day 3 03:13:30	FOIA	Rhythmic theta waves	Awake
Testing Seizures	Day 3 09:53:21	FBTC	Rhythmic theta waves	Awake
	Day 4 09:57:27	FOIA	Rhythmic theta waves	Awake

	Onset	Predicted	#False Alarms	Hours of analysed data
Testing Seizures	Day 3 09:53:21	No	0	≈6
	Day 4 09:57:27	No	0	≈24

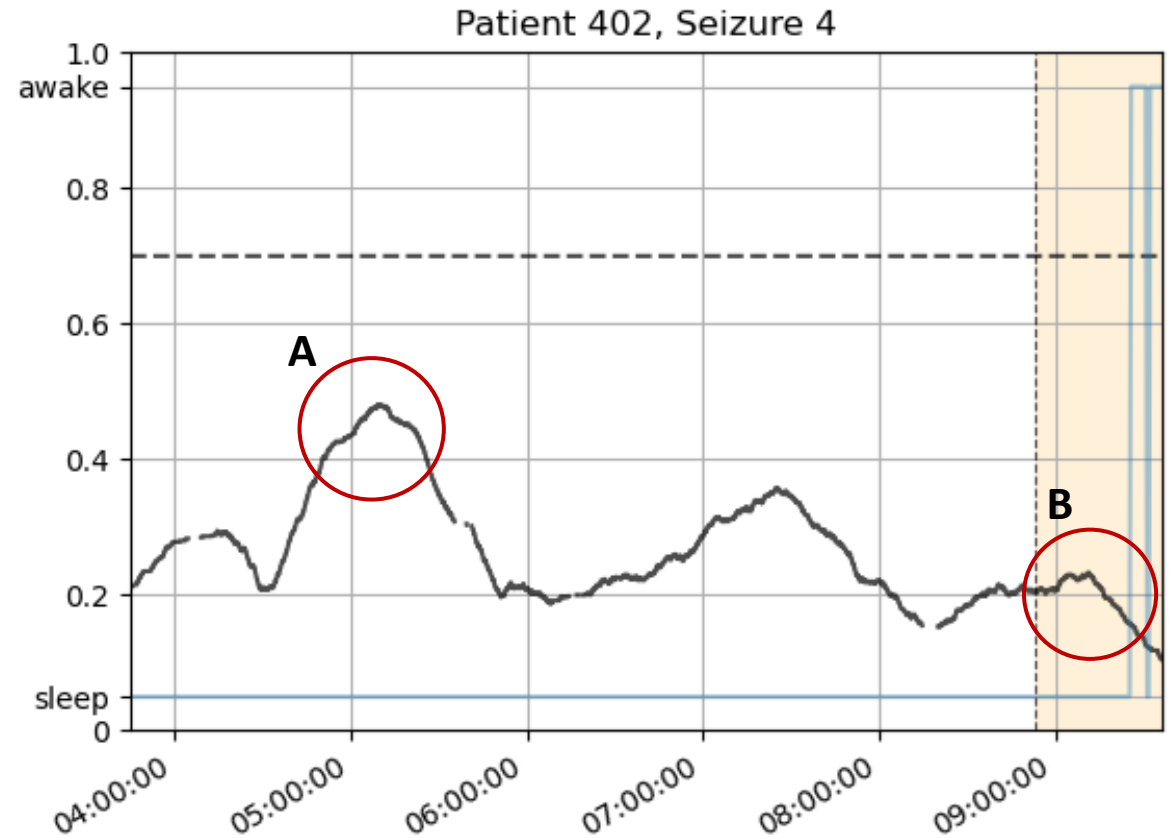
# Patient 402 – Analysing regression coefficients

- What can we say about these features?
- Different gama bands, kurtosis, activity, energy
- Influence of the gamma bands
- Electrodes: O1, F4 and T8.
  - Do these make sense?



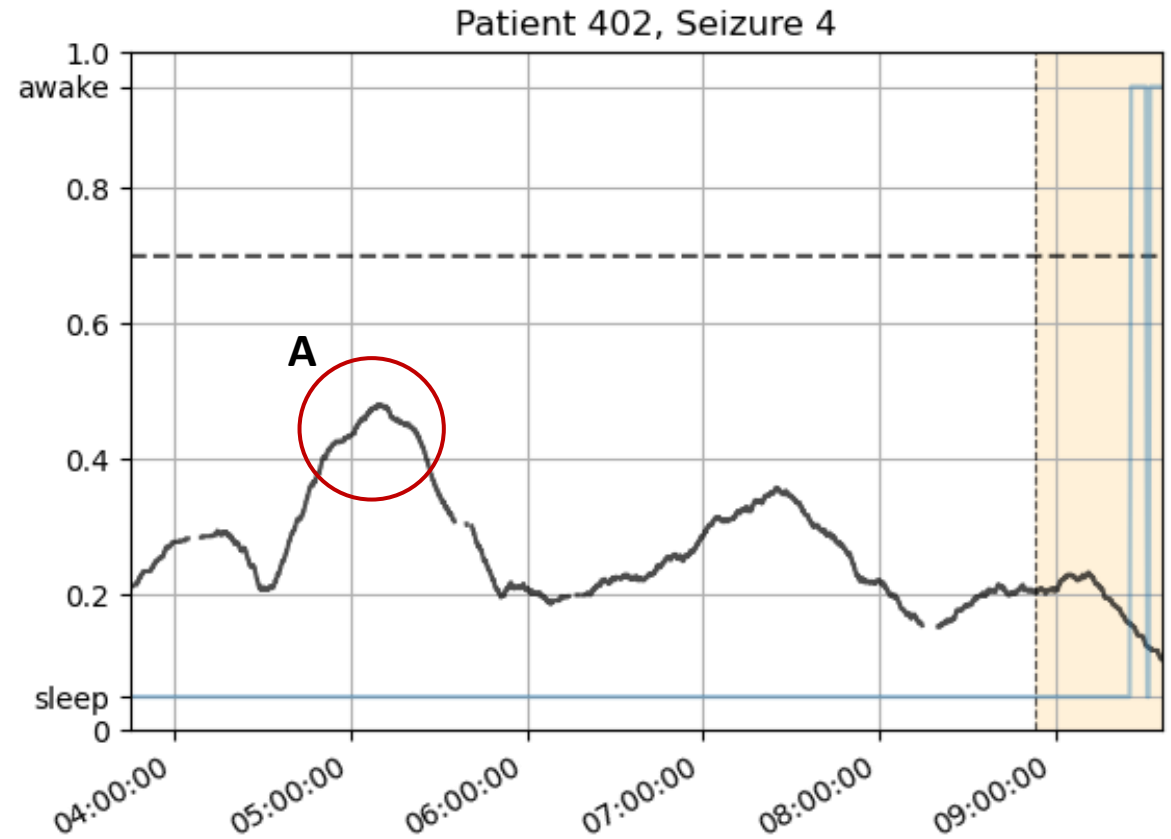
# Patient 402 – Over time

- Cluster A (04h30-05h10):
  - A peak of 0.47, sleep?
- Cluster B (08h55-09h35):
  - Did not predict the seizure



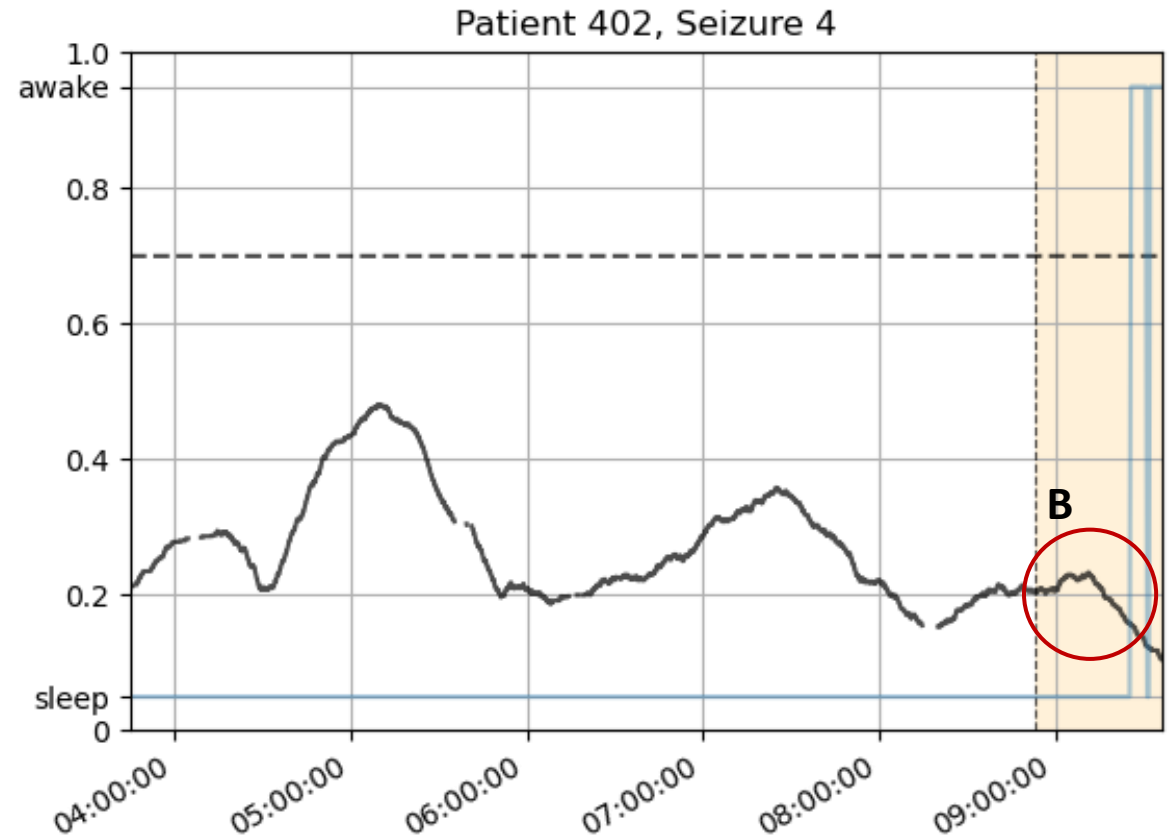
# Patient 402 – Counterfactual explanations

- For the prediction to be different (no peak), we would have to change the following features (top 3 by order of importance):
  - O1\_Ratio\_delta\_theta :↓
  - T8\_Ratio\_theta\_gamma2: ↑
  - F4\_Relative\_gamma1\_power: ↓
- Or, there was a peak because:
  - O1\_Ratio\_delta\_theta were high
  - T8\_Ratio\_theta\_gamma2 were low
  - F4\_Relative\_gamma1\_power were high



# Patient 402 – Counterfactual explanations

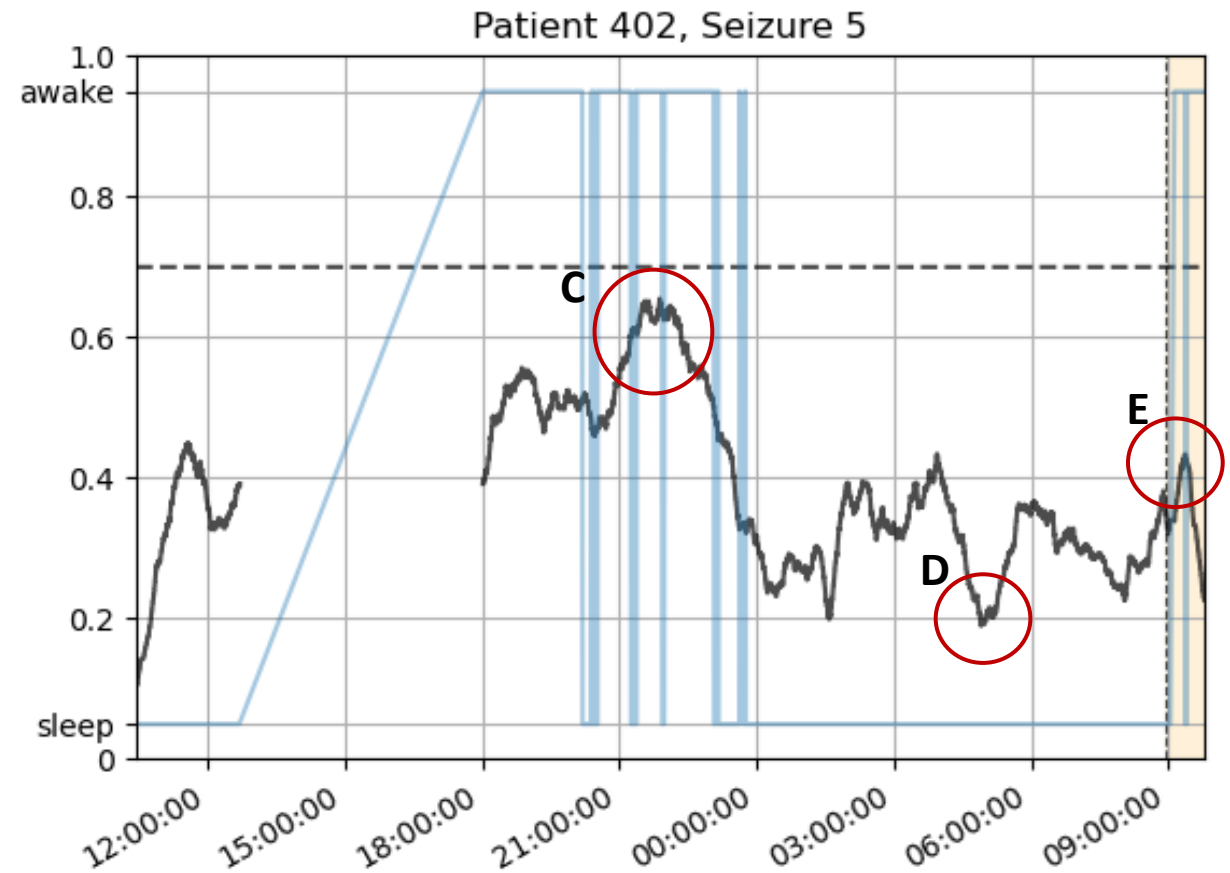
- For the prediction to be different (no peak), we would have to change the following features (top 3 by order of importance):
  - F4\_Kurtosis:  $\uparrow$
  - O1\_Ratio\_delta\_theta:  $\downarrow$
  - T8\_Ratio\_theta\_gamma2:  $\uparrow$
- Or, there was a peak because:
  - F4\_Kurtosis: were low
  - O1\_Ratio\_delta\_theta: were high
  - T8\_Ratio\_theta\_gamma2: were low





# Patient 402 – Over time

- Cluster C (21h00-00h00):
  - A peak of 0.66, sleep?
- Cluster E (08h10-09h40):
  - Again, at the same time, a seizure occurred but it was not predicted



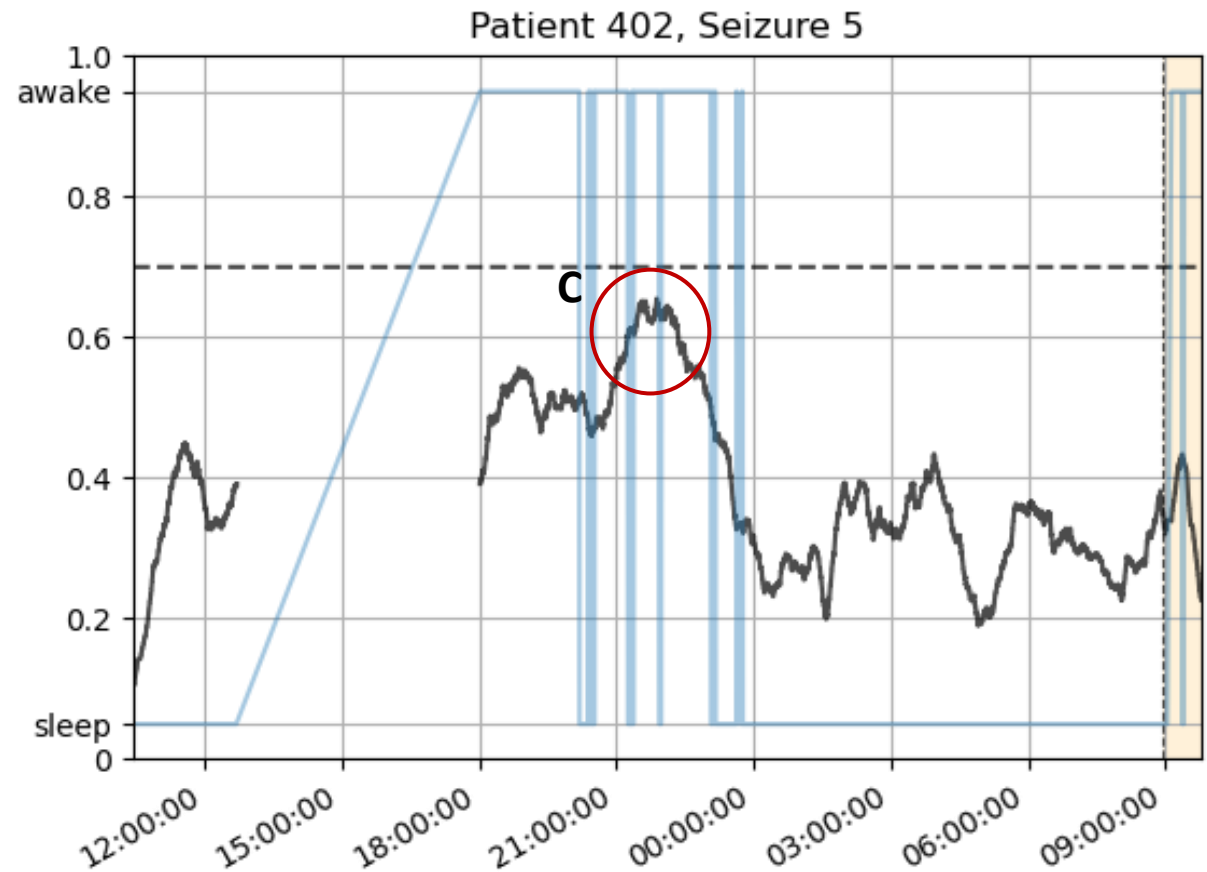
# Patient 402 – Counterfactual explanations

- For the prediction to be different (no peak), we would have to change the following features (top 3 by order of importance):

- F4\_Kurtosis: ↓
- O1\_Ratio\_delta\_theta: ↑
- T8\_Ratio\_theta\_gamma2: ↓

- Or, there was a peak because:

- F4\_Kurtosis was high
- O1\_Ratio\_delta\_theta was low
- T8\_Ratio\_theta\_gamma2 was high



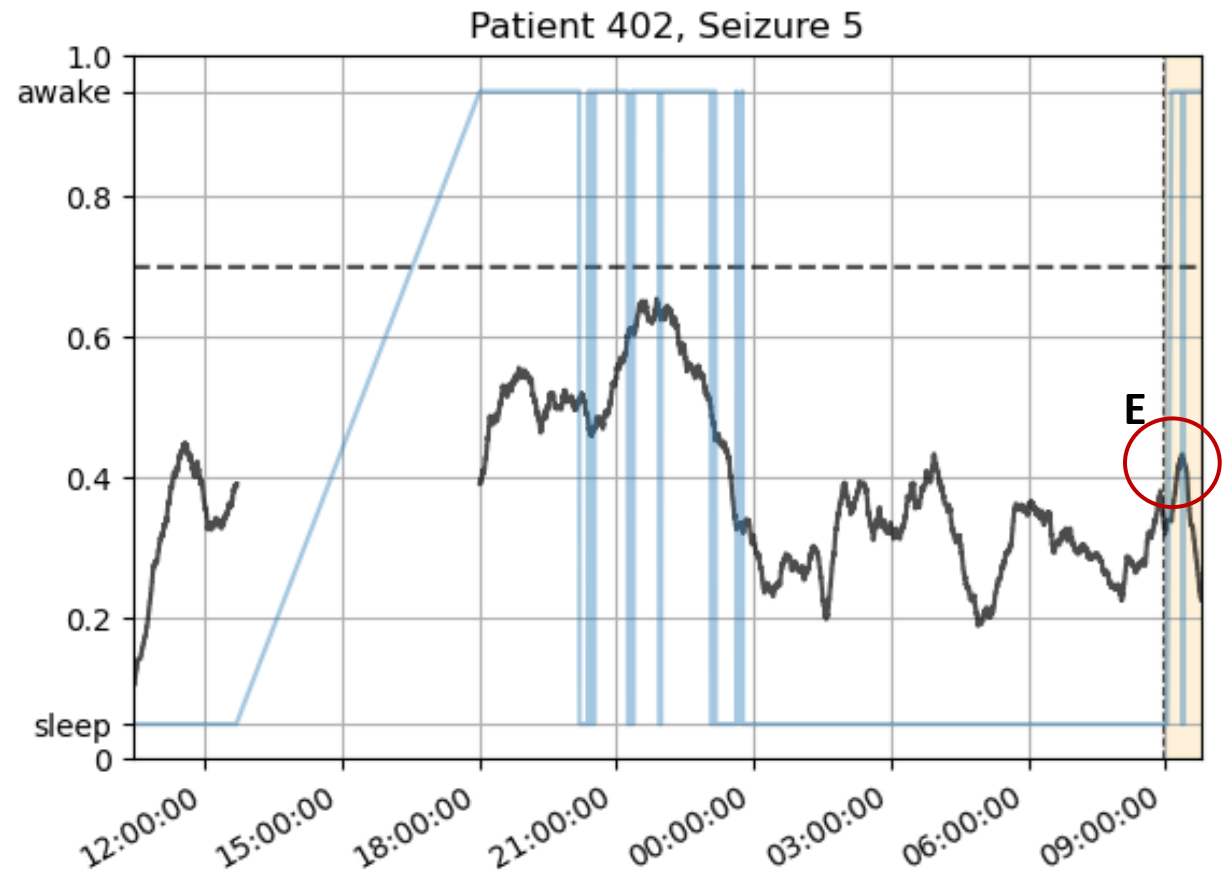
# Patient 402 – Counterfactual explanations

- For the prediction to be different (no peak), we would have to change the following features (top 3 by order of importance):

- F4\_Kurtosis:  $\uparrow$
- O1\_Ratio\_delta\_theta:  $\downarrow$
- T8\_Ratio\_theta\_gamma2:  $\uparrow$

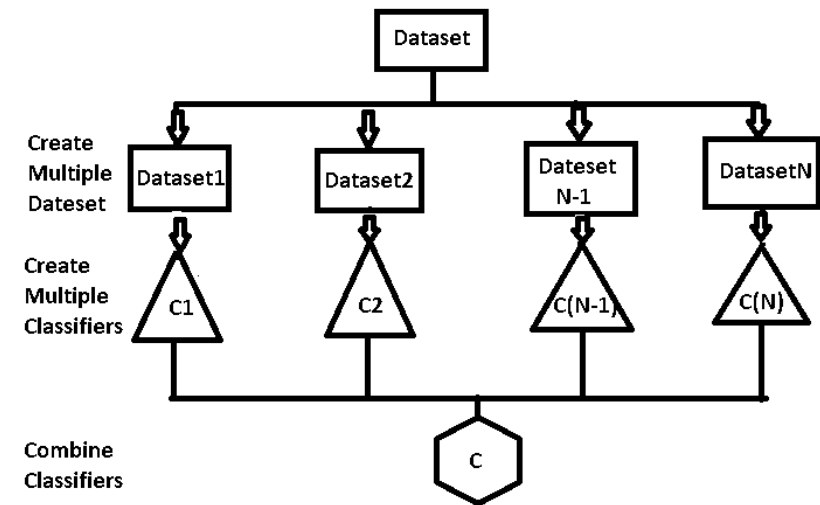
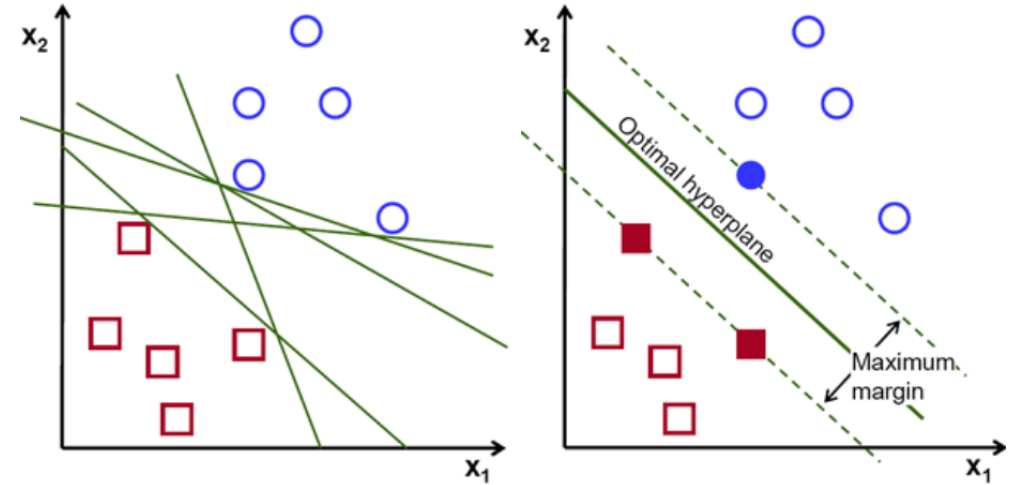
- Or, there was a peak because:

- O1\_Ratio\_delta\_theta was low
- T8\_Ratio\_theta\_gamma2 was high
- F4\_Relative\_gamma1\_power was low

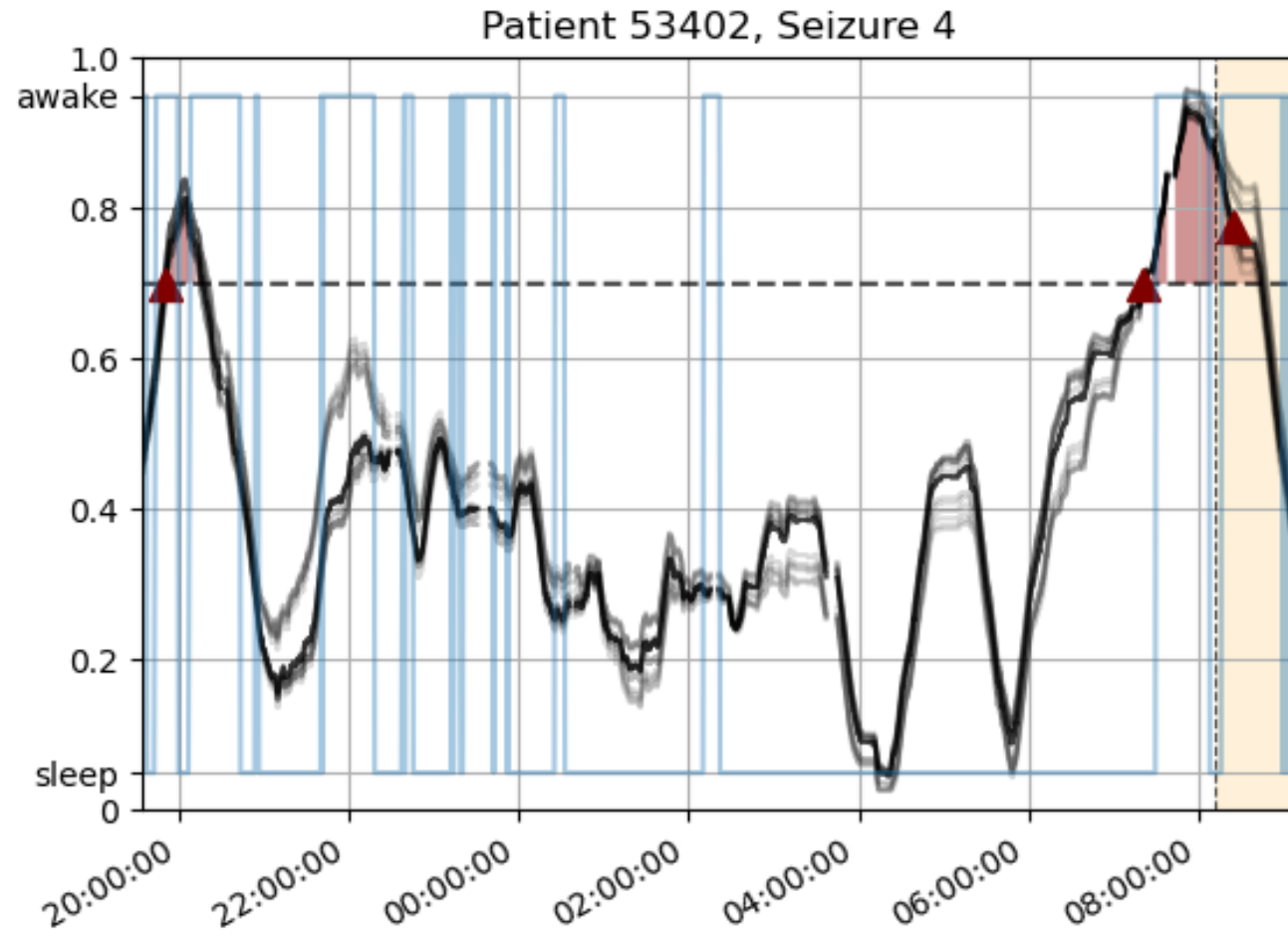


# Support Vector Machines

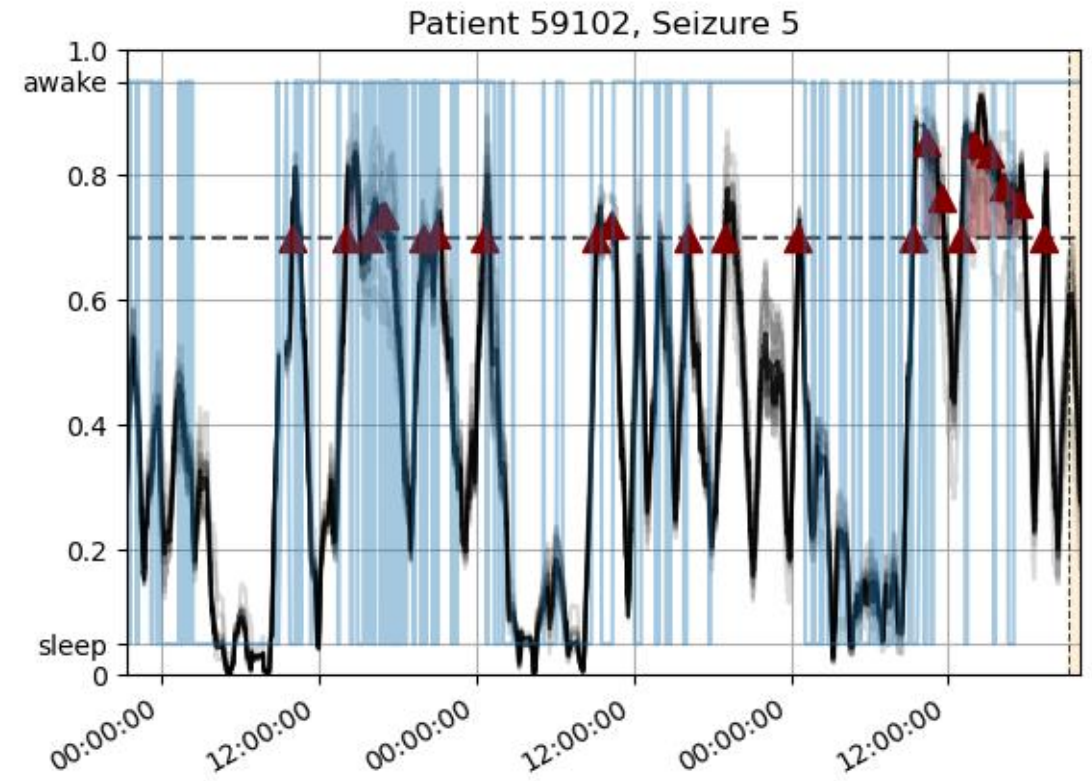
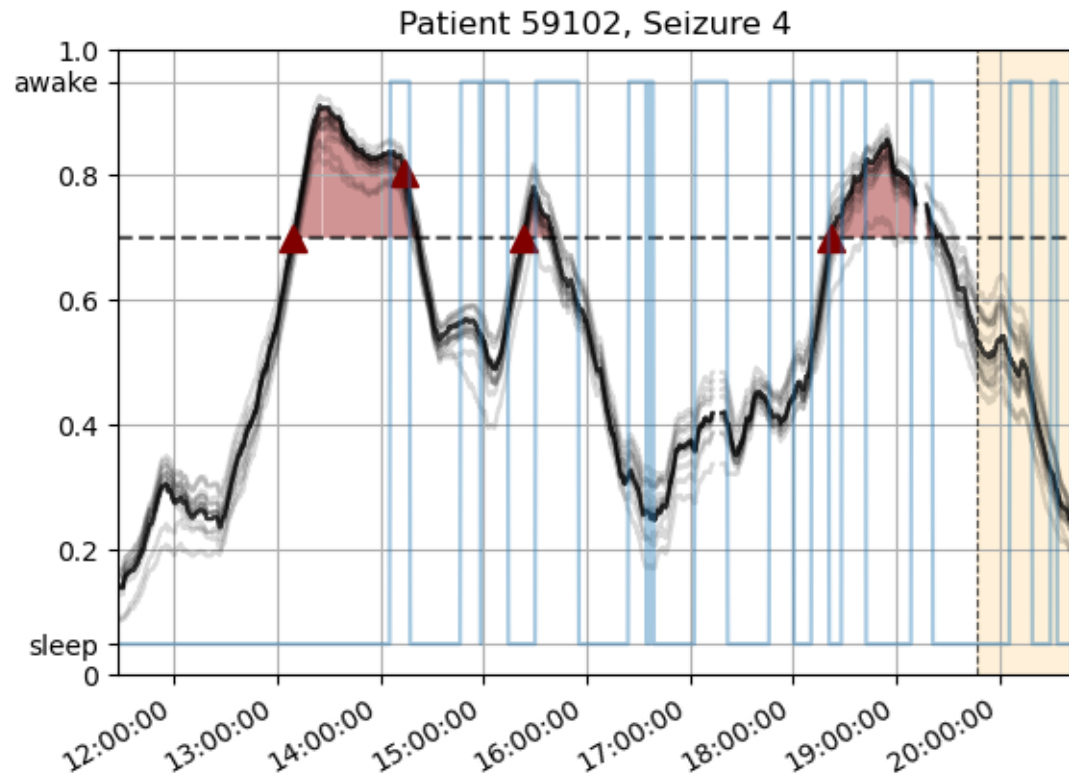
- 1) With the first 3 seizures:
  - Find the optimal pre-ictal period within a range of values: 30-60 minutes
  - Find the optimal number of features
- 2) Do the following 15 times:
  - Balance the data: choose random samples from inter-ictal
  - Train an SVM classifier
- 3) Each classifier then makes a prediction. The prediction is obtained by a voting system



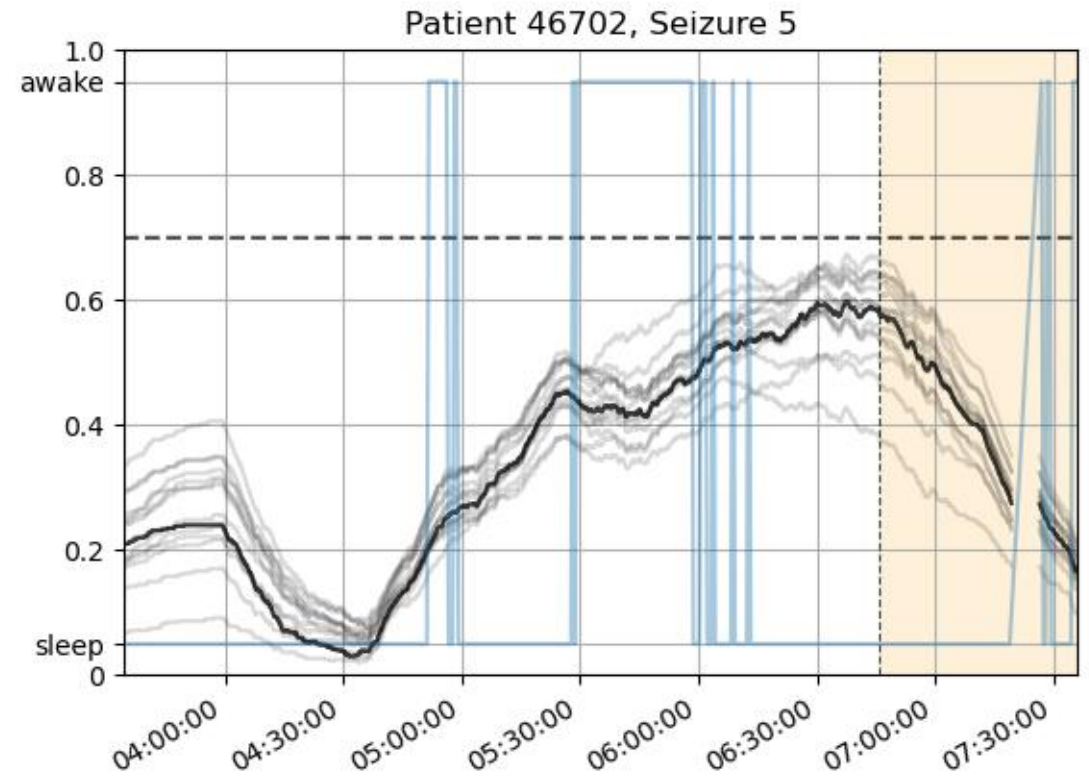
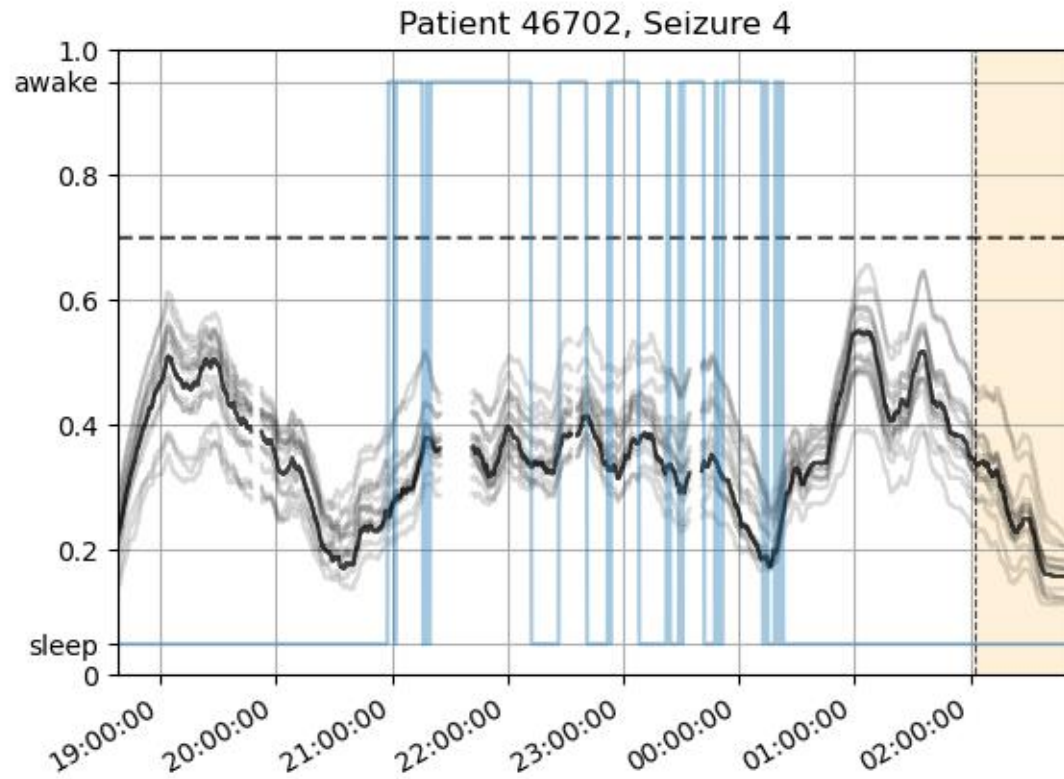
# Patient 53402 – Over time



# Patient 59102 – Over time

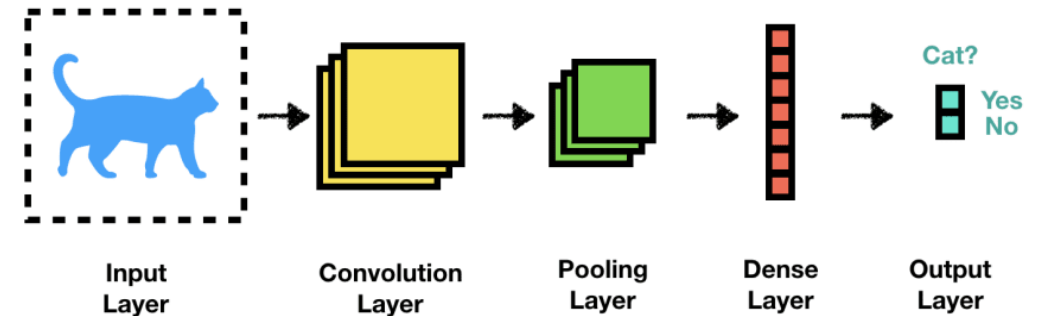
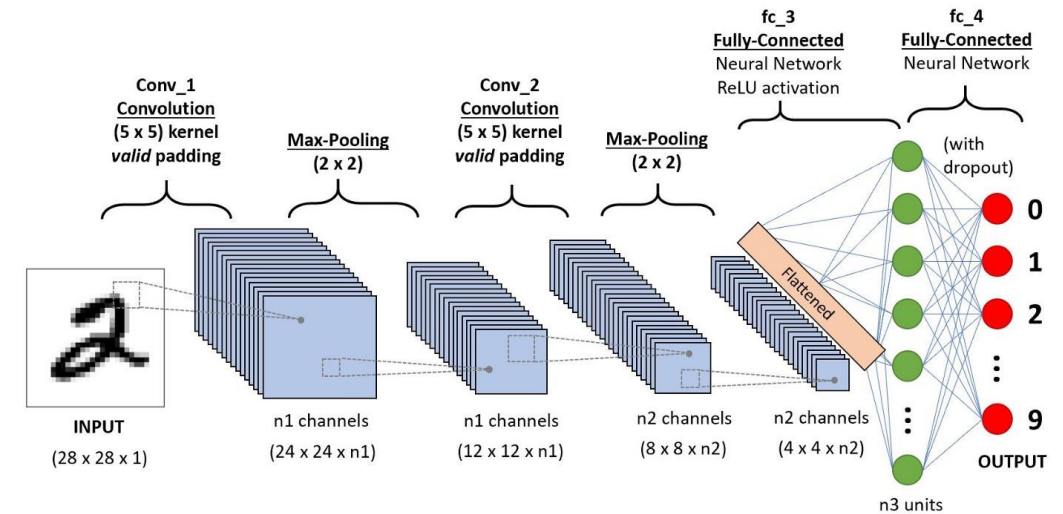


# Patient 46702 – Over time



# Convolutional Neural Networks

- 1) With the first 3 seizures:
  - Find the optimal pre-ictal period within a range of values: 30-60 minutes
  - Find the optimal number of features
- 2) Do the following 3 times:
  - Balance the data: choose random samples from inter-ictal
  - Train an CNN classifier
- 3) Each classifier then makes a prediction. The prediction is obtained by a voting system





# Patient 8902 – Overall performance

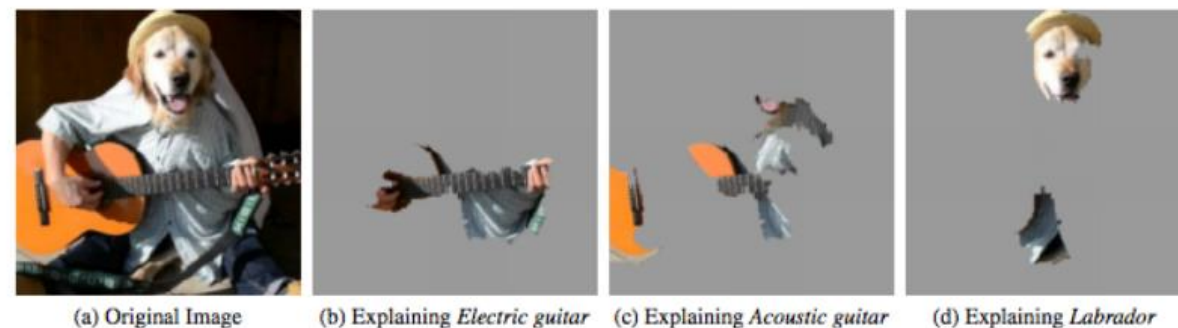
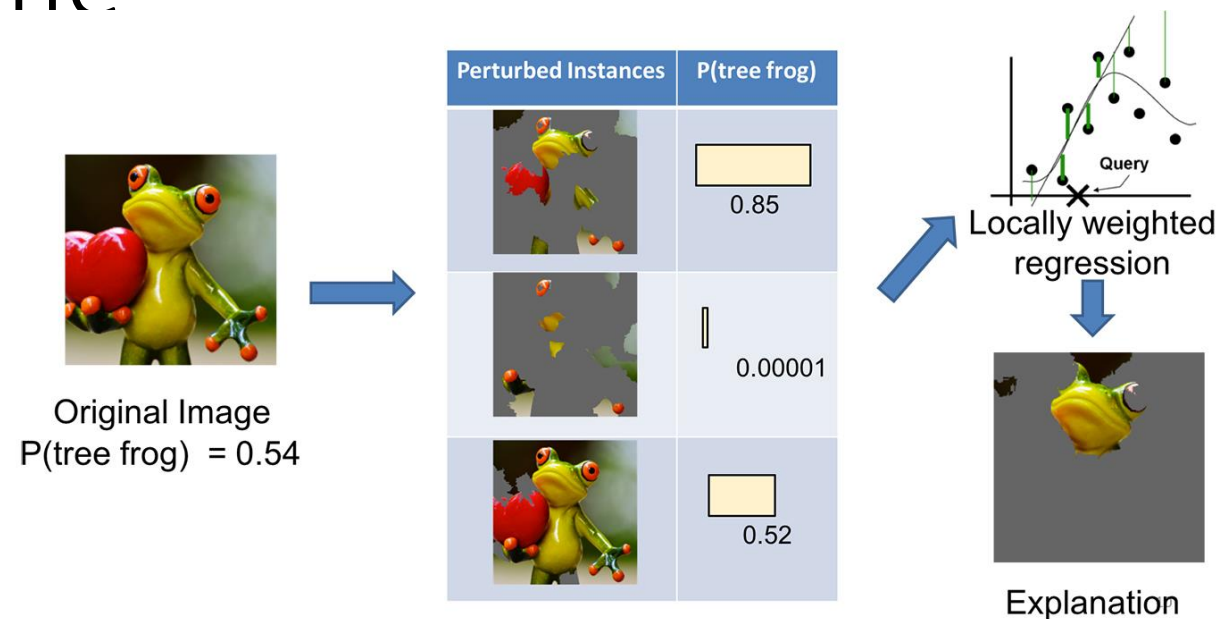
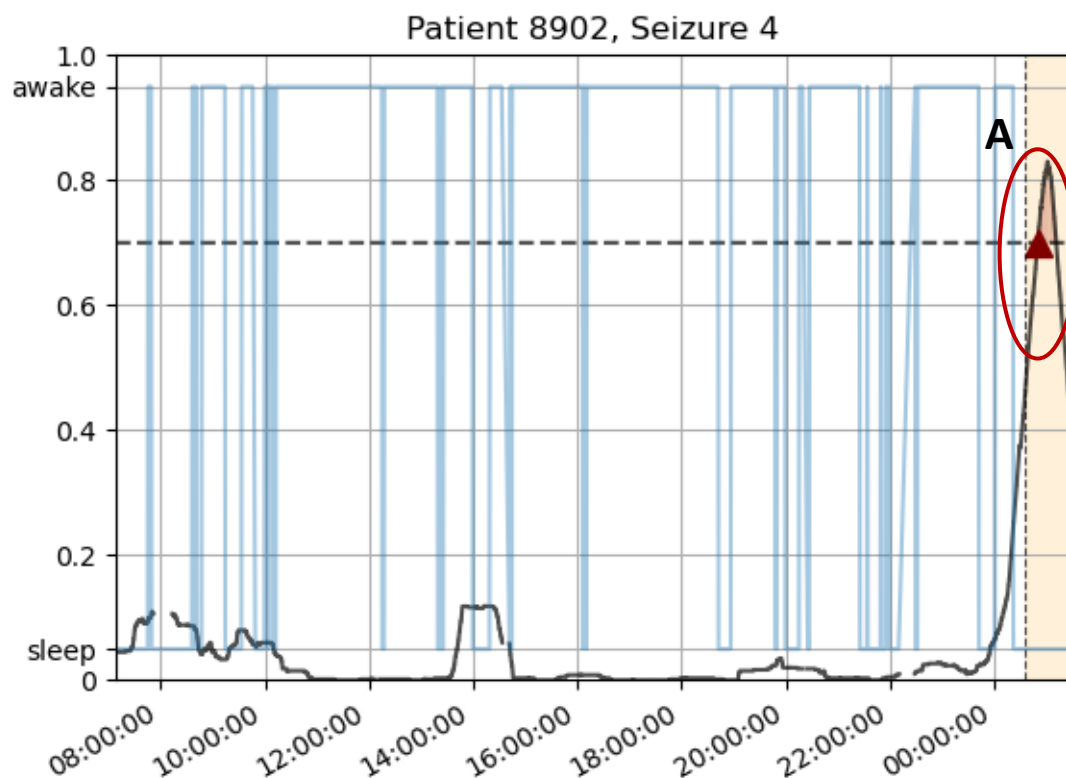
- SS:50%, FPR/h=0.00
- Above chance
- All seizures during the night.  
Sleep related?
- Surgery decision: not offered
- Focus localization=tpl  
Temporal polar left

	Onset	Type	Pattern	Vigilance state at onset
Training Seizures	Day 0 00:51:14	Unclassified	Rhythmic alpha waves	Awake
	Day 1 00:03:23	FOIA	Rhythmic beta waves	Awake
	Day 1 06:37:05	FOIA	Rhythmic alpha waves	Awake
Testing Seizures	Day 2 01:35:56	FOIA	Amplitude depression	Awake
	Day 2 06:10:26	FOIA	Rhythmic alpha waves	Awake

	Onset	Predicted	#False Alarms	Hours of analysed data
Testing Seizures	Seizure 4 Day 2 01:35:56	Yes	1	≈17
	Seizure 5 Day 2 06:10:26	Yes	1	≈4

# Patient 8902 – Over time

- Ver ficheiro 8902\_seizure\_4\_DL\_00h30\_01h00 para ver o que a rede usou para dizer que as janelas correspondiam a um período pre-ictal



**Figure 4: Explaining an image classification prediction made by Google's Inception network, highlighting positive pixels. The top 3 classes predicted are "Electric Guitar" ( $p = 0.32$ ), "Acoustic guitar" ( $p = 0.24$ ) and "Labrador" ( $p = 0.21$ )**