Entropy-Assisted Mind Wandering Detection Using EEG Signals with XGBoost Classifier

Speaker: CYS
Advisor: Prof. An-Yeu Wu
Date: 2019/07/11

What is Mind-wandering?

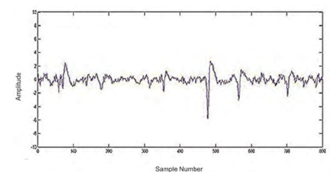
- A shift of attention away from a primary task toward internal information
  - Task-unrelated thought
  - Stimulus-independent thought
  - Mind pops
  - Zone outs

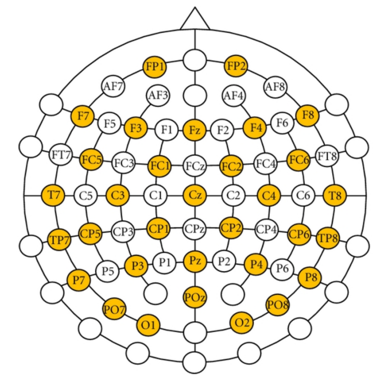
- Mind-wandering occurs anywhere between 20%-50%

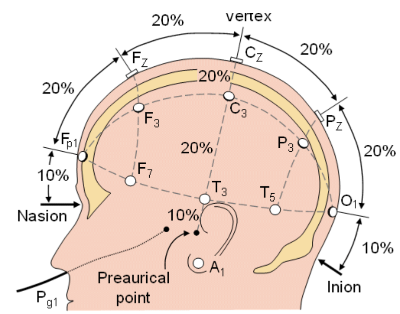
- Negative impact
  - Education
    - Degrade efficiency
  - Driving
    - Car accident

Electroencephalography (EEG)

- EEG is a technique that records the electrical activity of the brain
- Common application for EEG research
  - Social interaction
  - Psychology and neuroscience
    - Attention, learning, memory
  - Clinical and psychiatric studies
  - Brain computer interface

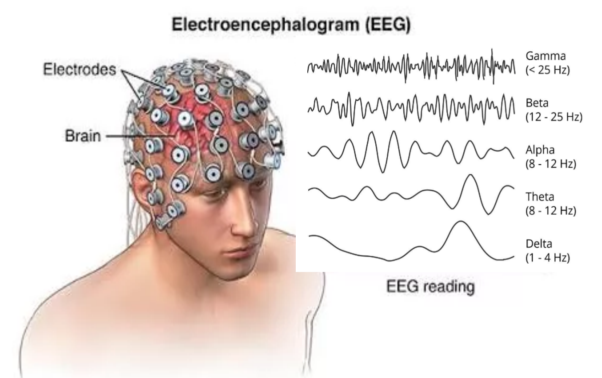






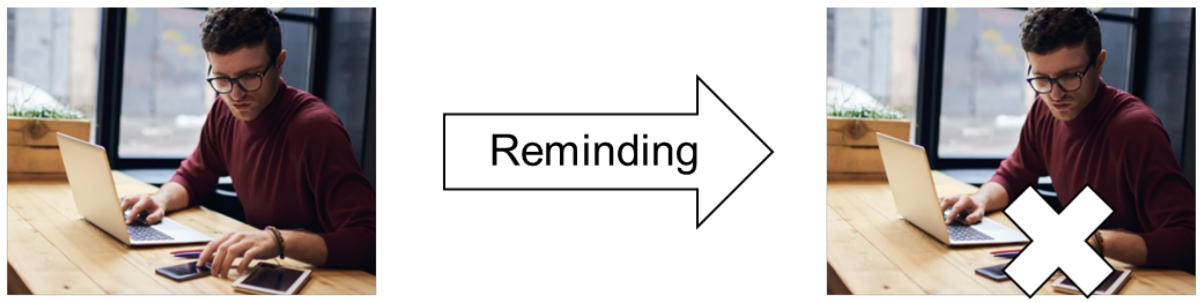
F=Frontal(額葉)
P=Parietal(頂葉)
T=Temporal(顳葉)
O=Occipital(枕葉)
C=Central(額葉中心溝)

Band	Remark
δ(2-3.5 Hz)	Deep sleep
θ(4-7 Hz)	Memory encoding and retrieval
α(9-11 Hz)	Quietly sitting in relaxed position with eyes closed
β(15-30 Hz)	Mental thought and activity



Problem Formulation

- Reminding from people help us to refocus to work, study



Attempt to use smartphone

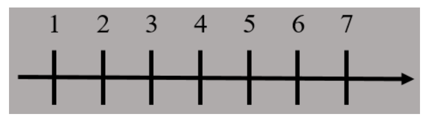
Focus On Work

- Collection of mind-wandering data

- Challenge of EEG mind wandering detector for continuously monitoring
  - Portable → least channel to get acceptable performance
  - Real-time → make the response as soon as possible
  - Low power → to continuously monitor mind wandering
  - Accurate → not to cause false alarm to influence user

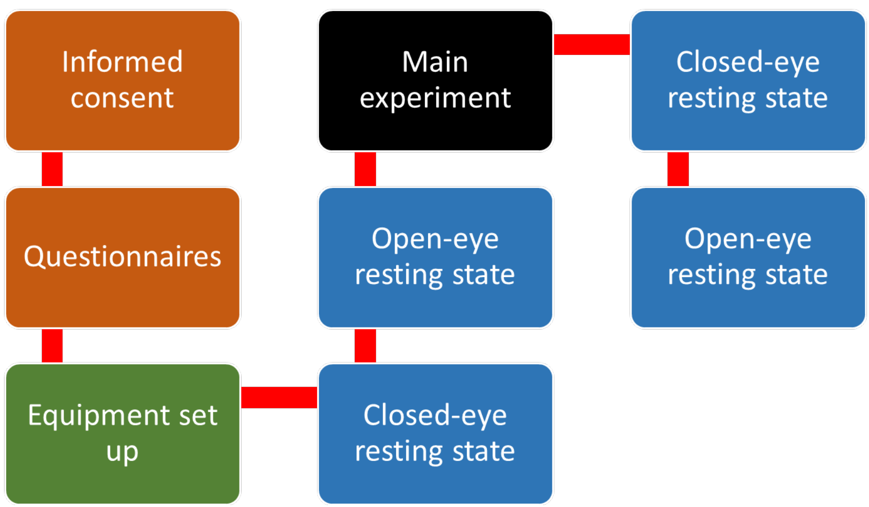
Sustained Attention to Response Task (SART)

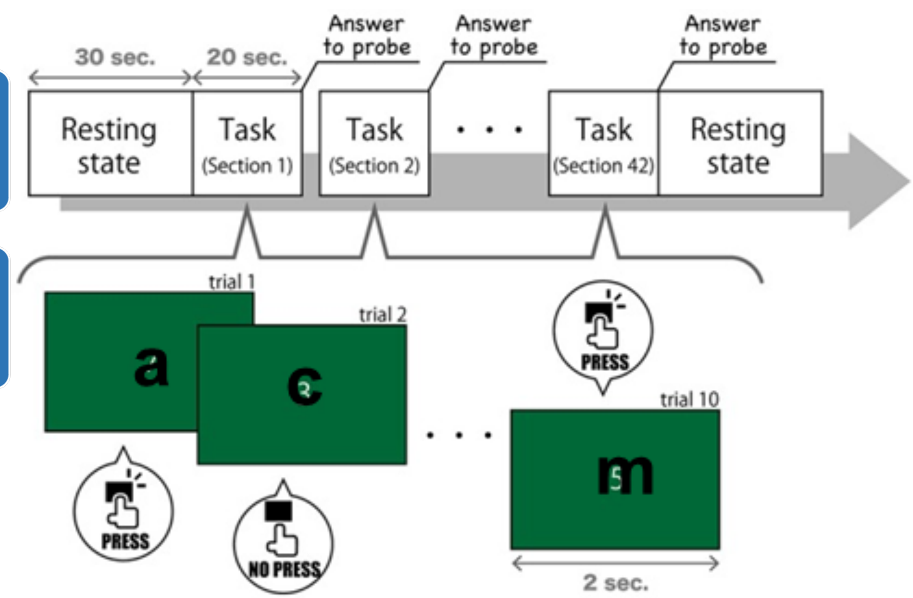
- Define 1 of 26 alphabets as target
  - Press when non-target digits appeared
- Dependent variables:
  - How focused are you?
  - What was in your mind?



1.專注在作業上
2.思考作業表現
3.被無關刺激吸引
4.思考無關的事
5.放空

- Experiment flowchart

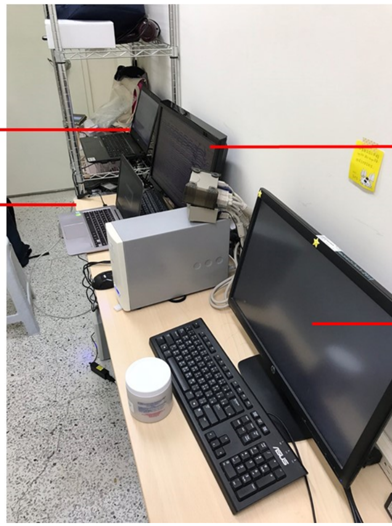




Experiment Instrument

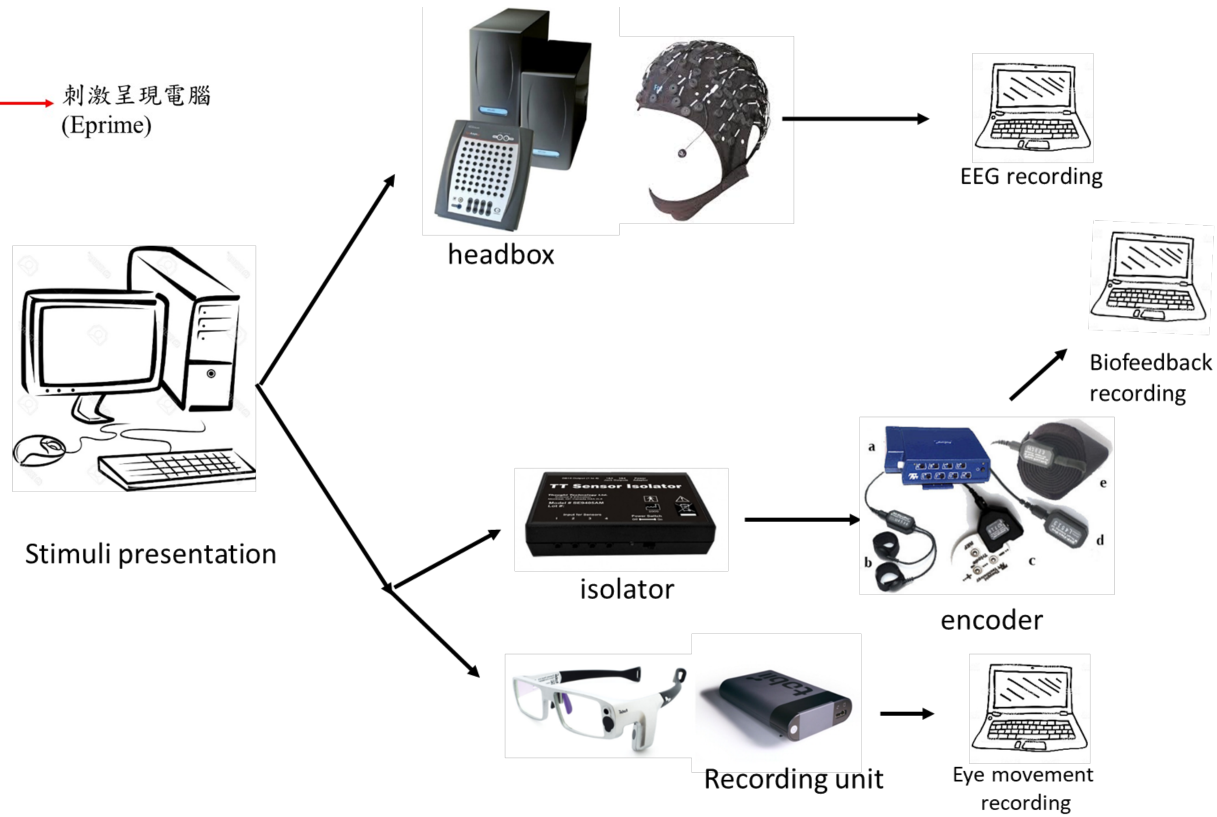
Infiniti 收錄電腦

Tobii Glass
controller電腦

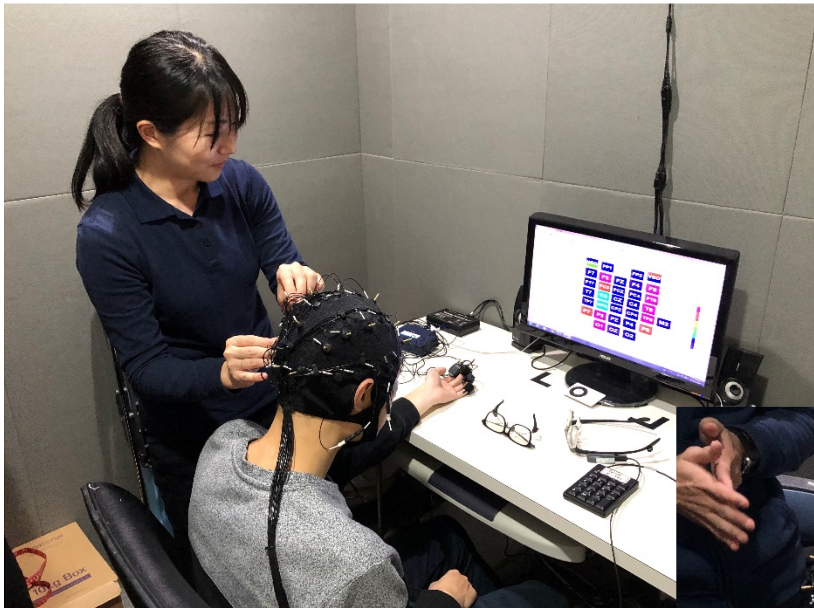


EEG 收錄電腦

刺激呈現電腦
(Eprime)



Experiments in NTU Psychology Dept.

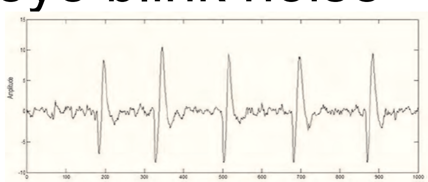


Comparison to Related Works

Related work

- Using EEG with lots of channels

- Independent component analysis (ICA) to remove eye blink noise



- Lack of feature variety
  - Extract $\theta, \alpha, \beta, \gamma$ power band
- Traditional classifier
  - SVM, KNN, Naive Bayesian

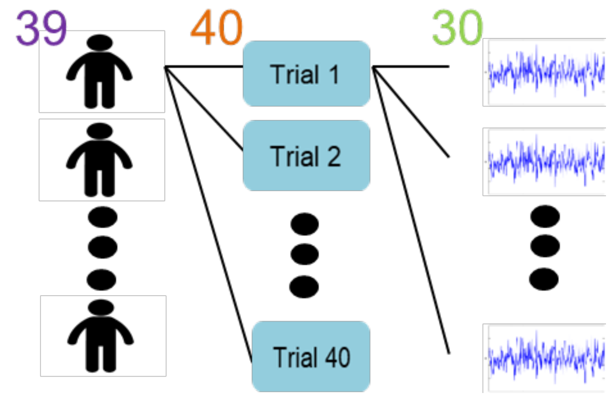
Ours

- Find the salient EEG channels first

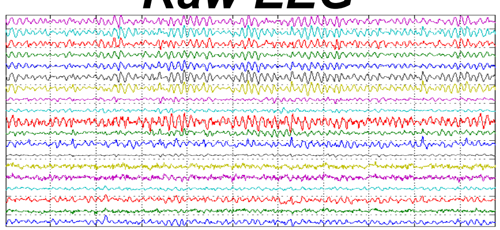
- No need of ICA to remove eye blink noise
- Extract non-linear features from three different domains
  - Find the substitution for multiscale entropy
- XGBoost classifier
  - Tree structure
  - Combining feature selection

Mind Wandering Prediction

- Data
  - 28-channel EEG + 2-channel EOG
- Label (how focused are you)
  - 0: self-rating < 4 (33.98%)
  - 1: self-rating > 4 (45.97%)



Raw EEG



→ **Preprocessing**: Bandpass Filter (4Hz-45Hz) | Epoch (10s before probe)

→ **Feature Extraction**: Time domain | Freq. domain | Wavelet domain

→ **Channel Selection**: p-value analysis

→ **Classification**: XGBoost

Feature Extraction

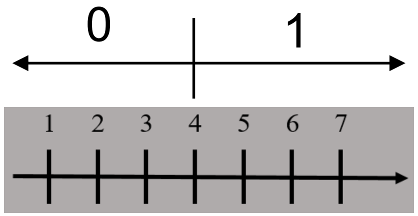
- Traditional features of EEG signals

	Time	Frequency	Wavelet
Basic	Mean, mean power, std, 1^{st} diff, 2^{nd} diff	$\theta, \alpha, \beta, \gamma$ power spectral density	Mean power, mean, std, a ratio of the absolute mean values of adjacent sub-bands.

- Non-linear entropy-related features

	Time	Frequency	Wavelet
Entropy	Multiscale permutation entropy, Multiscale dispersion entropy, Multiscale fluctuation-based dispersion entropy	Spectral entropy	Multiscale permutation entropy, Multiscale dispersion entropy, Multiscale fluctuation-based dispersion entropy

Channel Analysis



- Extracting ***40*** features in each channel
  - Frontal (F) and Central (C) are the salient channels

Channel (# of features p-value < .05)

Multiscale Entropy (MSE)

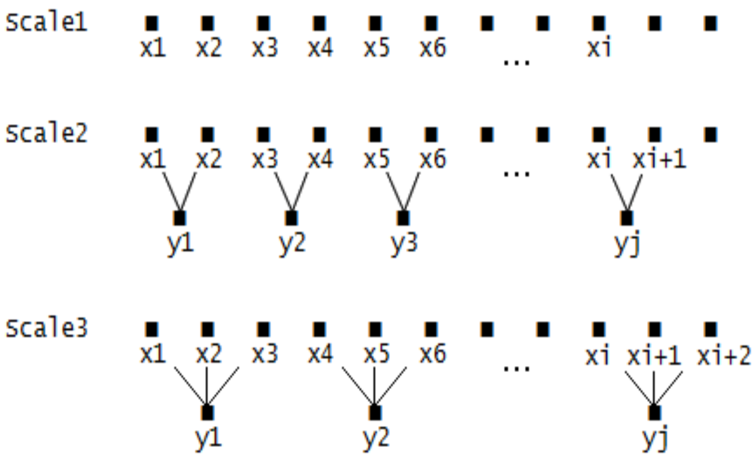
- Multiscale entropy (MSE) :
  - Complexity of time series
- Coarse graining (Mutliscale)
  - Non-overlapping average, total length N

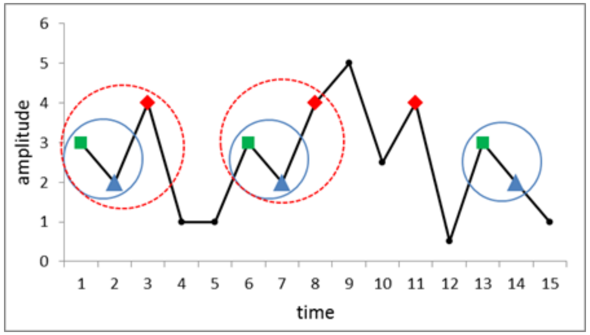
$$y_j^{(\tau)} = 1/\tau \sum_{i=(j-1)\tau+1}^{j\tau} x_i$$

τ: scale factor

  - τ set to 1~20, gain 20 new series
- Sample entropy calculation
  - Count the number of m -pt and $m+1$ -pt matches

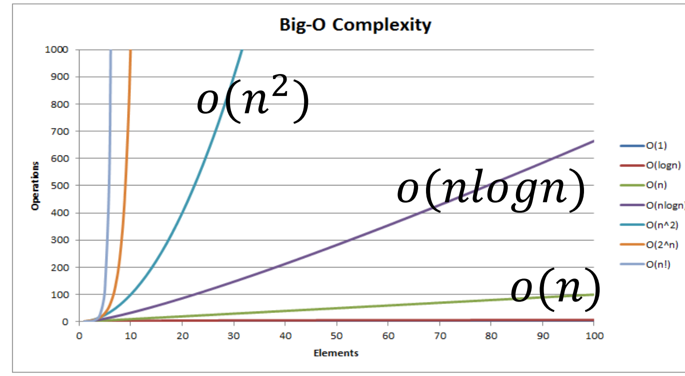
$$SampEn = -\ln\left(\frac{(m+1)\text{ - pt matches num}}{m\text{ - pt matches num}}\right)$$





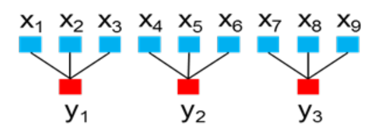
Multiscale Entropy (MSE)

- Widely used in biomedical applications
  - Diagnose depression
  - Parkinson's disease
  - Alzheimer's disease

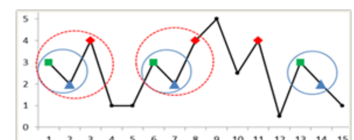


- Cons of MSE
  - High computational time **O(n²)** in calculating sample entropy
  - Lead to **undefined value** in short series
    - Hard to apply to wavelet domain

Coarse-graining of MSE



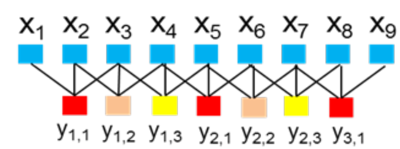
SampEn calculation of MSE



$$SampEn = -\ln\frac{n_\tau^{m+1}}{n_\tau^m}$$

- Cons of RCMSE [2]
  - **Higher** computational time than MSE
  - Still have undefined-value problem

Coarse-graining of RCMSE



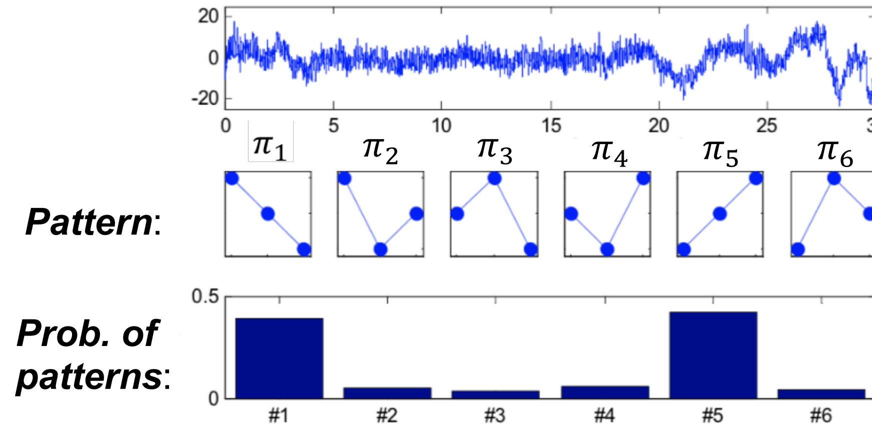
SampEn calculation of RCMSE

$$SampEn = -\ln\left(\frac{\sum_{k=1}^{\tau} n_{k,\tau}^{m+1}}{\sum_{k=1}^{\tau} n_{k,\tau}^{m}}\right)$$

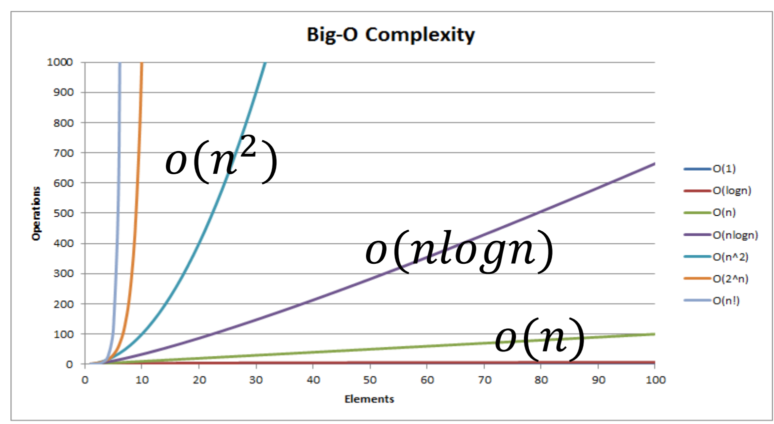
Permutation Entropy (PE)

- Quantifying complexity based on ordinal analysis
  - Step 1: Calculate probability of each pattern with length m
  - Step 2: Shannon entropy calculation

$$PermEn(x,m) = -\sum_{j=1}^{m!} p(\pi_j) \ln p(\pi_j)$$

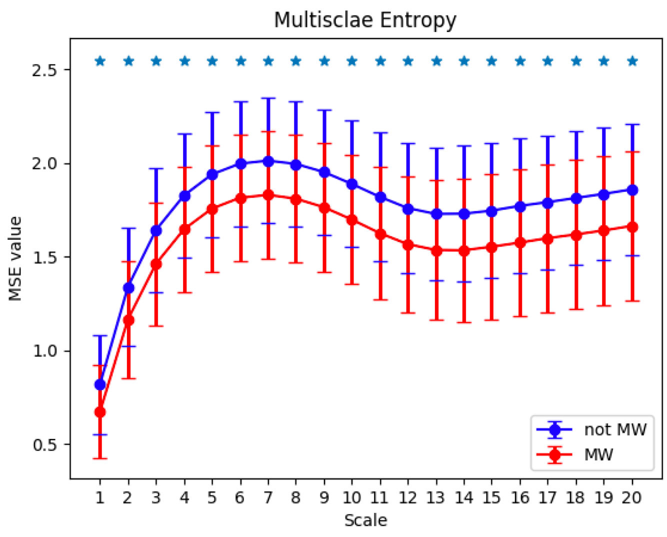


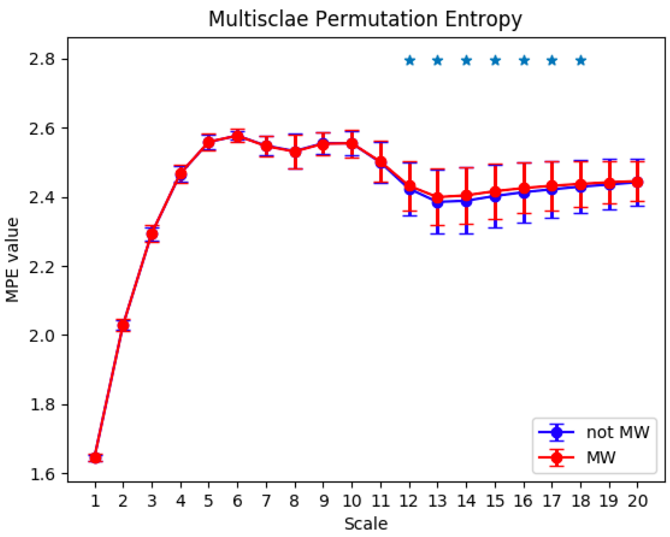
- Pros of permutation entropy
  - Computational time: O(n)
    - Much faster than MSE
  - No undefined value
- Cons of permutation entropy
  - Impact of equal value
  - Not consider difference between values
  - Not consider extremely high and low values



Comparison to MSE and MPE

- The number of features with p-value < 0.05 in Fz channel
  - Multiscale entropy: 20/20
  - Permutation entropy: 4/20

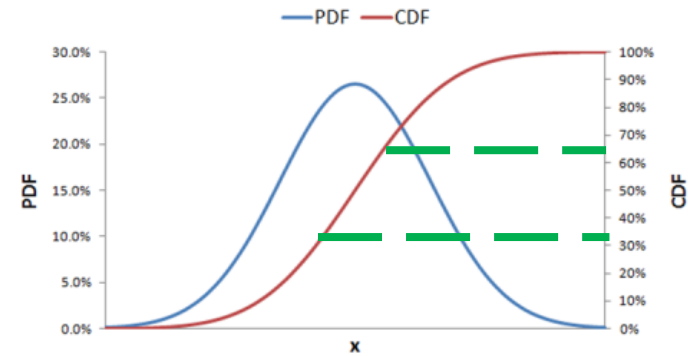




⋆ : p-value < 0.05

- Although permutation entropy have lower computation time, the performance is not good
- Entropy of mind wandering is lower than it of not mind wandering
  - EEG's complexity is higher when people concentrate on a task

Dispersion Entropy (DE)



- Quantify the regularity of time series
- Tackle limitations of SampEn, PermEn
- Calculation steps
  - Mapped sample point to ***c*** classes
    - Normal cumulative distribution function
    - Linearly mapping
  - Calculate dispersion pattern, if length=2
    - Pattern type: (1,1),(1,2),(1,3),(2,1),(2,2),(2,3),(3,1),(3,2),(3,3)
  - Shannon entropy calculation of patterns

Sample points: 9, 8, 1, 12, 5, -3, 1.5, 8.01, 2.99, 4, -1, 10

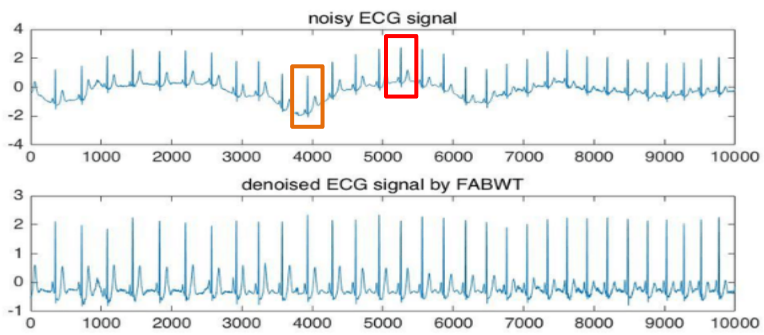
After class mapping: 3, 3, 1, 3, 2, 1, 1, 3, 2, 2, 1, 3

$$DE = -\sum_{n=1}^{k} p(pattern_n) ln(pattern_n)$$

- Computational time: **O(n)**

Fluctuation-based Dispersion Entropy (FDE) [5]

- Robust to the presence of baseline wanders and trends



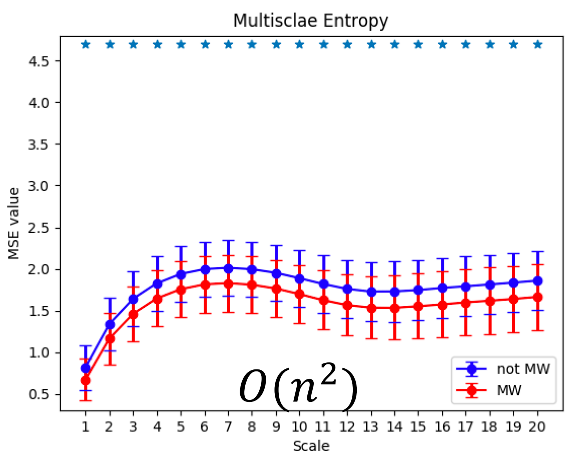
- When decide patterns
  - Define pattern by order (relative value)
  - For example: {1,5,1}, {4,8,4}
    - In DE: they are different pattern
    - In FDE: they are the same pattern (local de-trend)

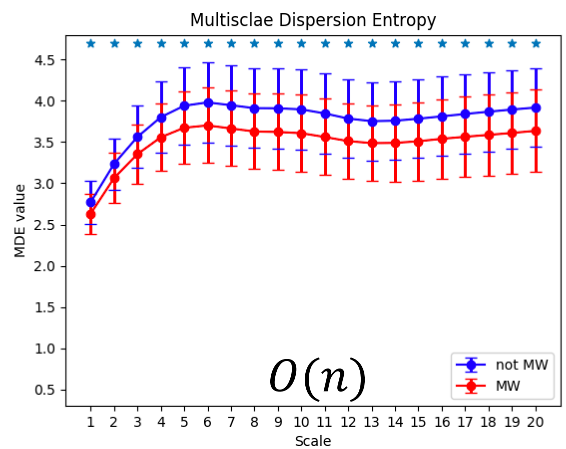
- Computational time: **O(n)**

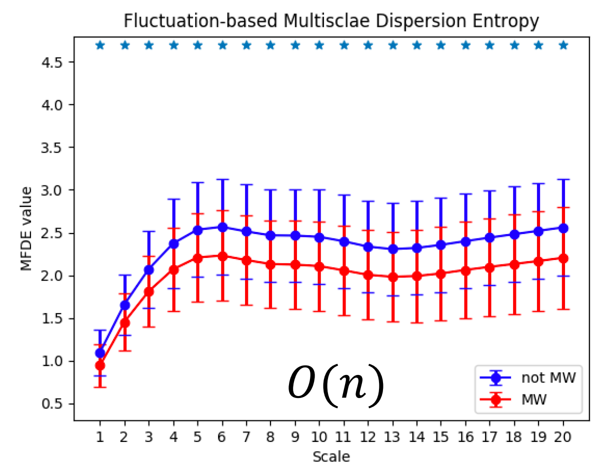
- Solve the problem of
  - High computational time in **sample entropy calculation (MSE)**
  - Not consider relative difference between value in **permutation entropy (PE)**
  - Trend and baseline waders noise in **dispersion entropy (DE)**

MSE v.s. MDE [6] v.s. MFDE

- The number of features with p-value < 0.05 in Fz channel
  - Multiscale entropy: 20/20
  - Multiscale dispersion entropy: 20/20
  - Fluctuation-based multiscale dispersion entropy: 20/20





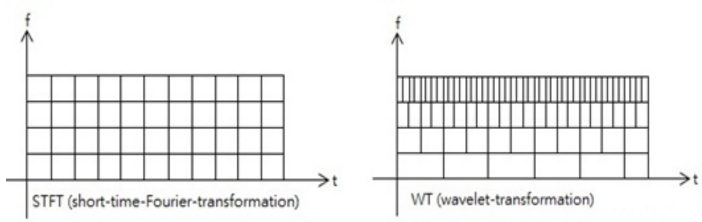


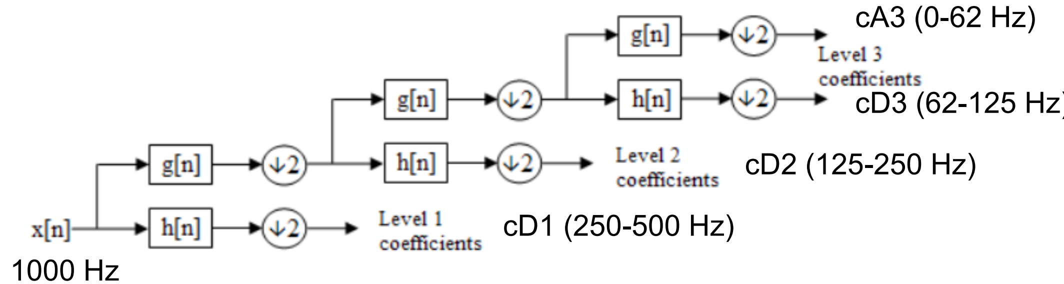
$*$: p-value < 0.05

- Multiscale fluctuation-based dispersion entropy
  - Faster computational time
  - Similar performance to MSE
  - Remove local trend

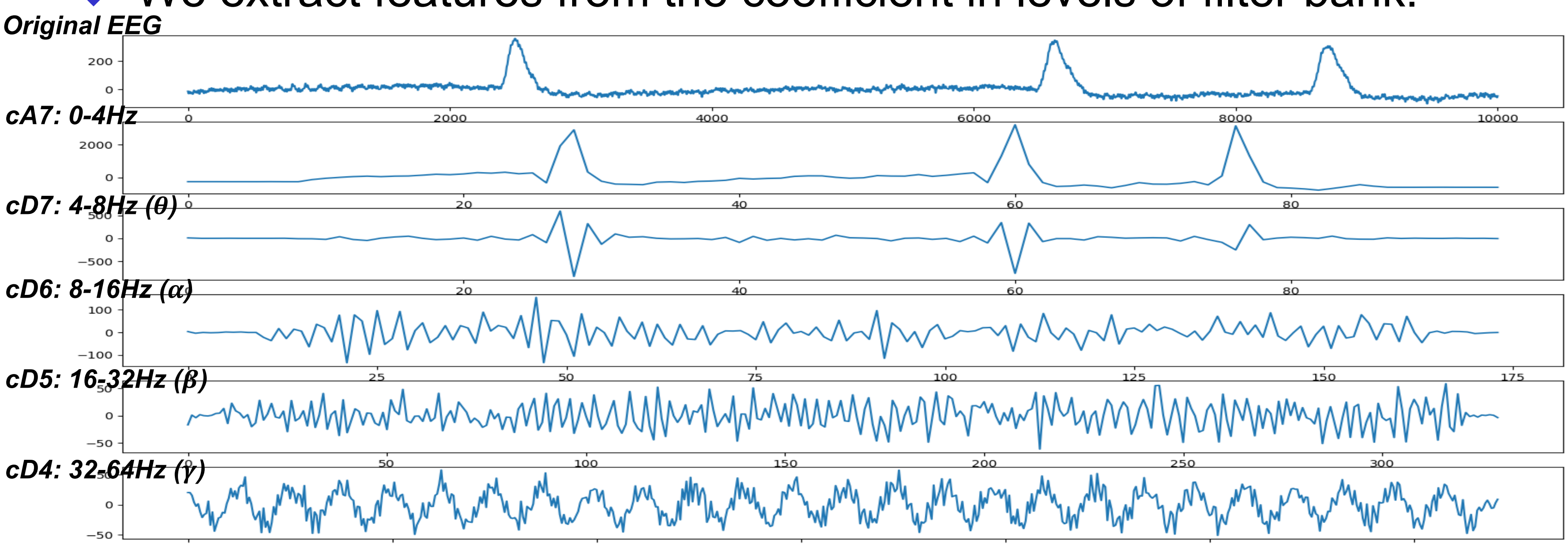
Wavelet Decomposition

- Discrete wavelet transform (DWT) can decompose the signals into frequency sub-bands





- We extract features from the coefficient in levels of filter bank.

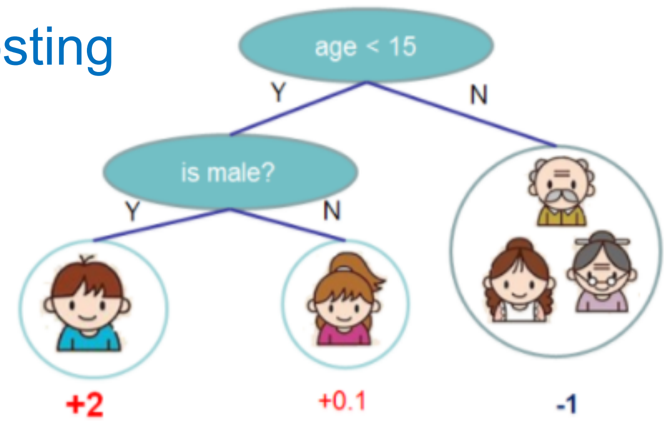


c: coefficient, A: Approximate, D: Detail

Extreme Gradient Boosting (XGBoost)

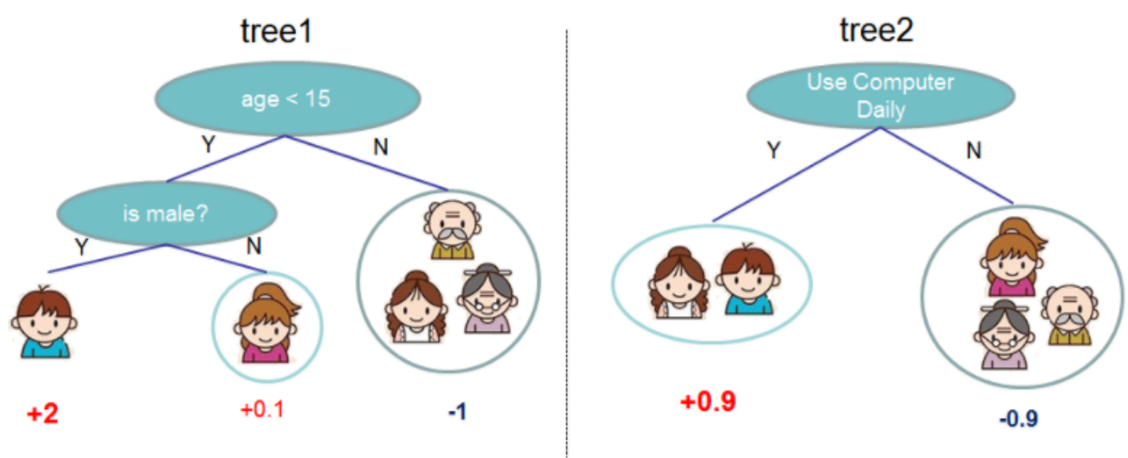
- XGBoost = Tree + Ensemble + Gradient Boosting
- Tree
  - Classification & Regression Tree (CART):
    - Both for classification and regression

$$\hat{y}_i = f(x_i)$$



- Ensemble

$$\hat{y}_i = \sum_{k=1}^{K} f_k(x_i), f_k \in F$$



f() = 2 + 0.9= 2.9 f()= -1 - 0.9 = -1.9

Extreme Gradient Boosting (XGBoost) (2/2)

Gradient Boosting

- Use gradient descent to optimize objective function
- Objective function = training loss + regularization

$$obj^{(t)} = \sum_{i=1}^{n} l\left(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)\right) + \Omega(f_t)$$

  - $l(\ )$: Loss function (least square, **cross entropy**)
  - y_i : Ground truth
  - $\hat{y}_i^{(t-1)}$: So far we did
  - $f_t(x_i)$: Next tree we would like to optimize
  - $\Omega(f)$: Regularization term

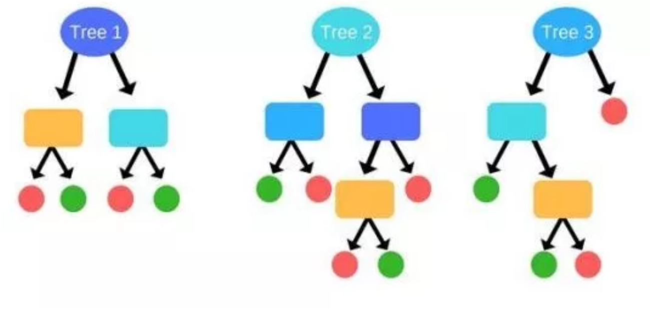
Taylor's Expansion

$$f(x+\Delta x) \approx f(x) + f'(x)\Delta x + \frac{1}{2} f''(x)\Delta x^2 \quad (1)$$

$$obj^{(t)} = \sum_{i=1}^{n} [\, l\left(y_i, \hat{y}_i^{(t-1)}\right) + g_i f_t(x_i) + \frac{1}{2} h_i f_t^{\,2}(x_i)] + \Omega(f_t) + \text{constant}$$

$$g_i = f'(x), h_i = f''(x)\ in\ (1)$$

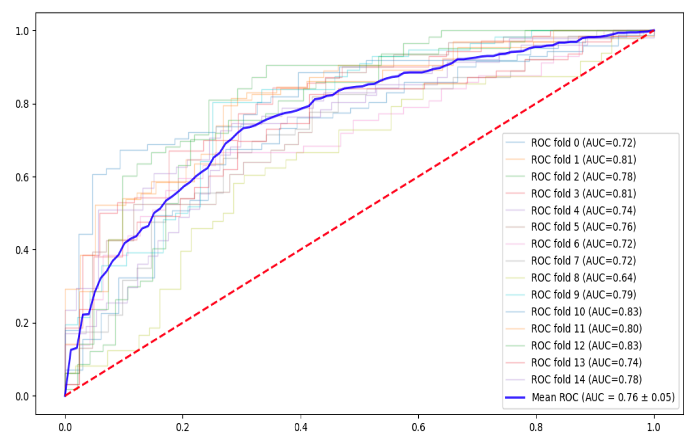
Classification Results



- Used channel: Fz, Fp1, Fp2
- Select top 15 in each scheme
  - According to the number of features used in XGBoost trees

Scheme	F1 Score (# of features)
Basic features	0.72 (28*3) 0.71 (15)
Entropy features	0.79 (395*3) 0.76 (15)
Basic + Entropy features	0.80 (423*3) 0.76 (15)

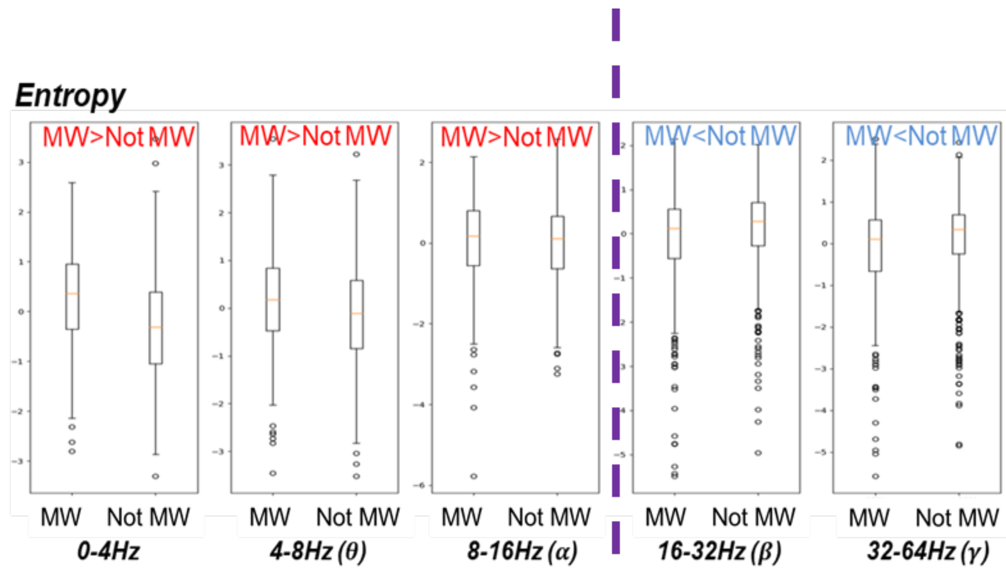
ROC curve using 15 Entropy features

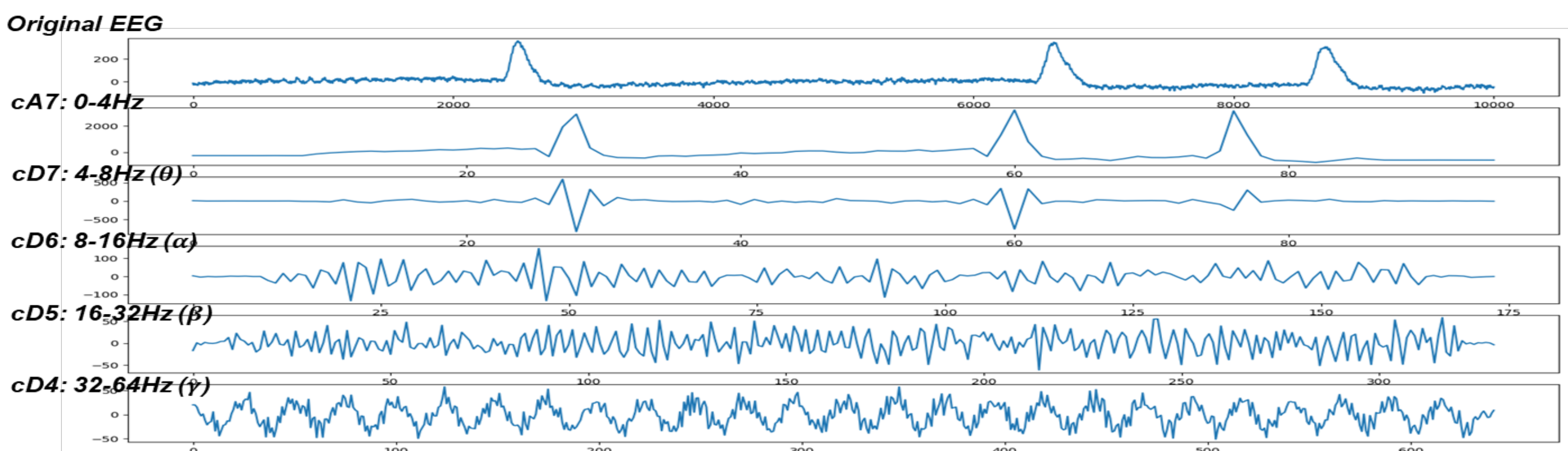


- Entropy-related features perform better than basic features set

Entropy Features in Wavelet Domain

- In STFT analysis
  - MW -> low entropy
- In wavelet analysis
  - Low frequency component
    - MW -> high entropy
  - High frequency component
    - MW -> low entropy





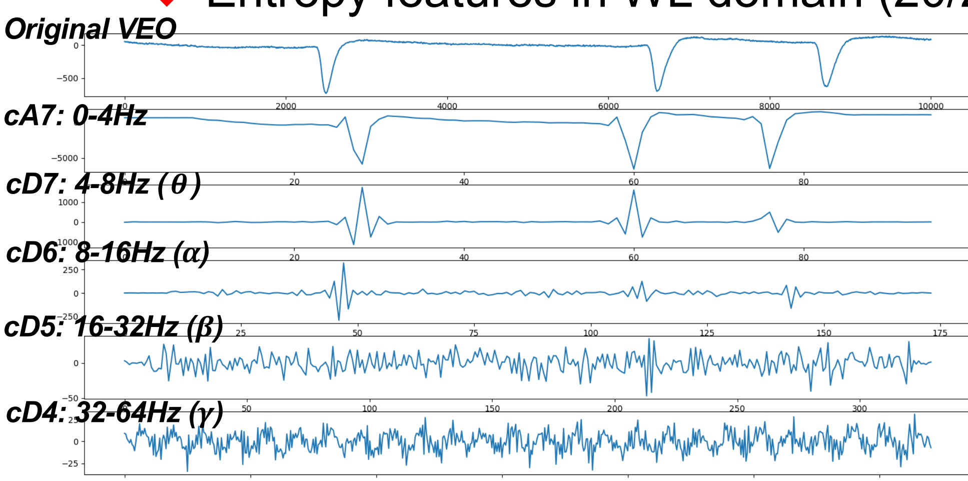
c: coefficient, A: Approximate, D: Detail

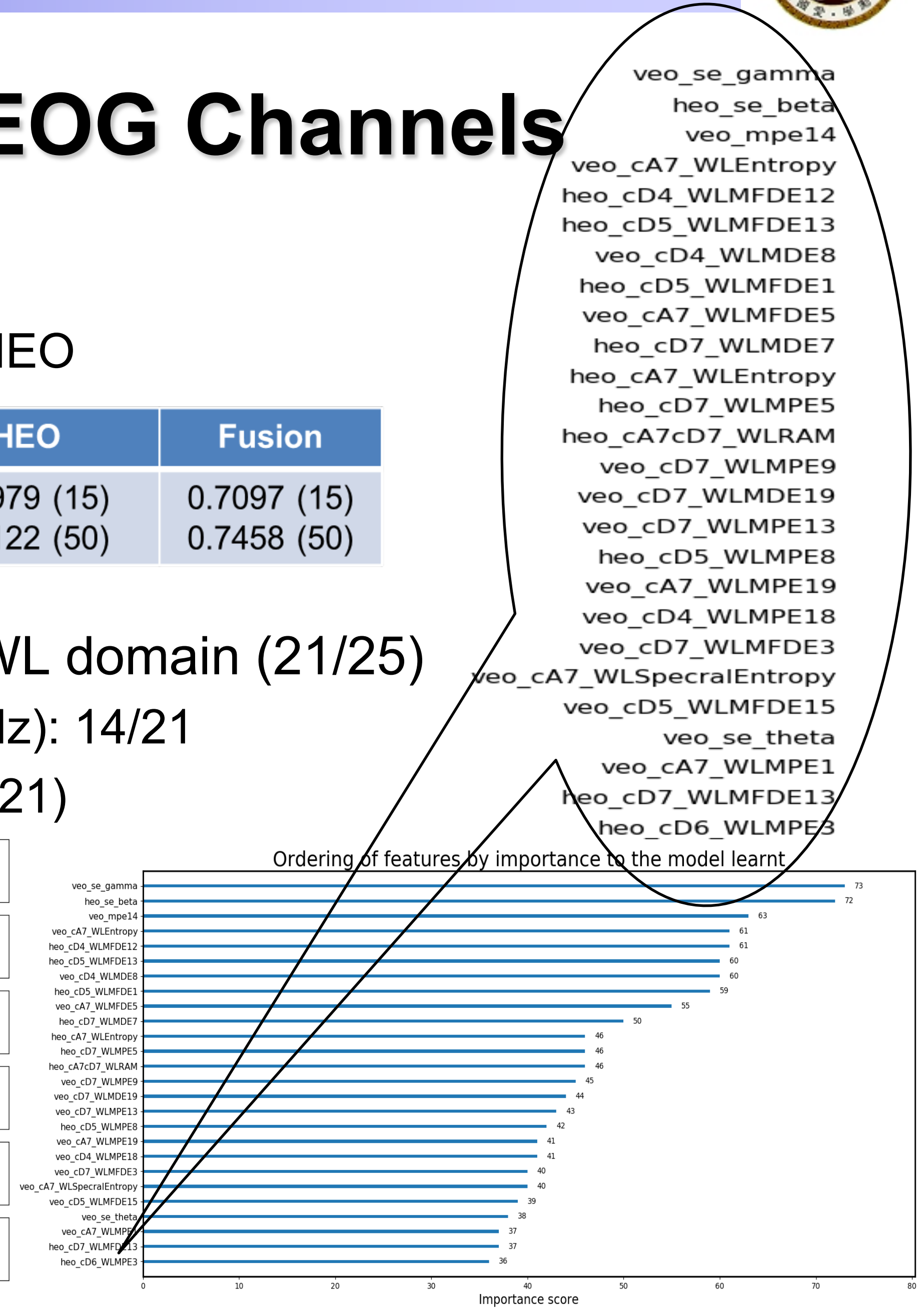
Results Using EOG Channels

- EOG can reach 0.7458 F1 score
  - F1 score of VEO is 2% more than HEO

Scheme	VEO	HEO	Fusion
Basic + Entropy Features	0.7183 (15) 0.7378 (50)	0.6979 (15) 0.7122 (50)	0.7097 (15) 0.7458 (50)

- The significant features are most WL domain (21/25)
  - Especially in lower frequency (<16Hz): 14/21
  - Entropy features in WL domain (20/21)





Result of Single-Channel EEG and EOG

- Combining EEG and EOG to get the brain and eye information

Scheme	3-channel EEG	2-channel EOG	Fz	Fz + 2 EOG
Basic + Entropy features	0.6932 (15) 0.7291 (50)	0.7097 (15) 0.7458 (50)	0.6743 (15) 0.7060 (50)	**0.7309 (15)** **0.7535 (50)**

- Top 50 features selected from Fz and 2 EOG

Channel	# of basic features	# of entropy features	Total # of features
Fz	1	17	18
VEO	0	16	16
HEO	2	14	16

  - The importance of each channel are almost equal
  - Entropy features are dominant

Conclusion

- To the best of our knowledge, we first apply entropy features on MW detection task
  - When mind wandering, the entropy of EEG is lower
- Find the substitution (DE) for MSE solve the problem of
  - High computational time
  - Undefined value
- Apply entropy feature on wavelet domain
  - Low-frequency component have better performance
- EOG signals are related to mind wandering
  - Vertical EOG perform better than horizontal EOG → eye blink is vertical direction
- The best result is achieved by combining single-channel EEG and EOG