การเปรียบเทียบวิธีการกำจัดสัญญาณรบกวนในสัญญาณคลื่นไฟฟ้าสมอง

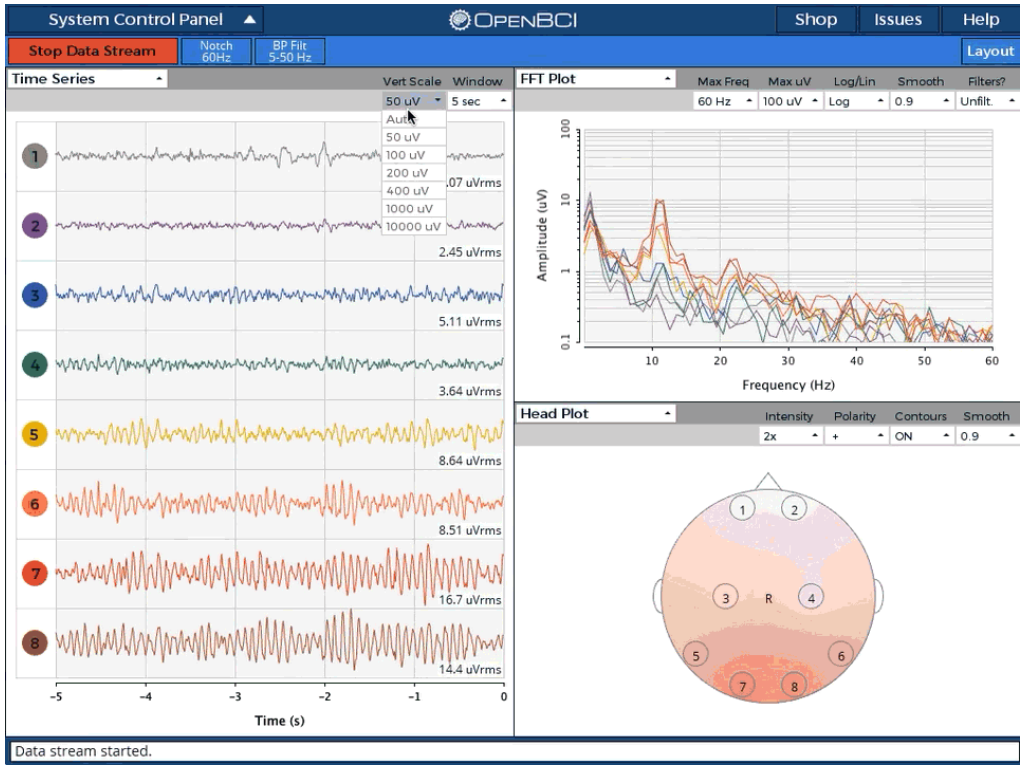
ธนกฤต แสนเจริญกิจ

อาจารย์ที่ปรึกษา ผศ.ดร.จิตโกมุท ส่งศิริ

ภาควิชาวิศวกรรมไฟฟ้า คณะวิศวกรรมศาสตร์

จุฬาลงกรณ์มหาวิทยาลัย

ที่มาและความสำคัญ



แหล่งที่มา : https://images.app.goo.gl/EdK1e8jiUowHy1VU9

แหล่งที่มา :
https://www.sciencephoto.com/media/904128/view/electroencephalography

แหล่งที่มา : https://www.directmedical.ie/product/somnowatch-plus/sow_eeg_03/

วัตถุประสงค์

- เพื่อศึกษาวิธีการกำจัดสัญญาณรบกวนที่ประยุกต์ใช้ได้กับการบันทึกสัญญาณคลื่นไฟฟ้าสมองหนึ่งช่องสัญญาณ
- เพื่อเปรียบเทียบผลลัพธ์ของวิธีการกำจัดสัญญาณรบกวนที่แตกต่างกัน ซึ่งมีวิธีการกำจัดสัญญาณรบกวนดังต่อไปนี้
  - Thresholding
  - Random forest
  - Convolutional neural network

ขอบเขตของโครงงาน

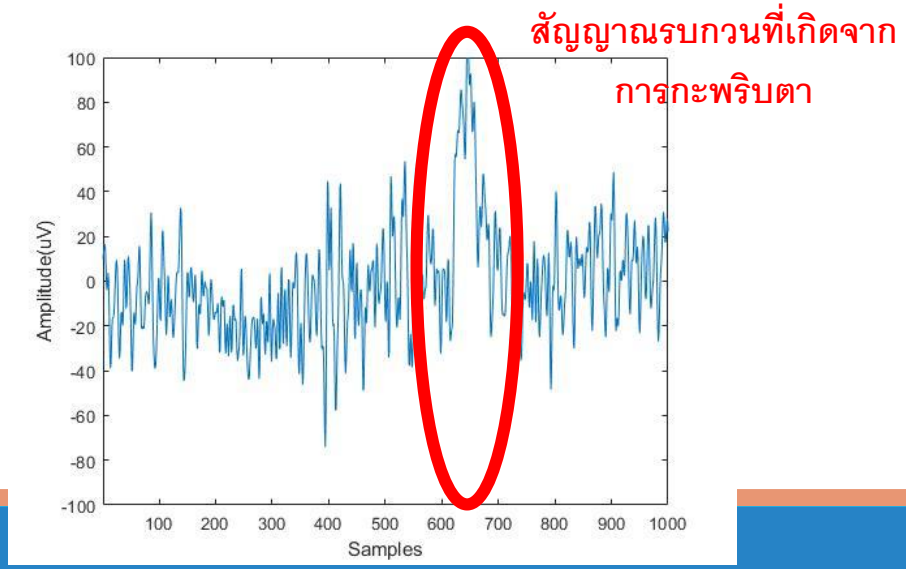
- กำจัด ***สัญญาณรบกวนที่เกิดจากตา*** เท่านั้น เช่น **การกะพริบตา**
- ข้อมูลที่นำมาใช้จะเป็น ***ข้อมูลสังเคราะห์*** จำลองการบันทึกสัญญาณจากเครื่องบันทึกคลื่นไฟฟ้าสมอง ***หนึ่งช่องสัญญาณ***

แหล่งที่มา : https://www.amazon.ca/NeuroSky-Single-Channel-Measurement-Relaxation-8rdsf-tg1324061/dp/B01M689ZAF

- การ ***เปรียบเทียบประสิทธิภาพ*** ของวิธีการกำจัดสัญญาณรบกวน จะเปรียบเทียบด้วย ***ความแตกต่างของ correlation coefficient*** และ ***normalized mean square error*** ของวิธีการต่าง ๆ

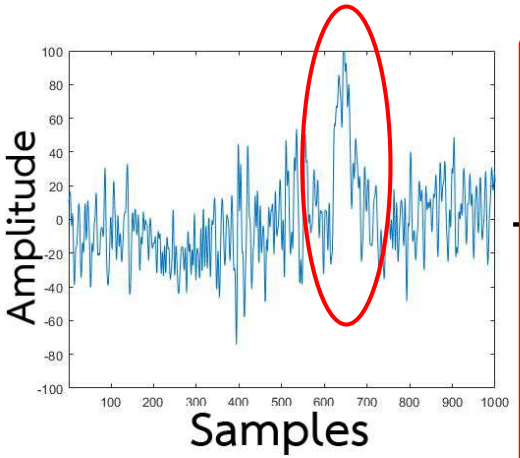
ความรู้พื้นฐานของสัญญาณคลื่นไฟฟ้าสมอง [3]

- สัญญาณคลื่นไฟฟ้าสมอง จะมีย่านความถี่ในช่วง 0.01 – 100 Hz
- มี Amplitude โดยทั่วไปไม่เกิน 100 uV
- สัญญาณรบกวนที่เกิดจากการเคลื่อนที่ของตา เช่น การกะพริบตา จะเกิดในช่วง 0.5 – 5 เฮิรตซ์



หลักการและทฤษฎีที่เกี่ยวข้อง

กระบวนการกำจัดสัญญาณรบกวน



สัญญาณคลื่นไฟฟ้าสมองที่มีสัญญาณรบกวนจากตา

การแยกหาคุณลักษณะเฉพาะ

1.Energy (0.5-5Hz)
2.Peak
3.Kurtosis
4.Statistical value of wavelet coefficients

คุณลักษณะเฉพาะต่างๆ

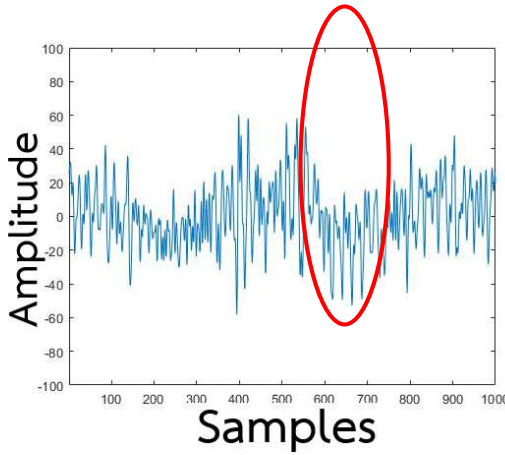
การตรวจจับบริเวณที่มีสัญญาณรบกวน

วิธีการ
1.Thresholding
2.Random forest
3.CNN

บริเวณที่ตรวจจับได้ว่ามีสัญญาณรบกวน

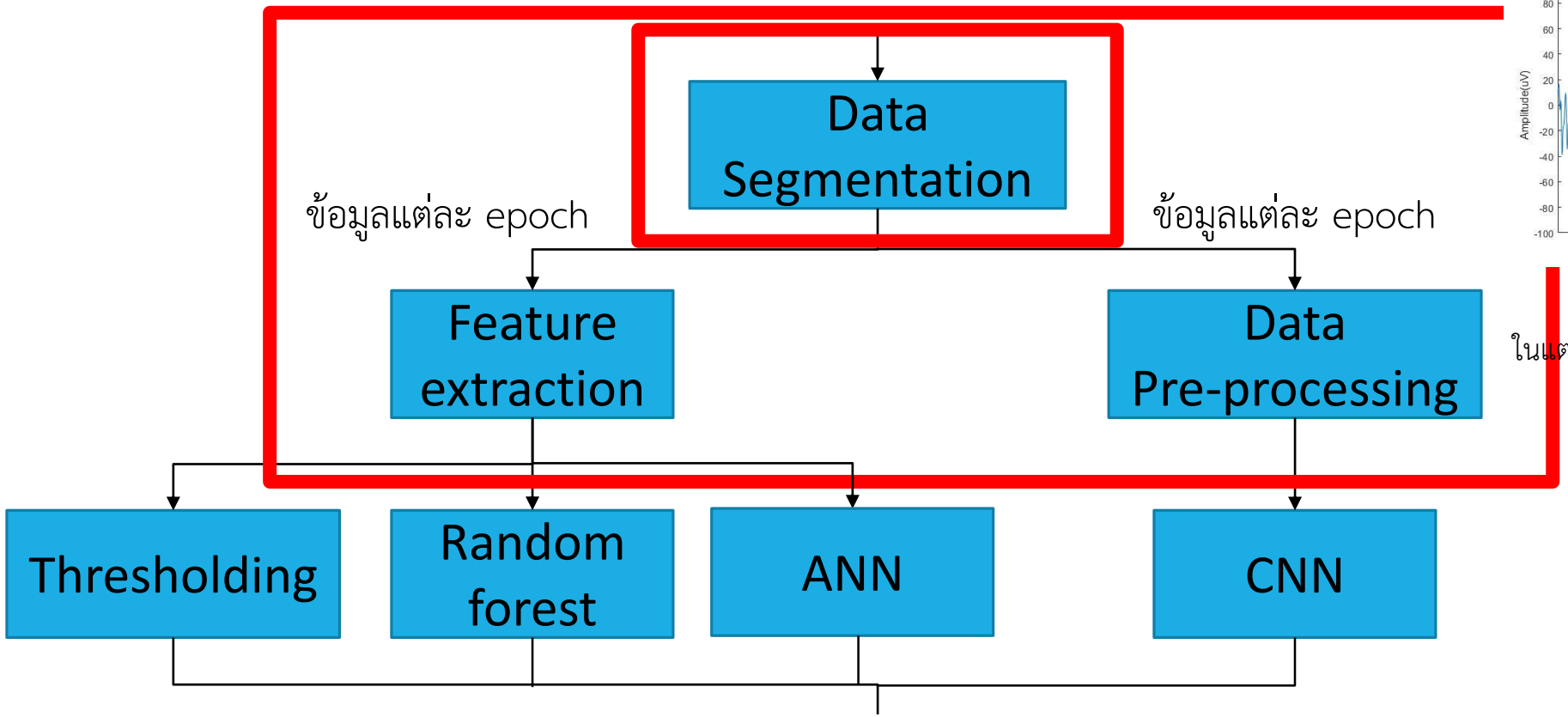
การกำจัดสัญญาณรบกวนในบริเวณที่ตรวจจับได้

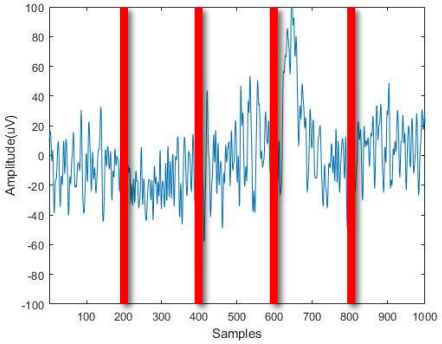
วิธีการ
Savitzky-Golay filter



สัญญาณคลื่นไฟฟ้าสมองที่สัญญาณรบกวนจากตาถูกกำจัด

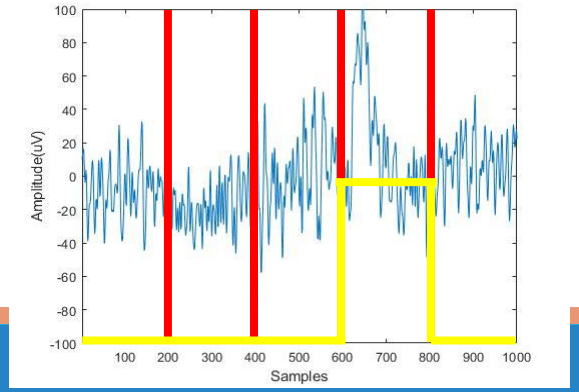
การแยกหาคุณลักษณะเฉพาะ





แบ่งสัญญาณเป็น epoch

ในแต่ละ epoch จะมี N data point



Features extraction

สัญญาณคลื่นไฟฟ้าสมอง
ที่มีสัญญาณรบกวน

↓

Data Segmentation

↓

Features extraction

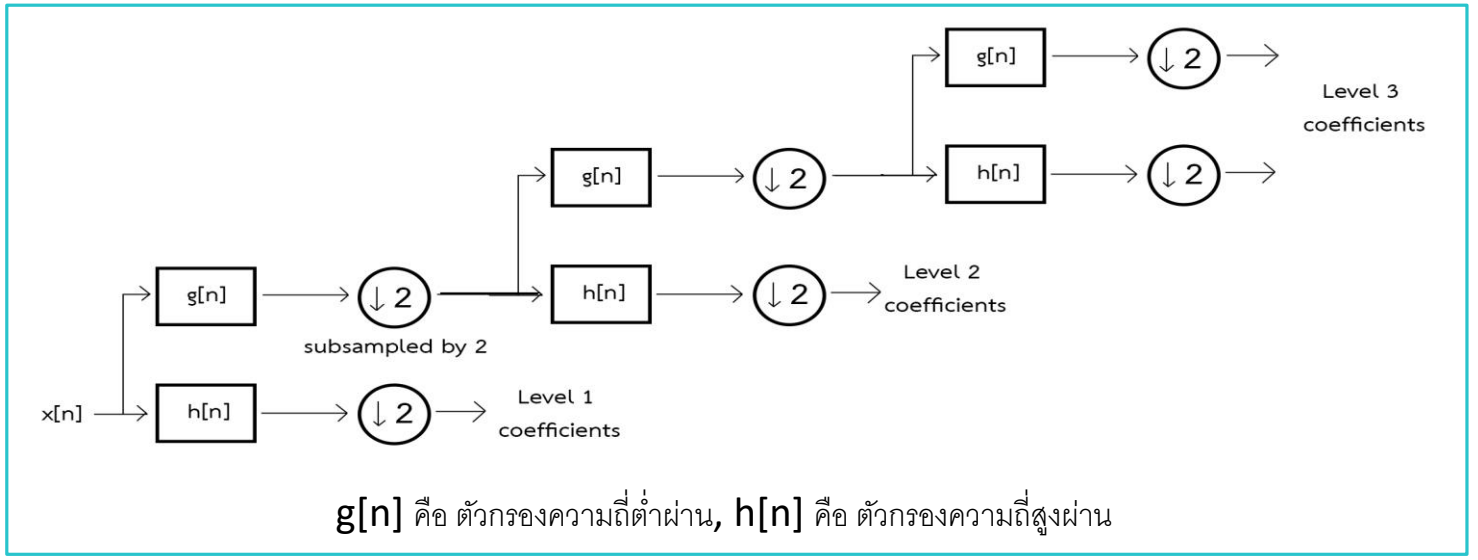
↓

Thresholding, Random forest, ANN

↓

บริเวณที่พบว่า
มีสัญญาณรบกวนหรือไม่

- Feature 1 พลังงานในย่านความถี่ 0.5 - 5 Hz ของแต่ละ epoch
- Feature 2 ขนาดของสัญญาณที่สูงที่สุด ในแต่ละ epoch
- Feature 3 Kurtosis ของการกระจายตัวของสัญญาณ ในแต่ละ epoch
- Feature 4 ค่าทางสถิติของค่าสัมประสิทธิ์เวฟเล็ต ของแต่ละ epoch



- Standard deviation
- 75th percentile
- 25th percentile
- Kurtosis
- Mean
- Median

Data
pre-processing

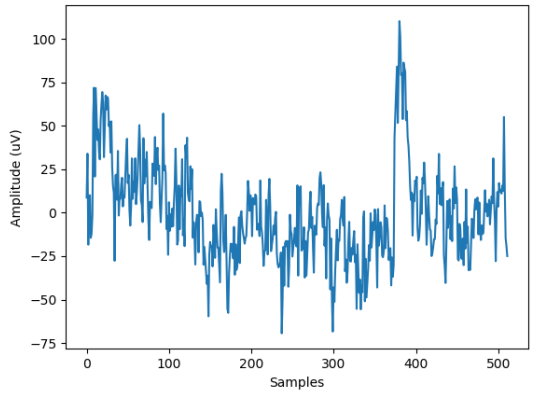
สัญญาณคลื่นไฟฟ้าสมอง
ที่มีสัญญาณรบกวน

Data Segmentation

Data pre-processing

CNN

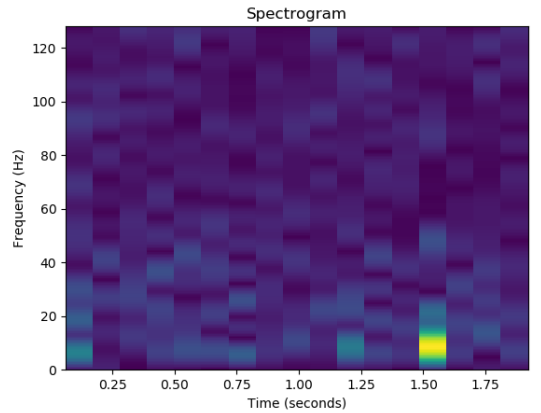
บริเวณที่พบว่า
มีสัญญาณรบกวนหรือไม่



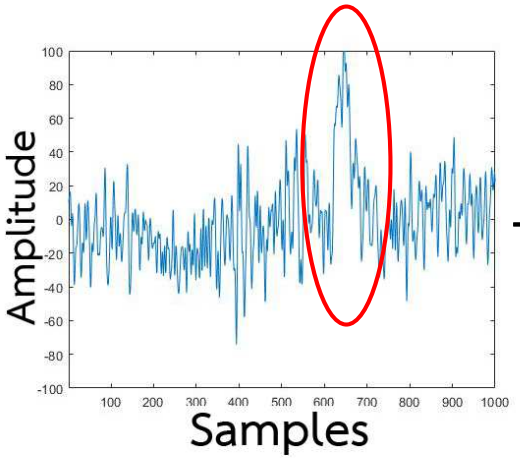
Short-time Fourier Transform

2D array

Magnitude squared of the STFT



กระบวนการกำจัดสัญญาณรบกวน



สัญญาณคลื่นไฟฟ้าสมองที่มีสัญญาณรบกวนจากตา

→

การแยกหา
คุณลักษณะเฉพาะ

1.Energy (0.5-5Hz)
2.Peak
3.Kurtosis
4.Statistical value of wavelet coefficients

คุณลักษณะเฉพาะต่างๆ →

การตรวจจับบริเวณที่มีสัญญาณรบกวน

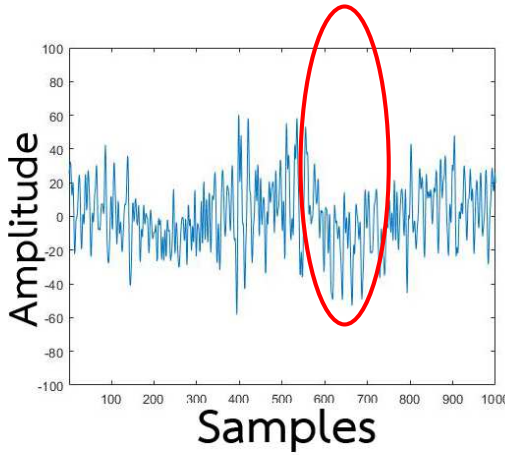
วิธีการ
1.Thresholding
2.Random forest
3.CNN

บริเวณที่ตรวจจับได้ว่ามีสัญญาณรบกวน →

การกำจัดสัญญาณรบกวนในบริเวณที่ตรวจจับได้

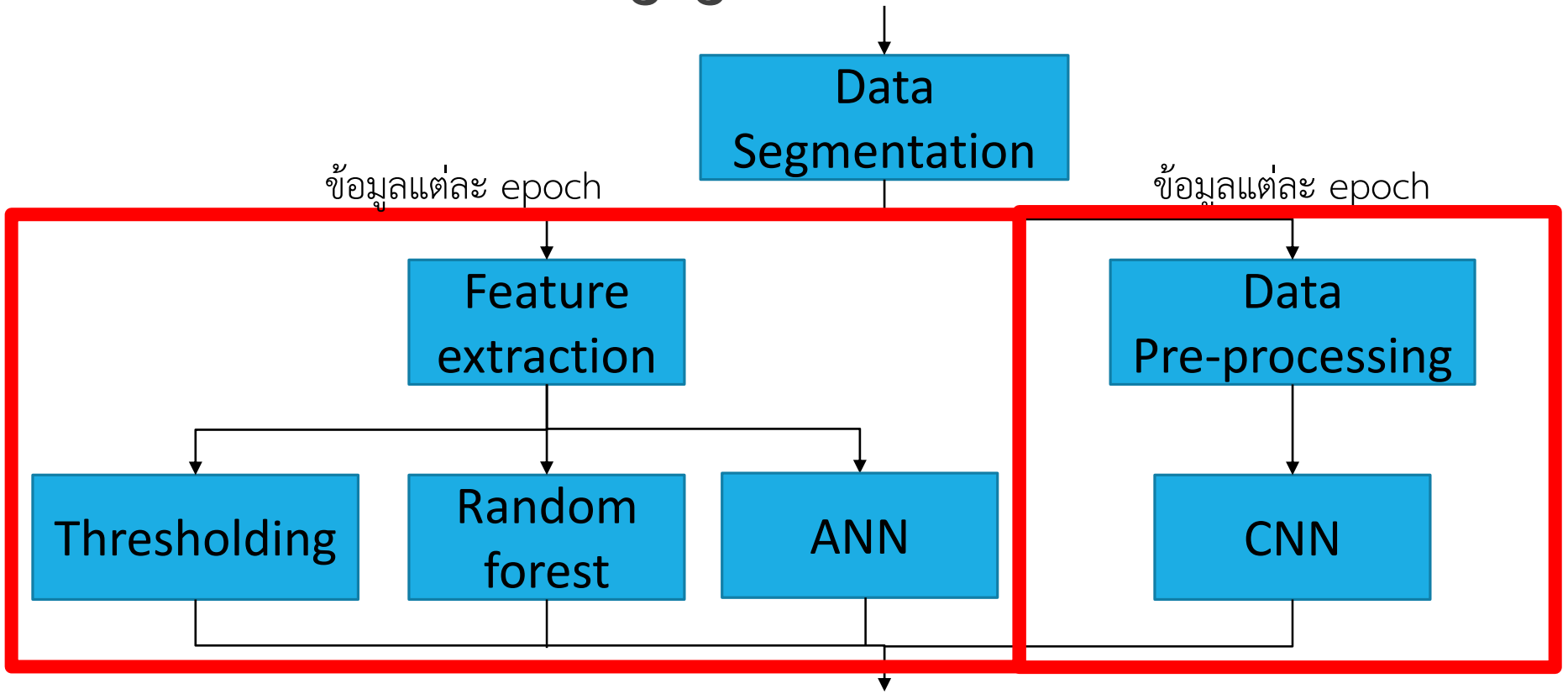
วิธีการ
Savitzky-Golay filter

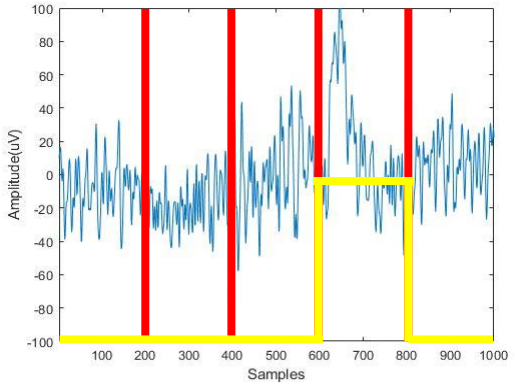
→



สัญญาณคลื่นไฟฟ้าสมองที่สัญญาณรบกวนจากตาถูกกำจัด

การตรวจจับบริเวณที่มีสัญญาณรบกวน

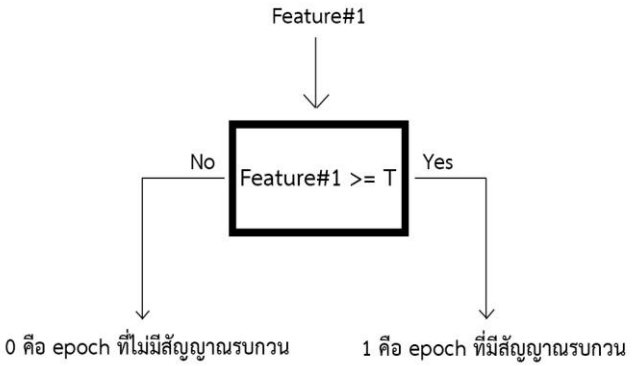




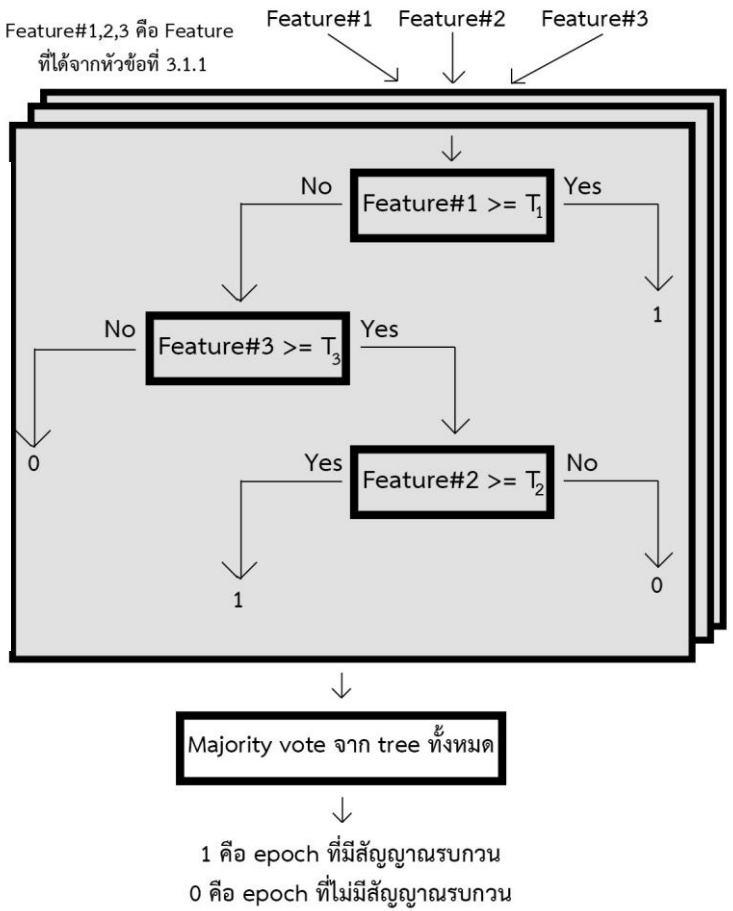
Thresholding

Feature#1 คือ ค่าพลังงานในย่านความถี่ 0.5 - 5 Hz ของแต่ละ epoch

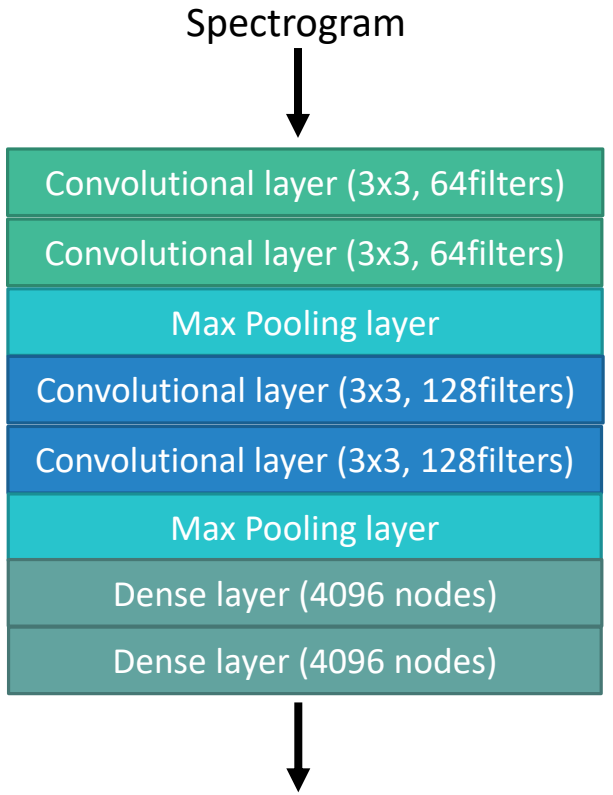
T คือ ค่า Threshold



Random forest



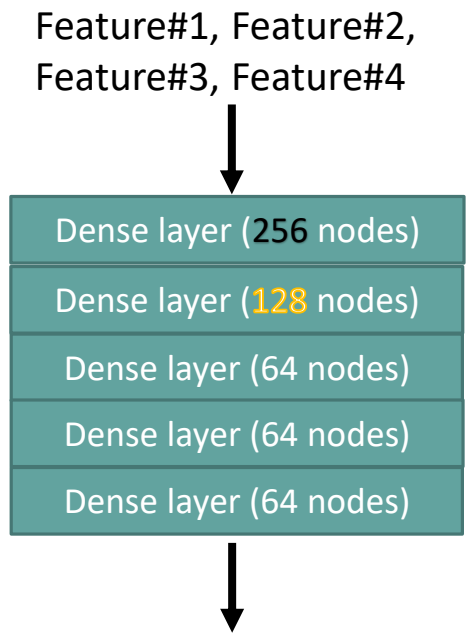
CNN



1 คือ epoch ที่มีสัญญาณรบกวน

0 คือ epoch ที่ไม่มีสัญญาณรบกวน

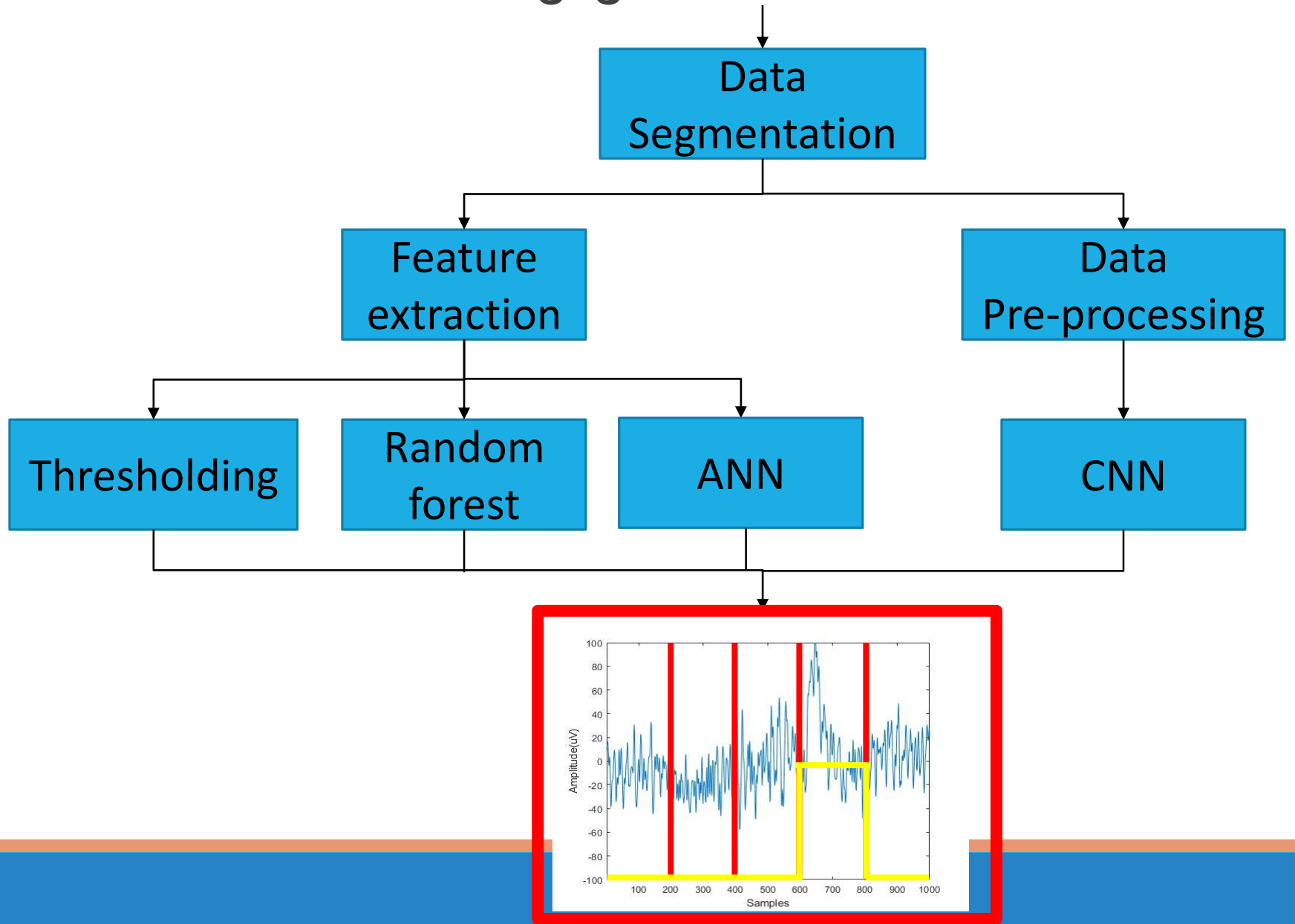
ANN



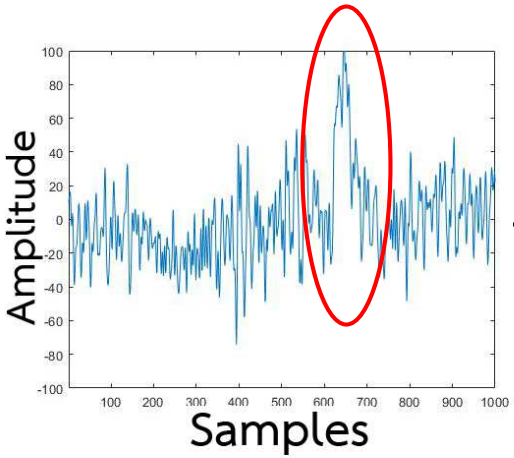
1 คือ epoch ที่มีสัญญาณรบกวน

0 คือ epoch ที่ไม่มีสัญญาณรบกวน

การตรวจจับบริเวณที่มีสัญญาณรบกวน



กระบวนการกำจัดสัญญาณรบกวน



สัญญาณคลื่นไฟฟ้าสมองที่มีสัญญาณรบกวนจากตา

การแยกหาคุณลักษณะเฉพาะ

1.Energy (0.5-5Hz)
2.Peak
3.Kurtosis
4.Statistical value of wavelet coefficients

คุณลักษณะเฉพาะต่างๆ

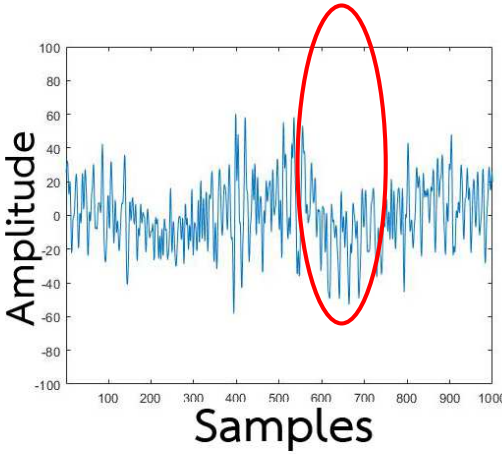
การตรวจจับบริเวณที่มีสัญญาณรบกวน

วิธีการ
1.Thresholding
2.Random forest
3.CNN

บริเวณที่ตรวจจับได้ว่ามีสัญญาณรบกวน

การกำจัดสัญญาณรบกวนในบริเวณที่ตรวจจับได้

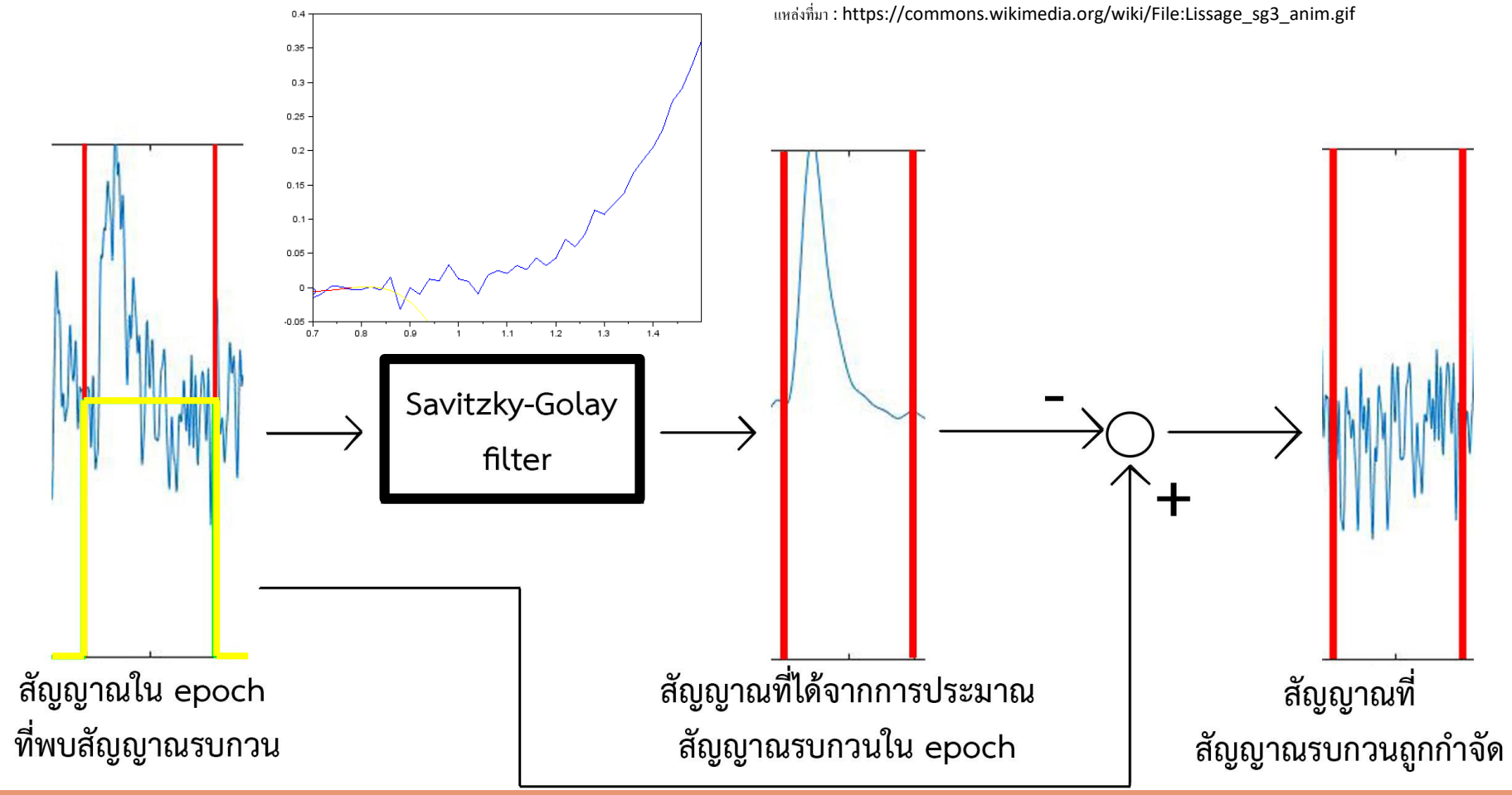
วิธีการ
Savitzky-Golay filter



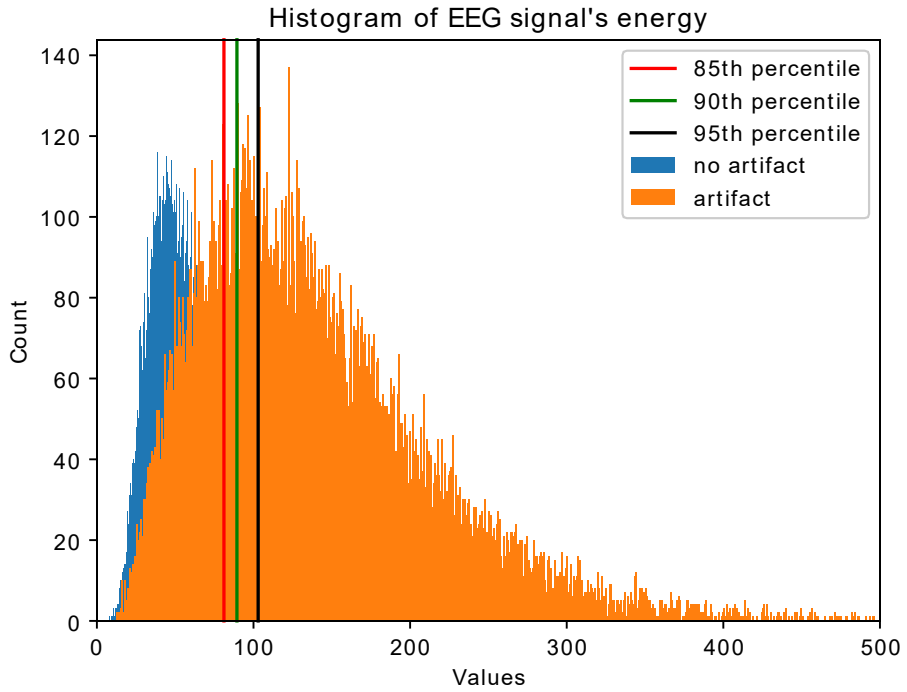
สัญญาณคลื่นไฟฟ้าสมองที่สัญญาณรบกวนจากตาถูกกำจัด

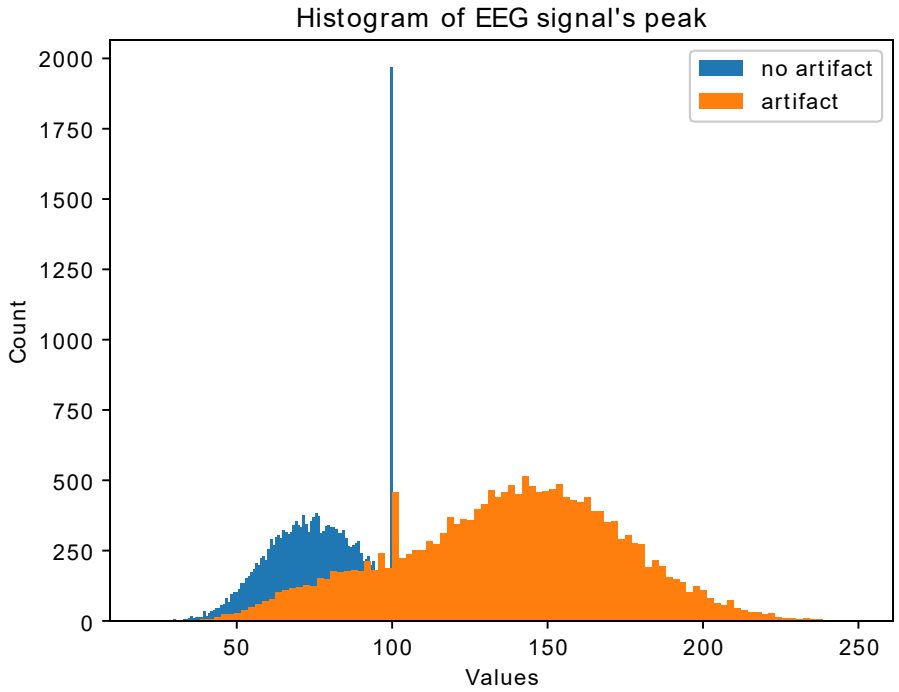
การกำจัดสัญญาณรบกวนในบริเวณที่ตรวจจับได้

แหล่งที่มา : https://commons.wikimedia.org/wiki/File:Lissage_sg3_anim.gif



ผลลัพธ์การดำเนินงาน

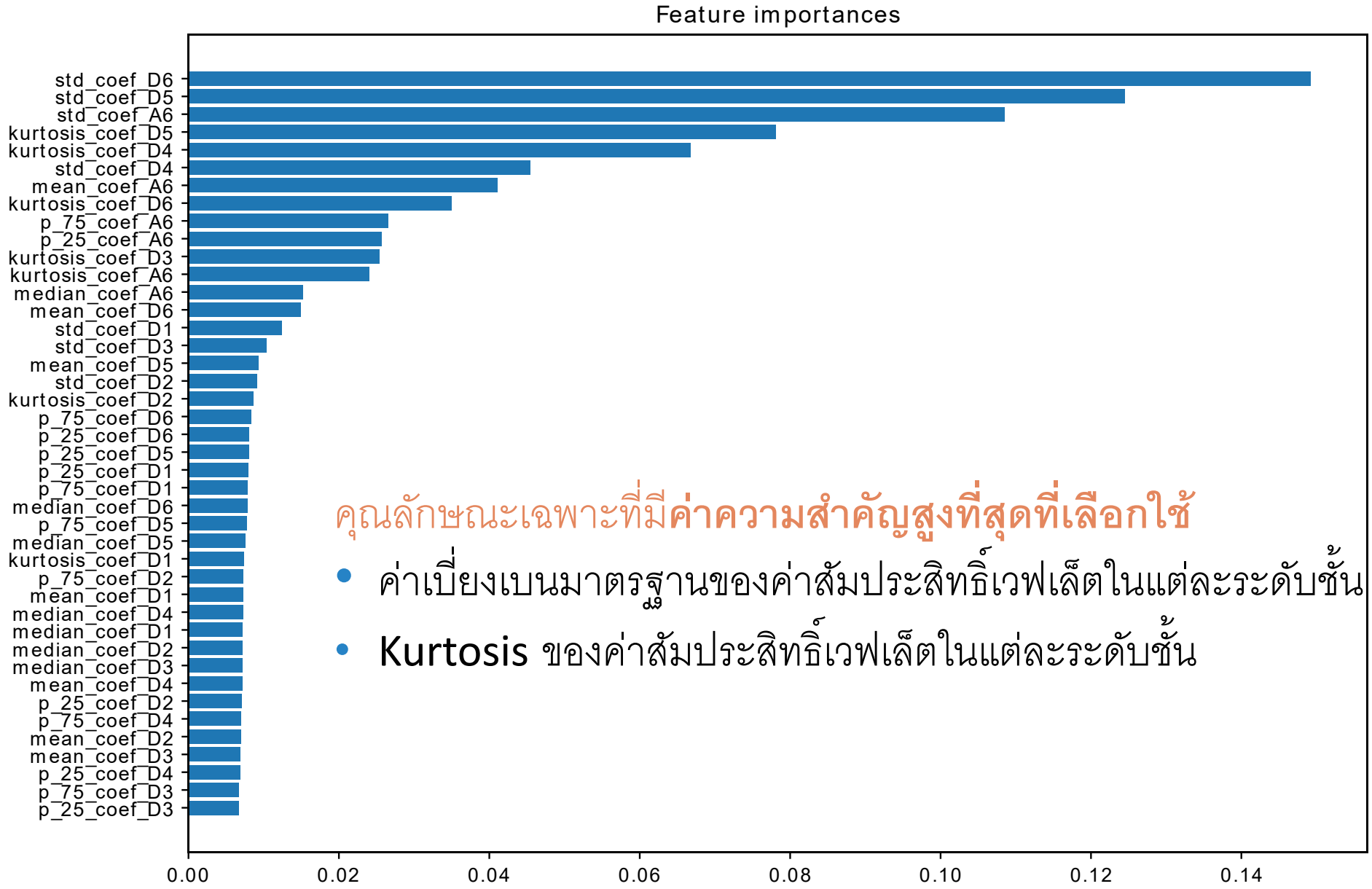




การคัดเลือกคุณลักษณะเฉพาะ

บริเวณที่**มีสัญญาณรบกวน**จะทำให้ **ค่าพลังงานในย่านความถี่** 0.5-5 เฮิรตซ์ และ **ขนาดของสัญญาณที่สูงที่สุด มีค่าสูงกว่า**บริเวณที่**ไม่มีสัญญาณรบกวน**

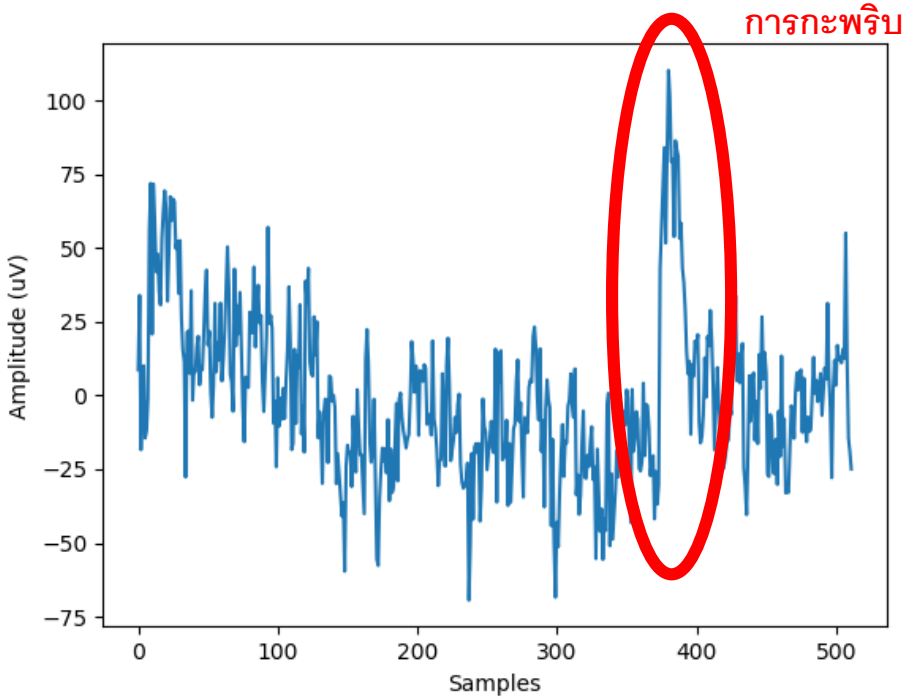
การคัดเลือกคุณลักษณะเฉพาะด้วยค่าความสำคัญ

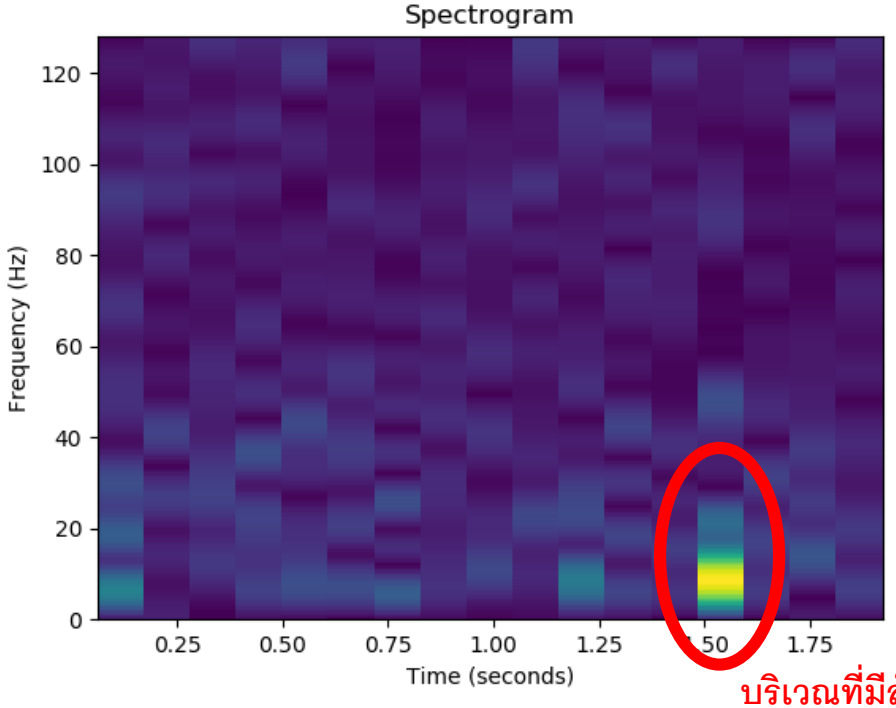


คุณลักษณะเฉพาะที่มี**ค่าความสำคัญสูงที่สุดที่เลือกใช้**

- ค่าเบี่ยงเบนมาตรฐานของค่าสัมประสิทธิ์เวฟเล็ตในแต่ละระดับชั้น
- Kurtosis ของค่าสัมประสิทธิ์เวฟเล็ตในแต่ละระดับชั้น

สัญญาณรบกวนที่เกิดจาก
การกะพริบตา





บริเวณที่มีสัญญาณรบกวน
ที่เกิดจากการกะพริบตา

บริเวณที่มีสัญญาณรบกวน แผนภาพสเปกตรัม ณ ช่วงความถี่และตำแหน่งของสัญญาณรบกวนนั้น จะมีค่าสูงแตกต่างจากบริเวณที่ไม่มีสัญญาณรบกวน

แผนภาพสเปกตรัมบริเวณที่มีสัญญาณรบกวนและไม่มีสัญญาณรบกวน

คุณลักษณะเฉพาะที่ใช้ในแต่ละแบบจำลอง

Thresholding

พลังงานในย่านความถี่ 0.5-5 เฮิรตซ์

Random forest & ANN

- พลังงานในย่านความถี่ 0.5-5 เฮิรตซ์
- ขนาดสัญญาณสูงที่สุด
- Kurtosis ของค่าสัมประสิทธิ์เวฟเล็ตในแต่ละระดับชั้น
- ค่าเบี่ยงเบนมาตรฐานของค่าสัม-ประสิทธิ์เวฟเล็ตในแต่ละระดับชั้น

CNN

Spectrogram

การเลือกใช้ตัวแปรของแบบจำลองต่าง ๆ

Thresholding

Threshold : 90th percentile

Random forest

hyperparameter	
n_estimators	200
criterion	gini
max_depth	20
max_features	sqrt
min_samples_leaf	2
min_samples_split	2

Threshold : 50 percent

CNN

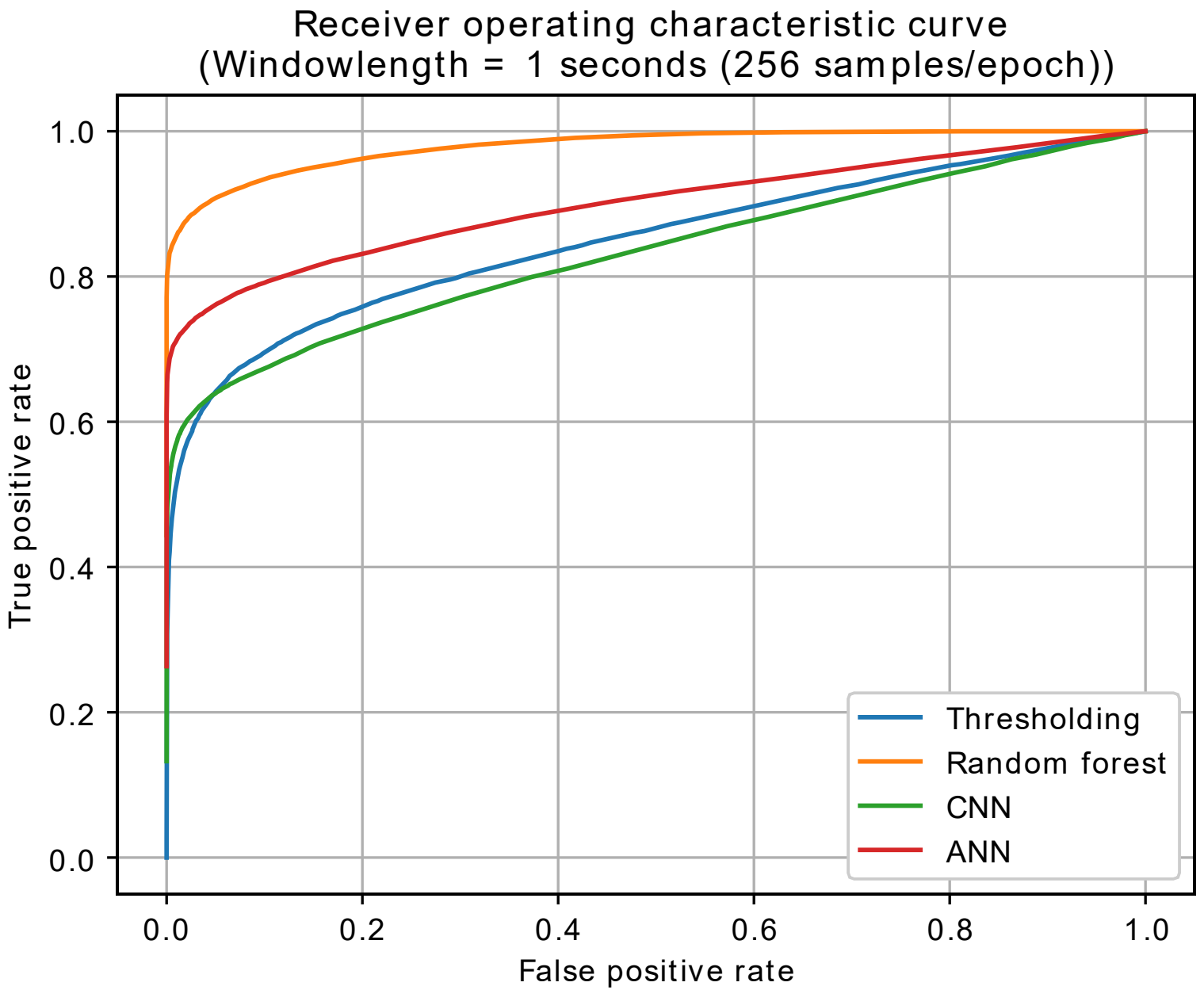
Model
C layer (3x3, 64filters)
C layer (3x3, 64filters)
MaxPooling layer
C layer (3x3, 128filters)
C layer (3x3, 128filters)
MaxPooling layer
Dense layer (4096nodes)
Dense layer (4096nodes)

Threshold : 50 percent

ANN

Model
Dense layer (256nodes)
Dense layer (128nodes)
Dense layer (64nodes)
Dense layer (64nodes)
Dense layer (64nodes)

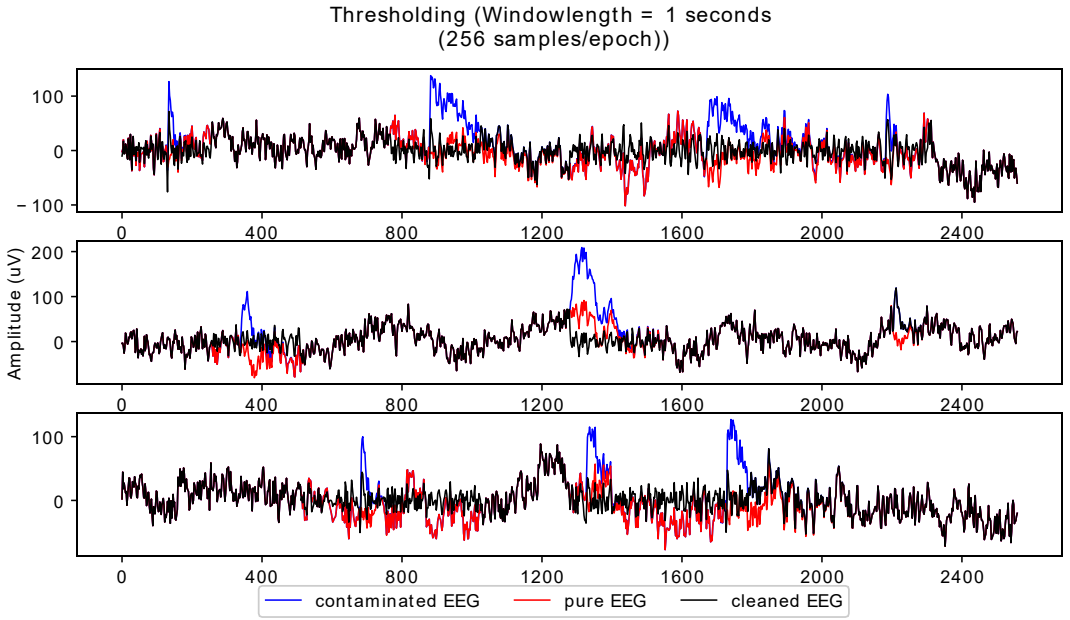
Threshold : 50 percent



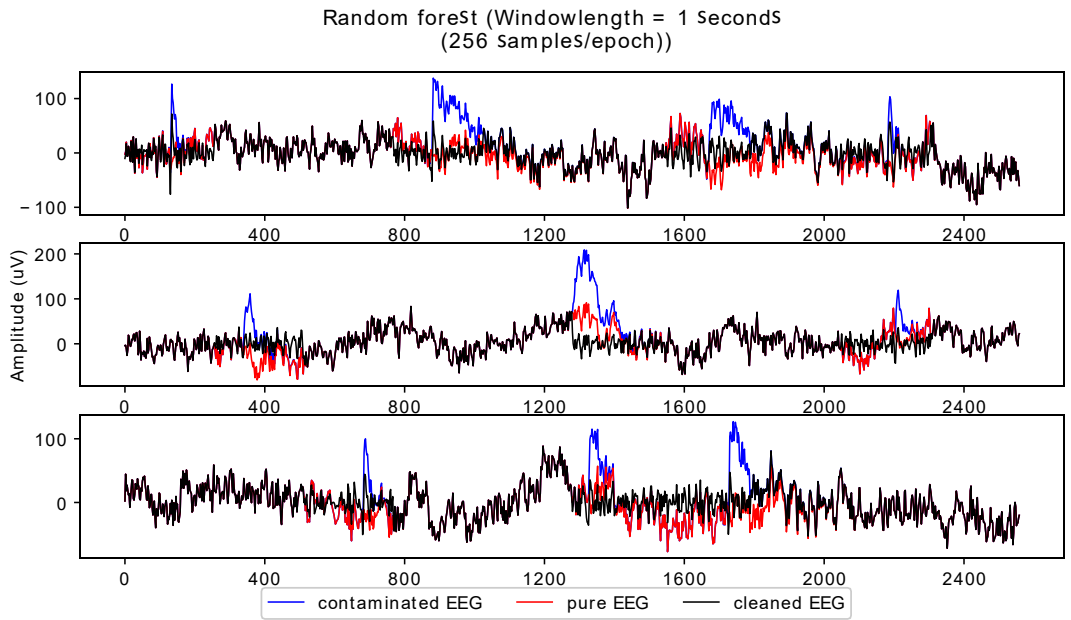
ROC curve ของแต่ละแบบจำลอง

Random forest เป็นแบบจำลองที่**มีความสามารถในการทำนาย**บริเวณที่มีสัญญาณรบกวน**สูงที่สุด**

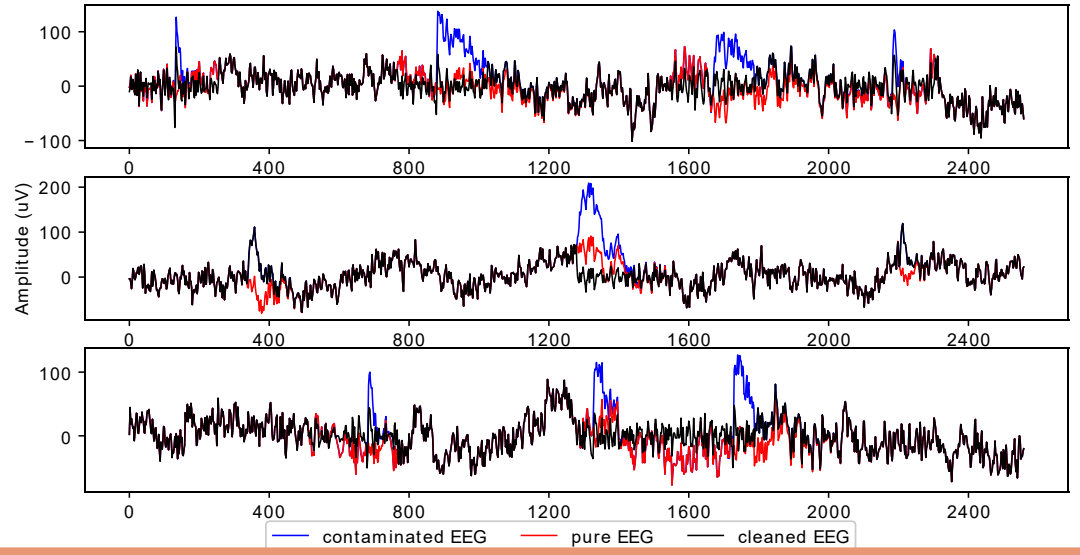
Thresholding



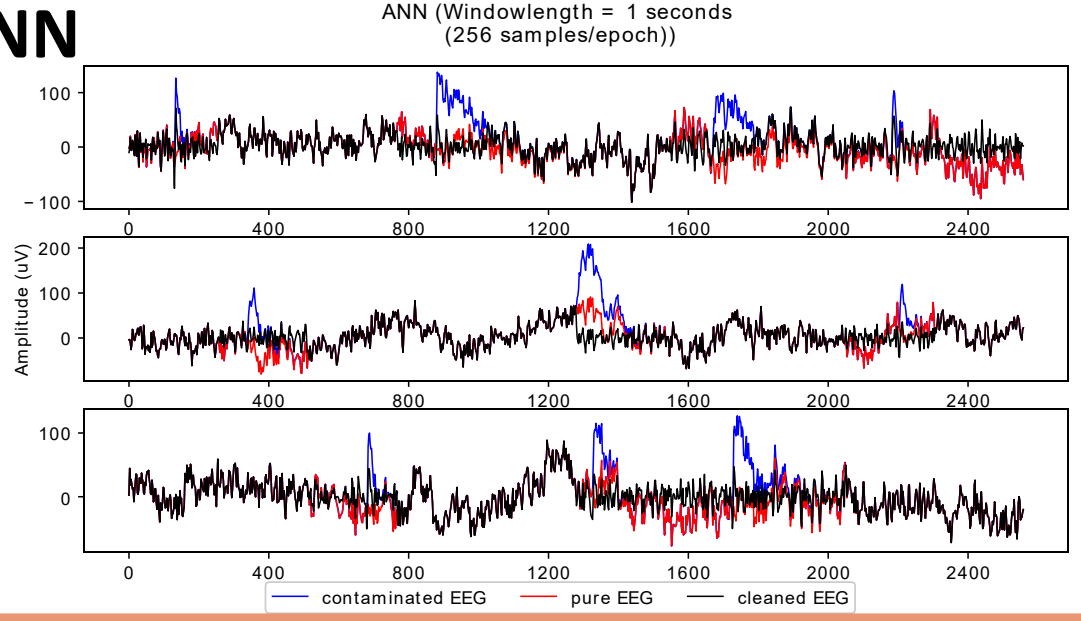
Random forest

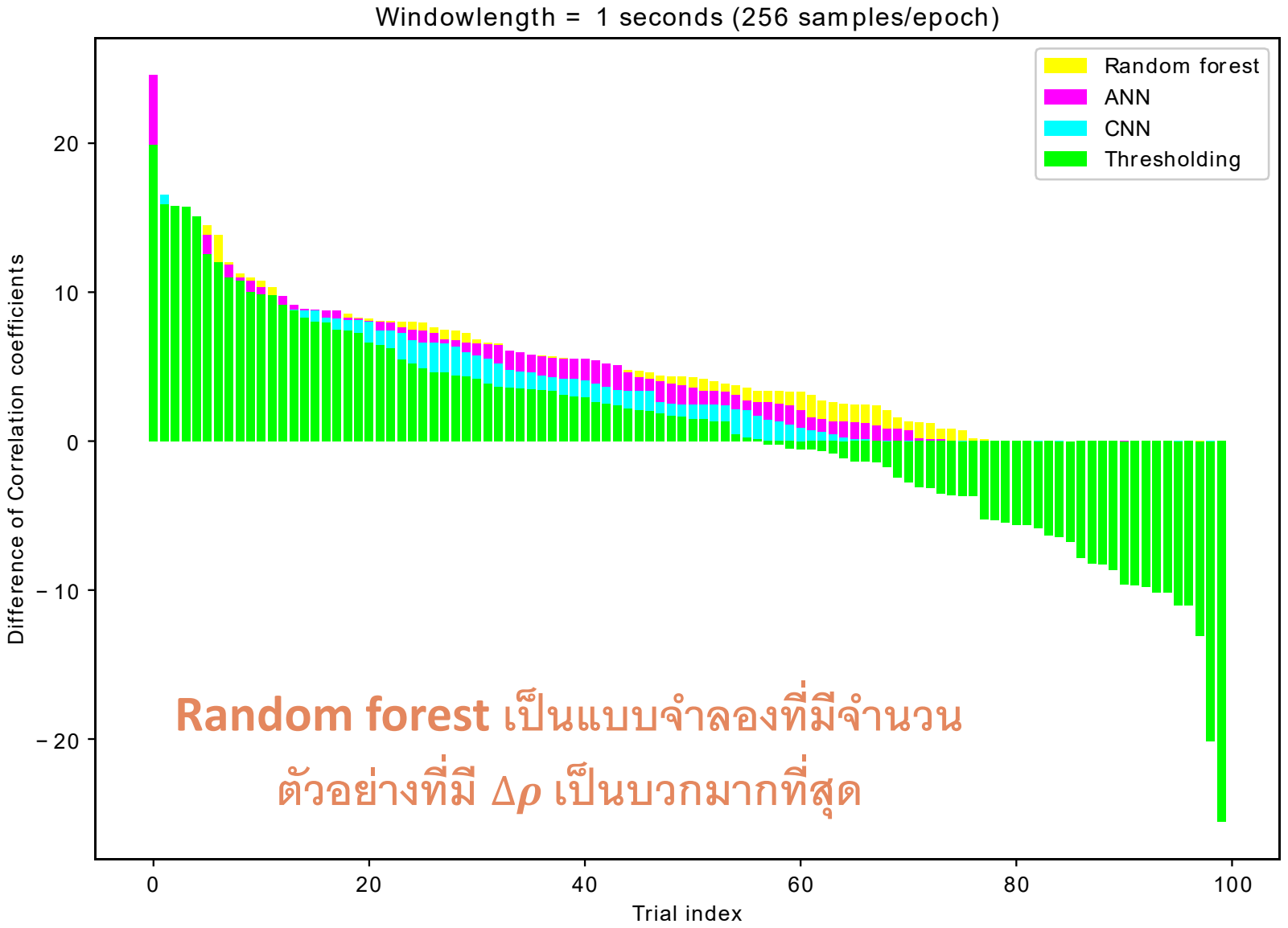


CNN



ANN





ความแตกต่างของค่าสัมประสิทธิ์สหสัมพันธ์ $\Delta\rho$

ค่า $\Delta\rho$ เป็นบวกมากแสดงว่าสัญญาณที่ถูกกำจัดสัญญาณรบกวนจะมีลักษณะทางกายภาพคล้ายกับสัญญาณคลื่นไฟฟ้าสมองที่ไม่มีสัญญาณรบกวนมาก

ค่าสมรรถนะของการตรวจจับและกำจัดสัญญาณรบกวน

	Test Accuracy (%)	Normalized Mean Squared Error (%)
Thresholding	81.09	29.50
Random forest	**87.94**	**24.29**
CNN	81.69	30.31
ANN	87.07	24.35

สรุปผลการทดลอง

- แบบจำลอง **Random forest มีสมรรถนะสูงที่สุด** เนื่องจากการที่จะกำจัดสัญญาณรบกวนได้ดี แบบจำลองจะต้องตรวจจับสัญญาณรบกวนได้ดีด้วย ซึ่งแบบจำลอง Random forest มีความแม่นยำในการตรวจจับสัญญาณรบกวนที่สูงที่สุด มี**ค่าความแม่นยำ (Test accuracy) เท่ากับ 87.94 เปอร์เซ็นต์** จึงส่งผลให้**ค่าความคลาดเคลื่อนกำลังสองเฉลี่ยแบบบรรทัดฐาน (NMSE) ต่ำที่สุด มีค่าเท่ากับ 24.29 เปอร์เซ็นต์**
- แบบจำลอง **CNN มีความแม่นยำที่ต่ำกว่า ส่งผลให้ค่าความคลาดเคลื่อนกำลังสองเฉลี่ยแบบบรรทัดฐานสูงกว่า** เนื่องจากข้อมูล**แผนภาพสเปกตรัมเพียงอย่างเดียวไม่เพียงพอ** แบบจำลองจะต้องใช้คุณลักษณะเฉพาะชนิดอื่นร่วมด้วย ดังที่แสดงให้เห็นในแบบจำลอง ANN

Q&A

อ้างอิง

[1] Wei-Yen Hsu, Chao-Hung Lin, Hsien-Jen Hsu, Po-Hsun Chen, and I-Ru Chen. Wavelet-based envelope features with automatic EOG artifact removal: Application to single trial EEG data. Expert Systems with Applications, 39(3):2743–2749, 2012.

[2] Nejedly, P. et al. Intracerebral EEG Artifact Identification Using Convolutional Neural Networks. Neuroinformatics. https://doi.org/10.1007/s12021-018-9397-6 (2018).

[3] Jose Antonio Urigüen and Begoña Garcia-Zapirain. EEG artifact removal—state of the art and guidelines. Journal of neural engineering, 12(3):031001, 2015.

Recap

การเปรียบเทียบวิธีการกำจัดสัญญาณรบกวนในสัญญาณคลื่นไฟฟ้าสมอง

ธนกฤต แสนเจริญกิจ

อาจารย์ที่ปรึกษา ผศ.ดร.จิตโกมุท ส่งศิริ

ภาควิชาวิศวกรรมไฟฟ้า คณะวิศวกรรมศาสตร์

จุฬาลงกรณ์มหาวิทยาลัย

- ข้อมูลสังเคราะห์
- สัญญาณรบกวนจากการกะพริบตา (ย่านความถี่ 0.5-5 เฮิรตซ์)

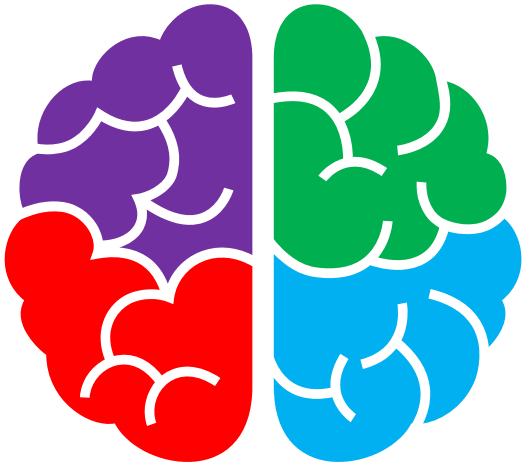
- การตรวจจับบริเวณที่มีสัญญาณรบกวน
  - ✓ Thresholding
  - ✓ Random forest
  - ✓ CNN
  - ✓ ANN
- การกำจัดสัญญาณรบกวน
  - ✓ Savitzaky-Golay filter

ประสิทธิภาพในการตรวจจับสัญญาณรบกวน

Random forest
ANN
CNN ≈ Thresholding

ประสิทธิภาพในการกำจัดสัญญาณรบกวน

- เพื่อศึกษาวิธีการกำจัดสัญญาณรบกวนในสัญญาณคลื่นไฟฟ้าสมองหนึ่งช่องสัญญาณ
- เพื่อเปรียบเทียบผลลัพธ์ของวิธีการกำจัดสัญญาณรบกวนแต่ละวิธี



การเปรียบเทียบวิธีการกำจัดสัญญาณรบกวนในสัญญาณคลื่นไฟฟ้าสมอง

Random forest is the **BEST!!!**

Q&A

Additional slide

ความรู้พื้นฐานเกี่ยวกับสัญญาณคลื่นไฟฟ้าสมอง [3]

- สัญญาณคลื่นไฟฟ้าสมอง จะมีย่านความถี่ในช่วง 0.01 – 100 Hz สามารถแบ่งเป็นช่วงต่าง ๆ ได้ดังนี้

ชื่อของช่วงความถี่	ช่วงความถี่ (Hz)
Delta	0.5 – 4
Theta	4 – 7
Alpha	8 – 13
Beta	14 – 30
Gamma	30 – 100

- มี Amplitude โดยทั่วไปไม่เกิน 100 uV
- สัญญาณคลื่นไฟฟ้าสมอง จะเป็น non-stationary processes

ขั้นตอนสังเคราะห์สัญญาณคลื่นไฟฟ้าสมองที่มีสัญญาณรบกวน

- สัญญาณคลื่นไฟฟ้าสมองที่ไม่มีสัญญาณรบกวน (Pink noise)
  (อัตราสุ่มของข้อมูลเท่ากับ 256 ตัวอย่างต่อวินาที)
- สัญญาณรบกวนที่เกิดจากการกะพริบตา

$$y[n] = c(e^{-bn} - e^{-an})$$

i. $90 \leq y[n] \leq 150$

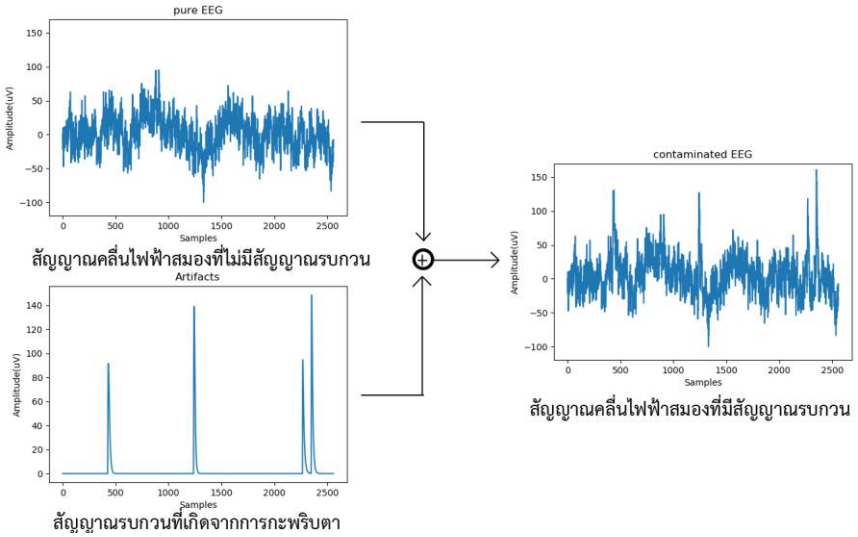
ii. $0.2 \times f_s \leq n_2 - n_1 \leq 2 \times f_s$

เมื่อ n_1 คือจุดเริ่มต้นของสัญญาณรบกวน

n_2 คือจุดที่ขนาดของสัญญาณรบกวนเป็น 1% ของ$90\ \mu V$

f_s คืออัตราการสุ่มของข้อมูล ซึ่งเท่ากับ 256 ตัวอย่างต่อวินาที

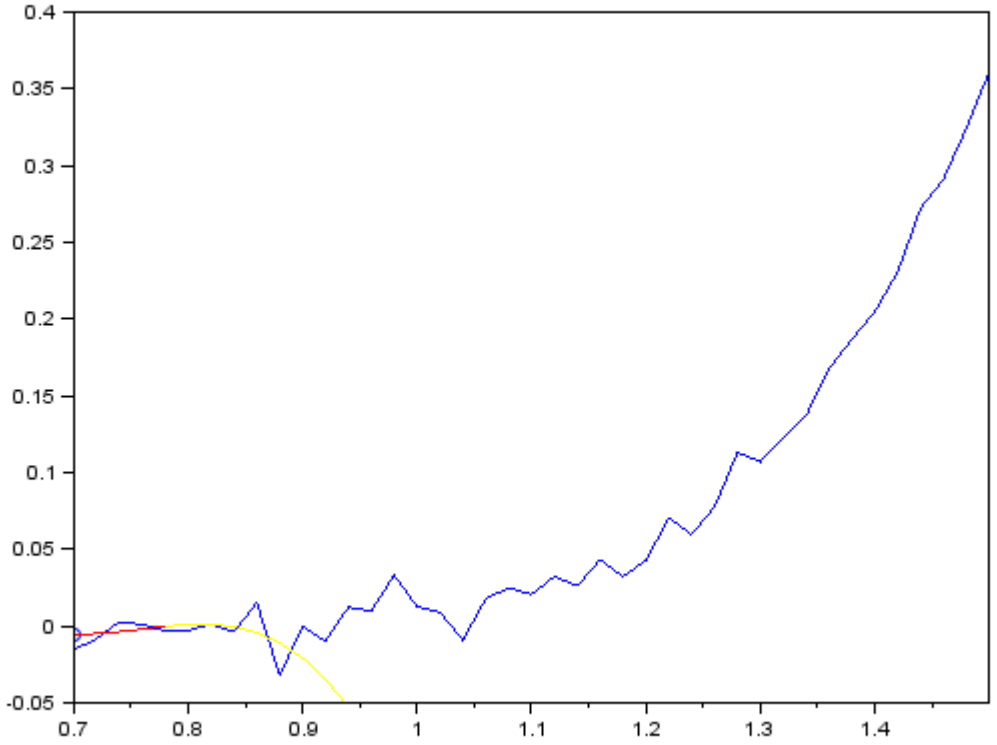
Pink noise



$$y[n] = c(e^{-bn} - e^{-an})$$

Savitzky-Golay filter

แหล่งที่มา : https://commons.wikimedia.org/wiki/File:Lissage_sg3_anim.gif



$$\hat{y}_j = \sum_{i=\frac{1-m}{2}}^{\frac{m-1}{2}} C_i y_{j+i} \quad , \frac{m-1}{2} \leq j \leq n - \frac{m-1}{2}$$

เมื่อ $\hat{y}_j$ คือ ค่าที่ถูกประมาณจากจุดข้างเคียง

n คือ จำนวนจุดของข้อมูลในบริเวณที่พบสัญญาณรบกวน

m คือ จำนวนจุดข้างเคียงที่นำมาใช้ประมาณ $\hat{y}_j$

y_{j+i} คือ ข้อมูลข้างเคียงที่นำมาใช้ประมาณ $\hat{y}_j$

C_i คือ สัมประสิทธิ์คอนโวลูชัน

Polynomial Degree	Quadratic or Cubic (2 or 3)			Quartic or Quintic (4 or 5)	
Window size	5	7	9	7	9
y_{j-4}			-21		15
y_{j-3}		-2	14	5	-55
y_{j-2}	-3	3	39	-30	30
y_{j-1}	12	6	54	75	135
y_j	17	7	59	131	179
y_{j+1}	12	6	54	75	135
y_{j+2}	-3	3	39	-30	30
y_{j+3}		-2	14	5	-55
y_{j+4}			-21		15
Normalization	35	21	231	231	429

ตัวอย่าง ค่าสัมประสิทธิ์คอนโวลูชัน

วิธีหาค่าสัมประสิทธิ์คอนโวลูชัน

จากคู่อันดับ (x_j, y_j)

แบบจำลองของ y คือ พหุนามอันดับที่ k $y_{i+j} = a_0 + a_1 z_i + a_2 z_i^2 + \ldots + a_k z_i^k$

โดย $z_i = \frac{x_{i+j} - x_j}{h}$ เมื่อ $h = x_{j+1} - x_j$

ถ้าจุดที่นำมาประมาณ y_j มีทั้งหมด m จุด $i = \frac{1-m}{2}, \frac{3-m}{2}, \ldots, 0, \ldots, \frac{m-3}{2}, \frac{m-1}{2}$

จะได้ $y = J\theta$ โดยที่แต่ละแถวของ J มีสมาชิก $1, z_i, z_i^2, \ldots, z_i^k$

$y = [y_{j+\frac{1-m}{2}} \;\; y_{j+\frac{3-m}{2}} \;\; \ldots \;\; y_{j+\frac{m-1}{2}}]^T$ สัมประสิทธิ์ $\theta = [a_0 \;\; a_1 \;\; \ldots \;\; a_k]^T$

วิธีหาค่าสัมประสิทธิ์คอนโวลูชัน

ต้องการ minimize $\|\hat{y} - y\|^2$

ซึ่งจะมีผลเฉลยคือ $\hat{\theta} = (J^T J)^{-1} J^T y$ แถวที่ 1 ของ $(J^T J)^{-1} J^T$ เท่ากับ $[C_{\frac{1-m}{2}} \ \dots \ C_0 \ \dots \ C_{\frac{m-1}{2}}]$

ซึ่ง y_j จะถูกประมาณจาก $\hat{y}_j = a_0 + a_1 z_0 + a_2 z_0^2 + \dots + a_k z_0^k$ ซึ่ง $z_0 = 0$

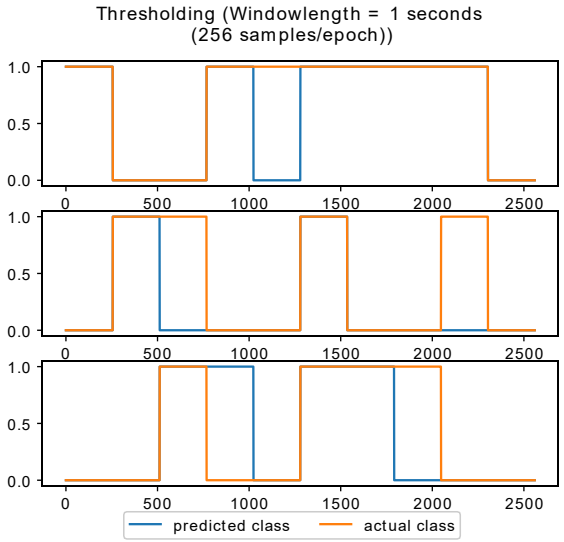
ดังนั้น $\hat{y}_j = a_0 = \hat{\theta} = C_{\frac{1-m}{2}} y_{j+\frac{1-m}{2}} + C_{\frac{3-m}{2}} y_{j+\frac{3-m}{2}} + \cdots + C_0 y_j + \dots + C_{\frac{m-3}{2}} y_{j+\frac{m-3}{2}} +$

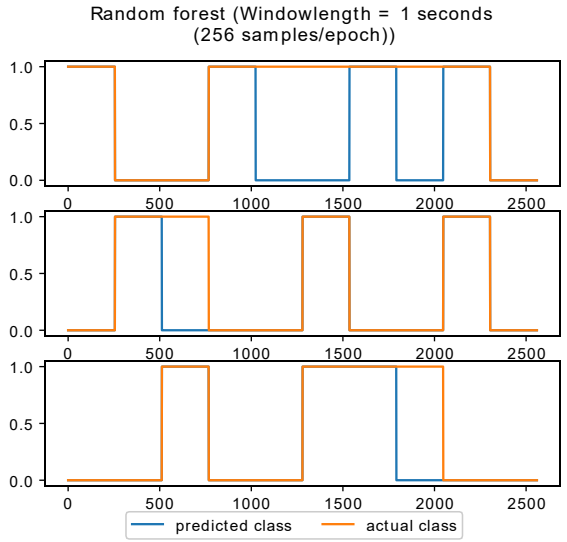
$C_{\frac{m-1}{2}} y_{j+\frac{m-1}{2}}$

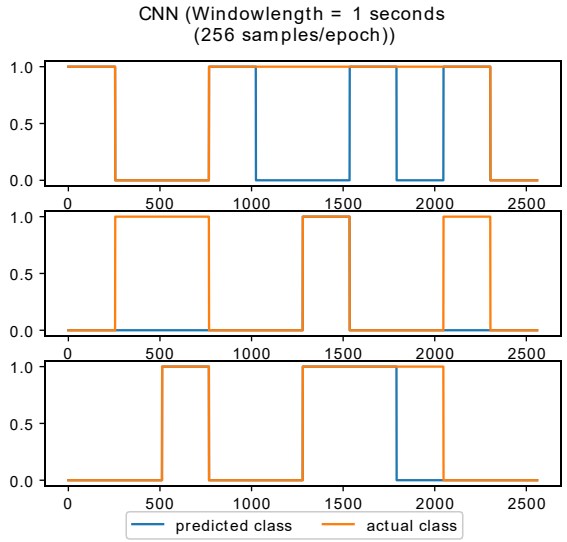
จึงได้ว่า $C_{\frac{1-m}{2}}, \dots, C_0, \dots, C_{\frac{m-1}{2}}$ คือสัมประสิทธิ์คอนโวลูชัน

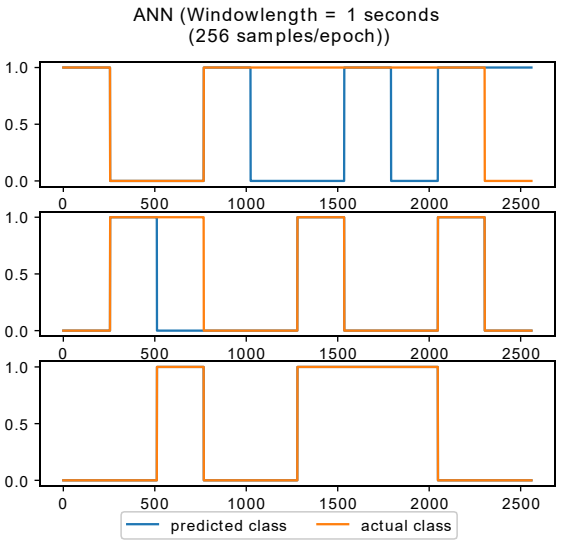
ตัวอย่างการตรวจจับบริเวณที่มีสัญญาณรบกวนของแต่ละแบบจำลอง

Thresholding **Random forest** **CNN** **ANN**









ค่าความแตกต่างของ Correlation coefficient

$$\Delta\rho_i = \widehat{\rho_i} - \rho_i$$

โดยที่ ρ_i คือ ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่าง สัญญาณคลื่นไฟฟ้าสมองที่***ไม่มีสัญญาณรบกวน***กับ สัญญาณคลื่นไฟฟ้าสมองที่***มีสัญญาณรบกวน ตัวอย่างที่ i***

$\widehat{\rho_i}$ คือ ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่าง สัญญาณคลื่นไฟฟ้าสมองที่***ไม่มีสัญญาณรบกวน***กับ สัญญาณคลื่นไฟฟ้าสมองที่***สัญญาณรบกวนถูกกำจัด ตัวอย่างที่ i***

งานที่จะทำในอนาคต

- ศึกษาคุณลักษณะเฉพาะที่จะช่วยให้แบบจำลองมีประสิทธิภาพสูงขึ้น
- ศึกษาวิธีการกำจัดสัญญาณรบกวนด้วยวิธีอื่น เนื่องจากตัวกรองสาวิทซกี-โกเลย์มีการประมาณสัญญาณ EEG ไปด้วยทำให้ค่า NMSE มีค่าเยอะ ในอนาคตจะหาตัวกรองที่สามารถประมาณสัญญาณรบกวนเพียงอย่างเดียว
- ศึกษาและพัฒนาให้แบบจำลองสามารถนำไปใช้แบบ real-time processing ได้ และนำไปประยุกต์ใช้กับ Brain-computer interface