UVIC
UNIVERSITAT
DE VIC

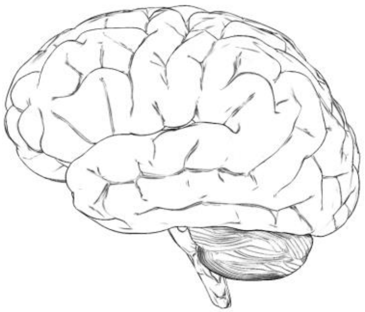
Diagnosis of Alzheimer's disease from EEG using machine learning algorithms

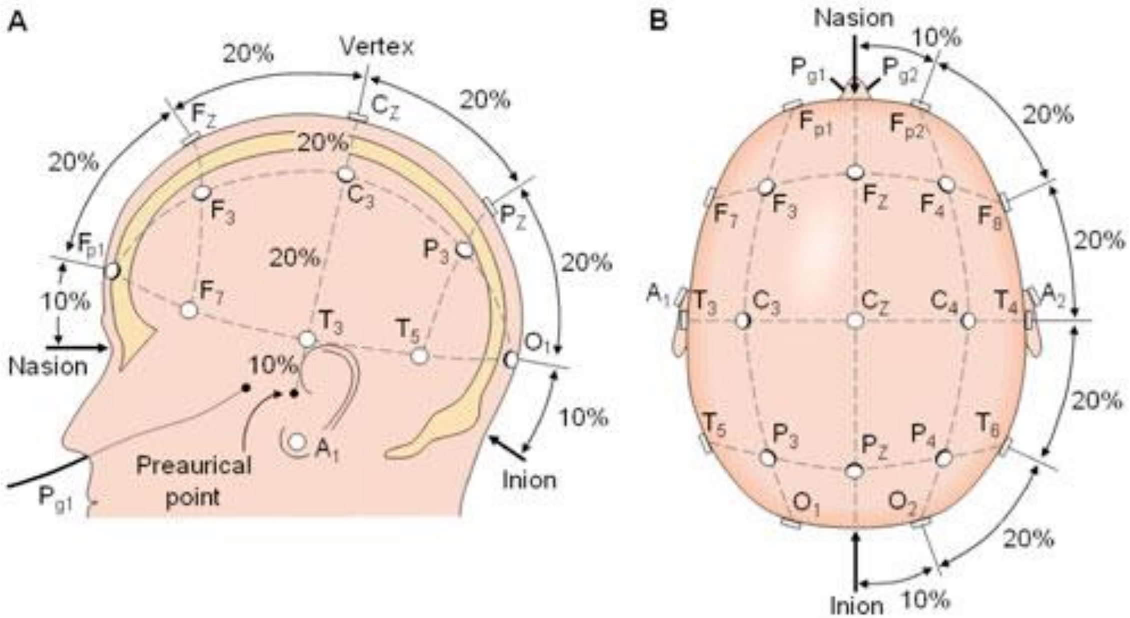
Dr. Jordi Solé-Casals

Brno University of Technology
Czech Republic
3rd SPLab Workshop 2013

Overview

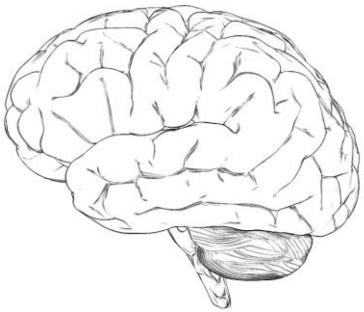
Signal acquisition





Overview

Signal acquisition → Signal Processing



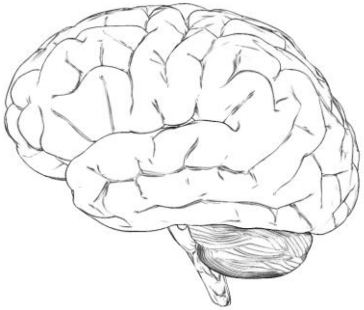
Parameterize the EEG signals with different measures:

- Frequency Power measure
- Synchrony measures

Select the best frequency range

Overview

Signal acquisition → Signal Processing → Classification

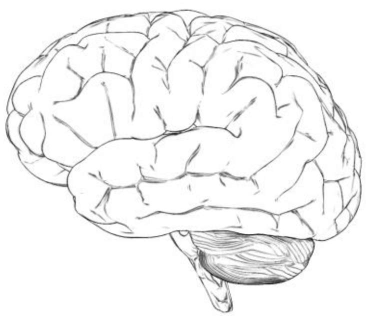


Use these features to build a classifier:

- Each measure individually
- Selecting the best measures and frequency ranges

Overview

Signal acquisition → Signal Processing → Classification → Application



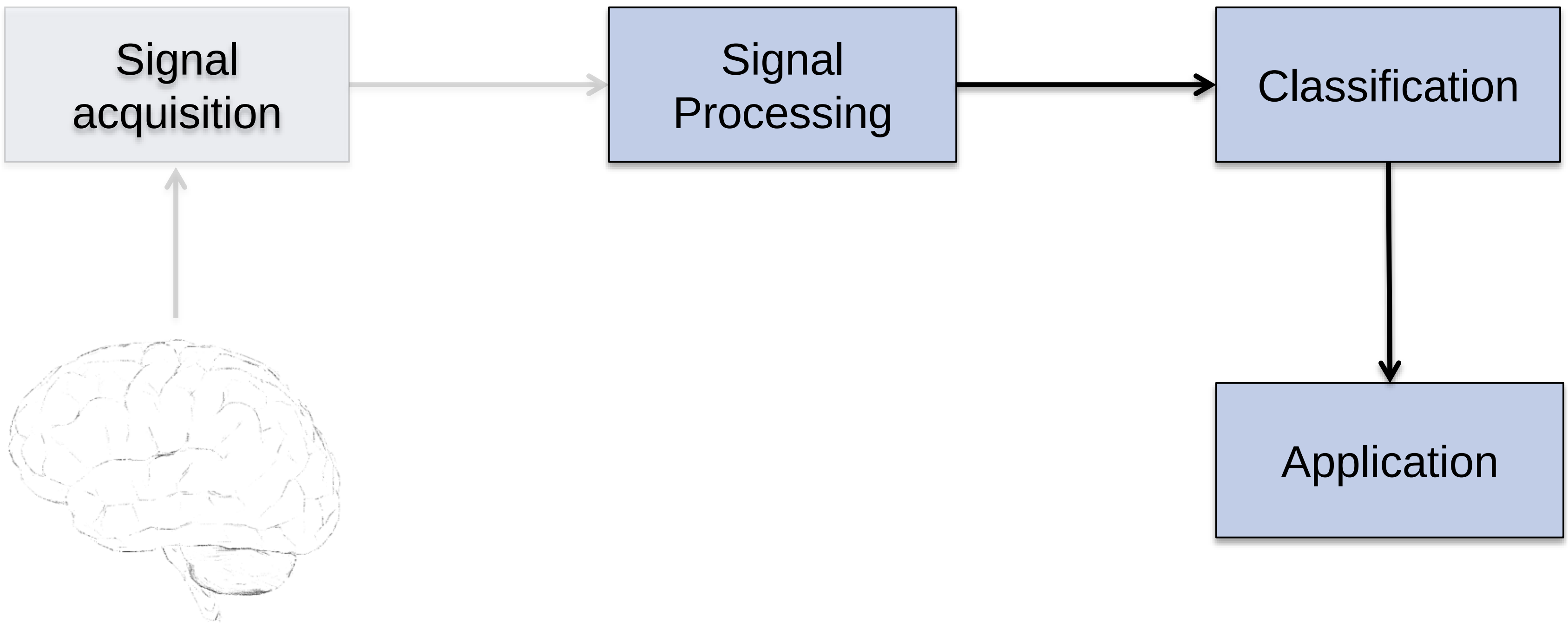
Help medical doctors to improve Alzheimer’s Disease diagnosis

Early detection of AD
Signal acquisition

EEG data sets:

- MCI Data set (*Musha*): 22 MCI subjects and 38 healthy control subjects
- Mild AD Data set (*Plymouth*): 17 Mild AD subjects and 24 healthy control subjects
- Both data sets acquired with 10-20 system but with different equipments and procedures

Early detection of AD
Signal processing



Early detection of AD

Signal processing

Proposed measures to characterize the EEG:

a) Power Measure: Relative Power

$$RP_i\,(f_1, f_2) = \frac{P_i(f_1, f_2)}{P_i(f_{min}, f_{max})}$$

b) Synchrony Measures

- Bivariate Measures

  Correlation

$$r = \frac{1}{N}\sum_{k=1}^{N}\frac{(x(k)-\tilde{x})}{\sigma_x}\frac{(y(k)-\tilde{y})}{\sigma_y}$$

  Coherence

$$c^2(f) = \frac{|\langle X(f)Y^*(f)\rangle|^2}{|\langle X(f)\rangle||\langle Y(f)\rangle|}$$

  Phase Synchrony

$$\gamma = \left|\langle e^{i(n\phi_x - m\phi_y)}\rangle\right|$$

- Multivariate Measures

  Granger Causality

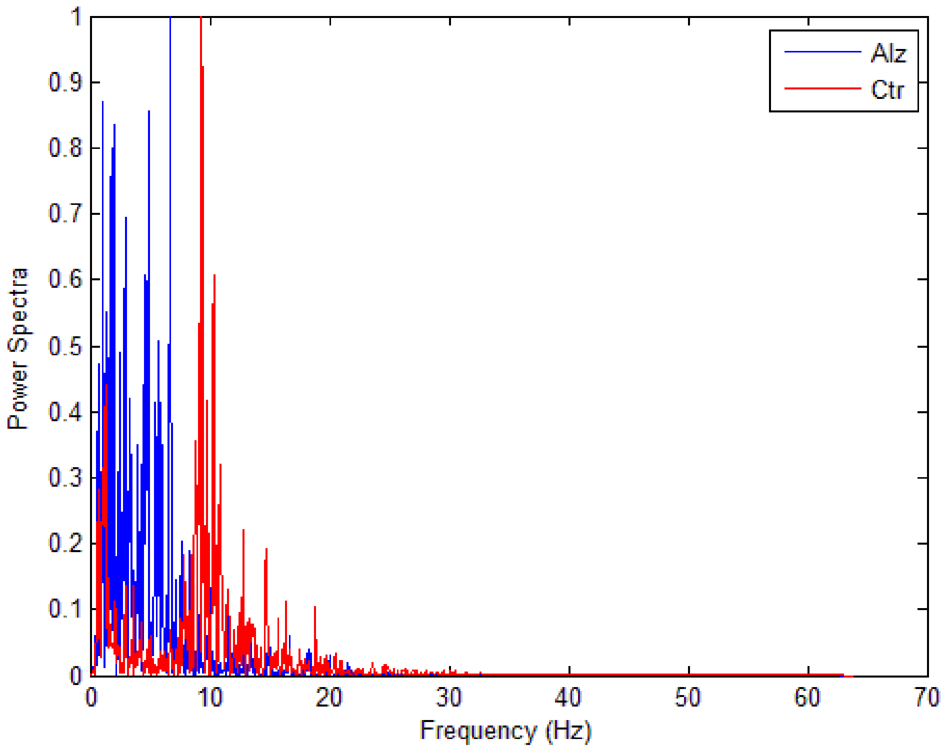
  Omega Complexity

$$\Omega = \exp\left(-\sum_{i=1}^{n}\lambda_i \log \lambda_i\right)$$

UVIC UNIVERSITAT DE VIC

Early detection of AD
Signal processing

AD introduce significant changes in the EEG power spectra:



UNIVERSITAT DE VIC

Early detection of AD

Signal processing

Proposed measures to characterize the EEG:

a) Power Measure: Relative Power

$$RP_i\,(f_1, f_2) = \frac{P_i(f_1, f_2)}{P_i(f_{min}, f_{max})}$$

b) Synchrony Measures

- Bivariate Measures

  Correlation

  $$r = \frac{1}{N}\sum_{k=1}^{N}\frac{(x(k)-\tilde{x})}{\sigma_x}\frac{(y(k)-\tilde{y})}{\sigma_y}$$

  Coherence

  $$c^2(f) = \frac{|\langle X(f)Y^*(f)\rangle|^2}{|\langle X(f)\rangle||\langle Y(f)\rangle|}$$

  Phase Synchrony

  $$\gamma = \left|\langle e^{i(n\phi_x - m\phi_y)}\rangle\right|$$

- Multivariate Measures

  Granger Causality

  Omega Complexity

  $$\Omega = \exp\left(-\sum_{i=1}^{n}\lambda_i \log \lambda_i\right)$$

Early detection of AD

Signal processing

Granger causality measures are derived from the multivariate autoregressive (MVAR) model of the multivariate time series:

$$x(n) = \sum_{l=1}^{p} \boldsymbol{A}(j)x(k-l) + e(k)$$

$$e(k) = \sum_{l=0}^{p} \widetilde{\boldsymbol{A}}(j)x(k-l)$$

$$\boldsymbol{E}(f) = \widetilde{\boldsymbol{A}}(f)\boldsymbol{X}(f)$$

$$\boldsymbol{X}(f) = \widetilde{\boldsymbol{A}}^{-1}(f)\boldsymbol{E}(f) = \boldsymbol{H}(f)\boldsymbol{E}(f)$$

$$\boldsymbol{S}(f) = \boldsymbol{H}(f)\boldsymbol{V}\boldsymbol{H}^{\dagger}(f)$$

x(*k*) are the time series, ***A***(*j*) are matrices and *e*(*k*) is a zero-mean Gaussian random vector

Early detection of AD
Signal processing

Family of synchrony measures derived from linear stochastic models:

1. Granger Coherence: $K_{ij}(f) = \frac{S_{ij}(f)}{\sqrt{S_{ii}(f)}\sqrt{S_{ij}(f)}}$
2. Partial Coherence: $C_{ij}(f) = \frac{M_{ij}(f)}{\sqrt{M_{ii}(f)}\sqrt{M_{ij}(f)}}$
3. Directed Transfer Function: $\gamma_{ij}^2(f) = \frac{\left|H_{ij}(f)\right|^2}{\sum_{j=1}^{m}\left|H_{ij}(f)\right|^2}$
4. Full Frequency Directed Transfer Function: $F_{ij}^2(f) = \frac{\left|H_{ij}(f)\right|^2}{\sum_f \sum_{j=1}^{m}\left|H_{ij}(f)\right|^2}$
5. Partial Directed Coherence: $P_{ij}(f) = \frac{\tilde{A}_{ij}(f)}{\sqrt{\sum_{i=1}^{m}\left|\tilde{A}_{ij}(f)\right|^2}}$
6. Direct Directed Transfer Function: $\chi_{ij}^2(f) = F_{ij}^2(f)C_{ij}^2(f)$

Early detection of AD

Signal processing

Proposed measures to characterize the EEG:

a) Power Measure: Relative Power

$$RP_i\,(f_1, f_2) = \frac{P_i(f_1, f_2)}{P_i(f_{min}, f_{max})}$$

b) Synchrony Measures

- Bivariate Measures

Correlation

$$r = \frac{1}{N}\sum_{k=1}^{N}\frac{(x(k)-\tilde{x})}{\sigma_x}\frac{(y(k)-\tilde{y})}{\sigma_y}$$

Coherence

$$c^2(f) = \frac{|\langle X(f)Y^*(f)\rangle|^2}{|\langle X(f)\rangle||\langle Y(f)\rangle|}$$

Phase Synchrony

$$\gamma = \left|\left\langle e^{i(n\phi_x - m\phi_y)}\right\rangle\right|$$

- Multivariate Measures

Granger Causality

Omega Complexity

$$\Omega = \exp\left(-\sum_{i=1}^{n}\lambda_i \log \lambda_i\right)$$

Early detection of AD

Signal processing

All the 11 measures are calculated in the following 435 different frequency ranges (from *F* to *F*+*W*):

$$
W \in \mathbb{N}[1,29] \quad \begin{bmatrix}
1-2 & & & & & \\
1-3 & 2-3 & & & & \\
1-4 & 2-4 & 3-4 & & & \\
\vdots & \vdots & \vdots & \ddots & & \\
1-29 & 2-29 & 3-29 & & 28-29 & \\
1-30 & 2-30 & 3-30 & \cdots & 28-30 & 29-30
\end{bmatrix}
$$

$$F \in \mathbb{N}[1,29]$$

Early detection of AD

Signal processing

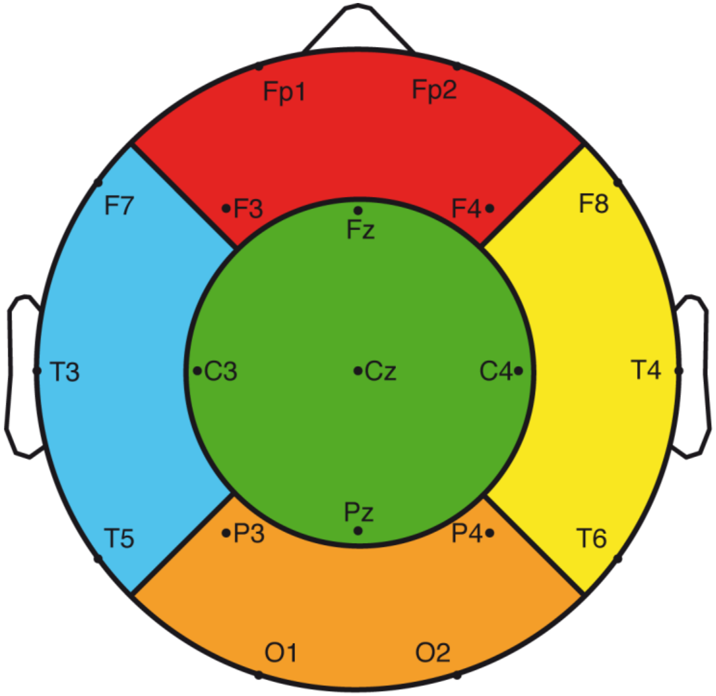
Global measures computation:

- RP was computed for each channel independently. To obtain a global measure for each subject, the RP for all the channels was averaged.
- For bivariate synchrony measures, EEG signals are aggregated into five regions (frontal, left temporal, central, right temporal and occipital).

Early detection of AD

Signal processing

Global Synchrony computation for bivariate measures (Correlation, Coherence and Phase Synchrony):



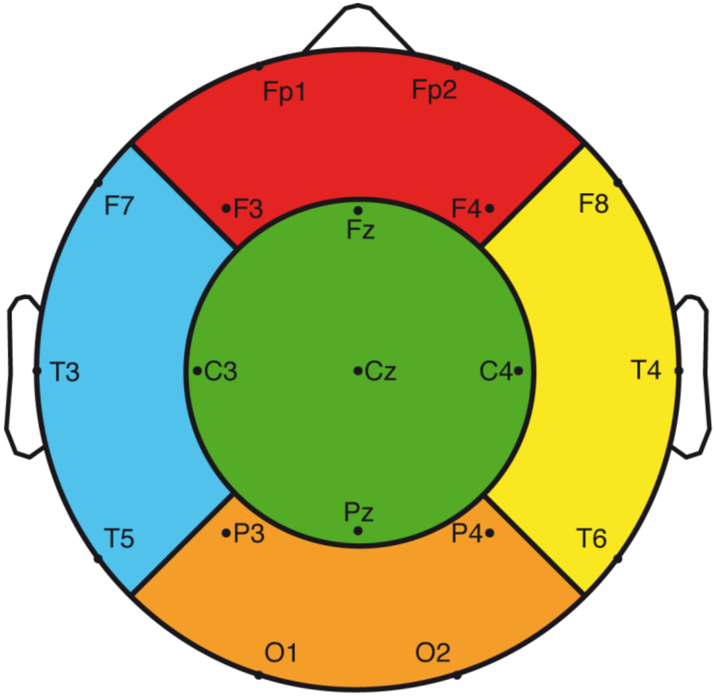
First, we compute the synchrony between each EEG signal from one region and each signal from the other.

Next, we evaluate synchrony by computing the average synchrony values of these signal pairs

UVIC UNIVERSITAT DE VIC

Early detection of AD
Signal processing

Global Synchrony computation for multivariate measures (Granger measures):



To avoid high-dimensional MVAR estimation, we calculated the time averaging between electrodes of the same region.

The Granger measures were then applied to these 5 averaged EEG signals.

The Granger values between the regions were then averaged.

Early detection of AD
Signal processing

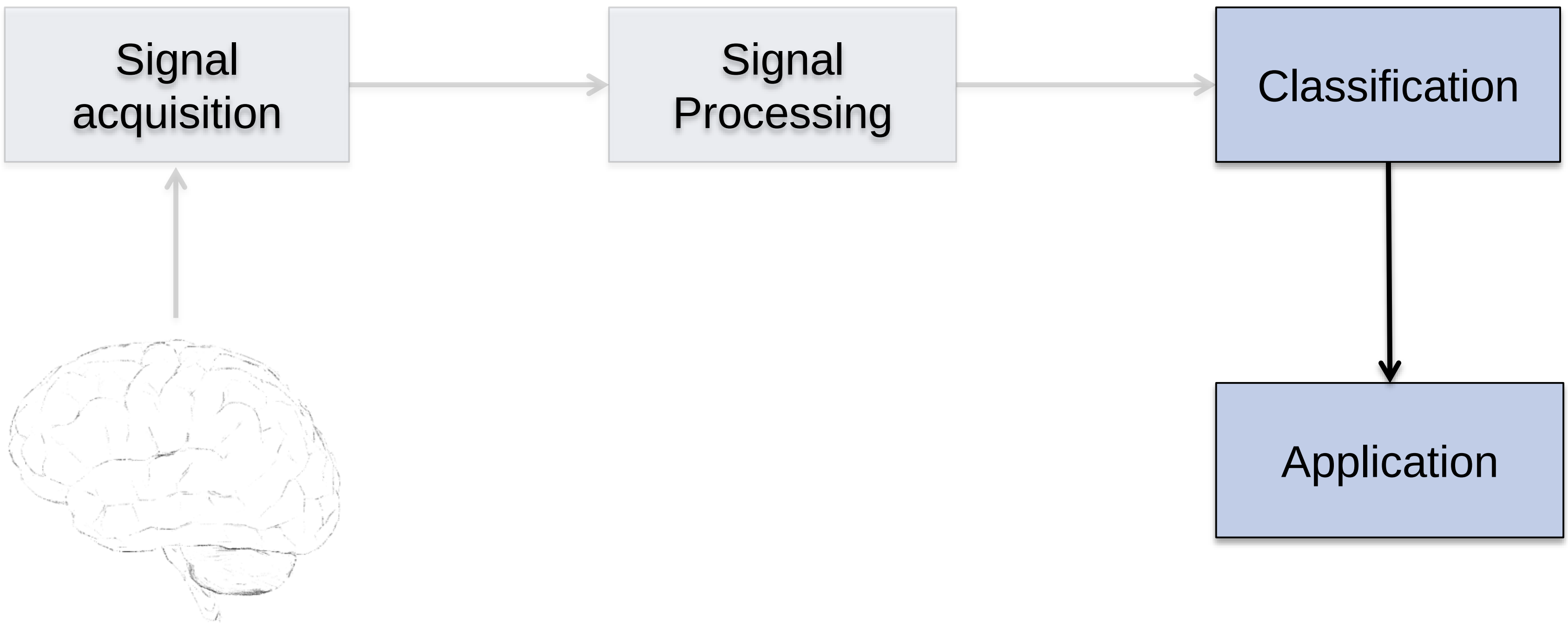
Separability criterion:

$$J(F, F+W) = \frac{|\mu_{Ctr}(F, F+W) - \mu_{Pat}(F, F+W)|}{(\sigma_{Ctr}(F, F+W) + \sigma_{Pat}(F, F+W))}$$

F and *(F+W)* refer to the start and end frequency of the study, respectively. σ refers to the mean and μ refers to the standard deviation.

We computed this separability criterion for each proposed measure

Early detection of AD
Classification



Early detection of AD

Classification

Two different studies were performed:

1. Classification using one single feature

   Exploring the best frequency band, the best signal length and the best Granger order.

2. Classification using multiple features

   Exploring the best combination of (few) features obtained trough Orthogonal Forward Regression (OFR).

In both cases, we used Linear Discriminant Analysis (LDA) with Leave One Out (LOO) cross-validation.

Early detection of AD

Classification (single feature)

MCI data set

Measures	Optimal frequency range (Hz)	p-value	CR optimal frequency range (%)	CR θ band (%)	CR α band (%)	CR β band (%)
RP	**2 - 9**	**0.0001**	**78.33**	70.00	53.33	71.67
Correlation	**1 - 8**	**0.0012**	**71.67**	70.00	61.67	71.67
Coherence	8 - 13	0.0132	68.33	55.00	68.33	8.33
Granger Coherence	2 - 8	0.0021	70.00	61.67	66.67	46.67
PC	21 - 27	0.0157	70.00	60.00	51.67	65.00
DTF	6 - 27	0.2727	65.00	35.00	58.33	56.67
ffDTF	8 - 30	0.0013	70.00	3.33	56.67	63.33
PDC	1 - 2	0.0085	66.67	61.67	56.67	41.67
dDTF	**14 - 16**	$\mathbf{8.88 \times 10^{-5}}$	**75.00**	71.67	46.67	70.00
Omega Complexity	8 - 10	0.0138	68.33	43.33	56.67	53.33
Phase Synchrony	4 - 5	0.0201	70.00	60.00	60.00	50.00

Early detection of AD

Classification (single feature)

Mild AD data set

Measures	Optimal frequency range (Hz)	p-value	CR optimal frequency range (%)	CR θ band (%)	CR α band (%)	CR β band (%)
RP	**4 - 7**	**8.38×10^{-8}**	**97.56**	87.80	90.24	78.05
Correlation	23 - 24	0.0699	68.29	21.95	53.66	78.05
Coherence	8 - 13	0.0003	75.61	70.73	75.61	51.22
Granger Coherence	**1 - 2**	**3.07×10^{-5}**	**82.93**	60.98	46.34	51.22
PC	3 - 4	0.0027	68.29	63.41	56.10	48.78
DTF	**5 - 6**	**2.64×10^{-6}**	**95.12**	87.80	21.95	14.63
ffDTF	1 - 2	3.00×10^{-6}	80.49	70.73	68.29	56.10
PDC	1 - 4	1.78×10^{-6}	80.49	78.05	60.98	60.98
dDTF	2 - 4	8.50×10^{-5}	78.05	68.29	53.66	63.41
Omega Complexity	7 - 8	0.0005	75.61	60.98	63.41	48.78
Phase Synchrony	9 - 10	5.45×10^{-5}	80.49	58.54	70.73	46.34

Early detection of AD
Classification (single feature)

Note that:

- The results shown in the optimal frequency range of each measure are always equal to or higher than the values obtained using the standard frequency ranges.
- The best CR was obtained in both data sets with RP.
- If a measure has several frequency ranges with the same CR, the one selected as the optimal frequency range is the one with the highest *J* for that measure.

Early detection of AD

Classification (multiple features)

- We want to determine which measures would be the most relevant for distinguishing MCI/Mild AD patients from healthy subjects.
- To control overfitting and to rank the input features by their significance, we use a procedure based on Gram-Schmidt Orthogonal Forward Regression (OFR).
- In order to control overfitting we applied the random probe method: random generations of data (extra feature) used to verify that the analyzed feature is more significant than random data.

Early detection of AD

Classification (multiple features)

The OFR algorithm used in this work can be summarized as follows:

1. Select the input features (if_i). For each if_i , select the frequency range that corresponds to the largest J. Repeat this procedure for all i measures.
2. Select the candidate feature (x_i) that best correlates to the output (o) to be modeled: $x_i = \mathrm{argmax}_\mathrm{i} \ \cos^2(if_i, o)$
3. Project the output vector on to the null space of the selected feature. Orthogonalize the rest of features using Gram-Schmidt orthogonalization in all the existing frequency ranges.
4. Remove the selected feature (x_i) from the list of input measures.
5. Return to (1) until all features have been selected.

Early detection of AD

Classification (multiple features)

MCI data set

Algorithm Order	Features	OFR Selected frequency ranges (Hz)	Standard frequency bands (Hz)
1	RP	2 - 8	4 - 8
2	Correlation	3 - 8	4 - 8
3	Coherence	1 - 6	1 - 4
4	PDC	1 - 3	1 - 4
5	ffDTF	9 - 29	13 - 30
6	Omega Complexity	24 - 25	13 - 30
7	Granger Coherence	1 - 30	1 - 4, 4 - 8, 8 – 13, 13 - 30
8	DTF	4 - 5	4 - 8
9	PC	1 - 10	1 - 4, 4-8
10	Phase Synchrony	28 - 30	13 - 30
11	dDTF	1 - 2	1 - 4

Early detection of AD

Classification (multiple features)

MCI data set

Algorithm Order	Features	OFR Selected frequency ranges (Hz)	Standard frequency bands (Hz)
1	RP	2 - 8	4 - 8
2	Correlation	3 - 8	4 - 8
3	Coherence	1 - 6	1 - 4
4	PDC	1 - 3	1 - 4
5	ffDTF	9 - 29	13 - 30
6	Omega Complexity	24 - 25	
7	Granger Coherence	1 - 30	13 - 30
8	DTF	4 - 5	4 - 8
9	PC	1 - 10	1 - 4, 4-8
10	Phase Synchrony	28 - 30	13 - 30
11	dDTF	1 - 2	1 - 4

CR: 81,67 %

Early detection of AD

Classification (multiple features)

Mild AD data set

Algorithm Order	Features	OFR Selected frequency ranges (Hz)	Standard frequency bands (Hz)
1	RP	4 – 7	4 - 8
2	Granger Coherence	1 – 2	1 - 4
3	Correlation	9 – 10	8 - 13
4	Phase Synchrony	25 – 26	13 - 30
5	PC	13 – 14	13 - 30
6	dDTF	2 – 6	4 - 8
7	Coherence	5 – 6	4 - 8
8	Omega Complexity	11 – 14	8 - 13
9	ffDTF	6 – 19	8 - 13
10	DTF	20 – 21	13 - 30
11	PDC	1 – 2	1 - 4

Early detection of AD

Classification (multiple features)

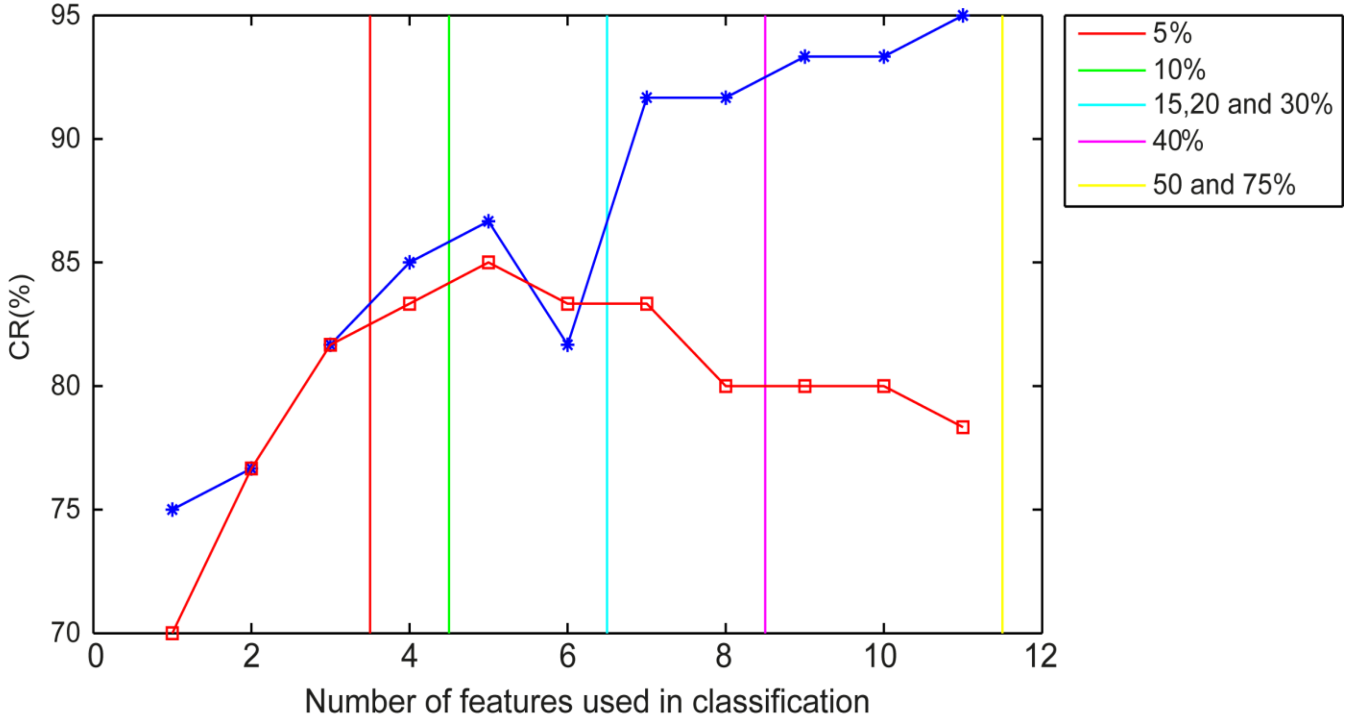
Mild AD data set

Algorithm Order	Features	OFR Selected frequency ranges (Hz)	Standard frequency bands (Hz)
1	RP	4 – 7	4 - 8
2	Granger Coherence	1 – 2	1 - 4
3	Correlation	9 – 10	8 - 13
4	Phase Synchrony	25 – 26	13 - 30
5	PC	13 – 14	13 - 30
6	dDTF	2 – 6	
7	Coherence	5 – 6	4 - 8
8	Omega Complexity	11 – 14	8 - 13
9	ffDTF	6 – 19	8 - 13
10	DTF	20 – 21	13 - 30
11	PDC	1 – 2	1 - 4

CR: 100 %

Early detection of AD
Classification (multiple features)

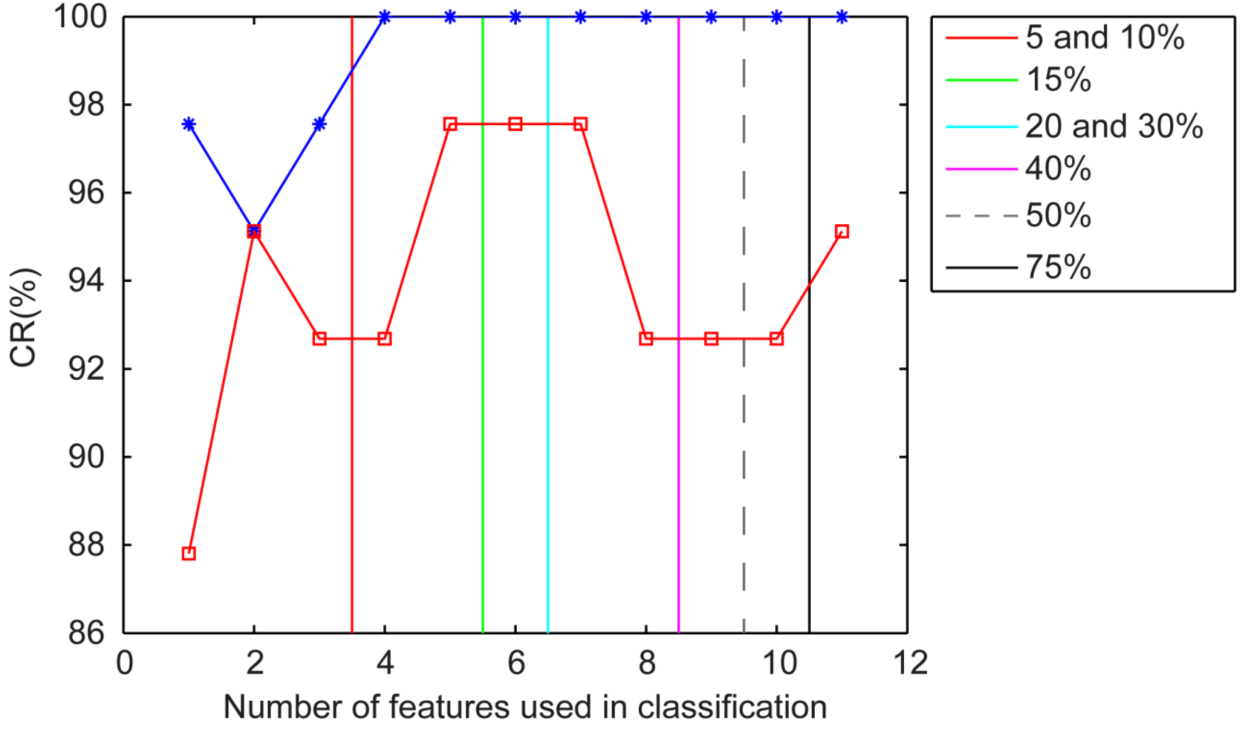
MCI data set



Early detection of AD

Classification (multiple features)

Mild AD data set



Early detection of AD

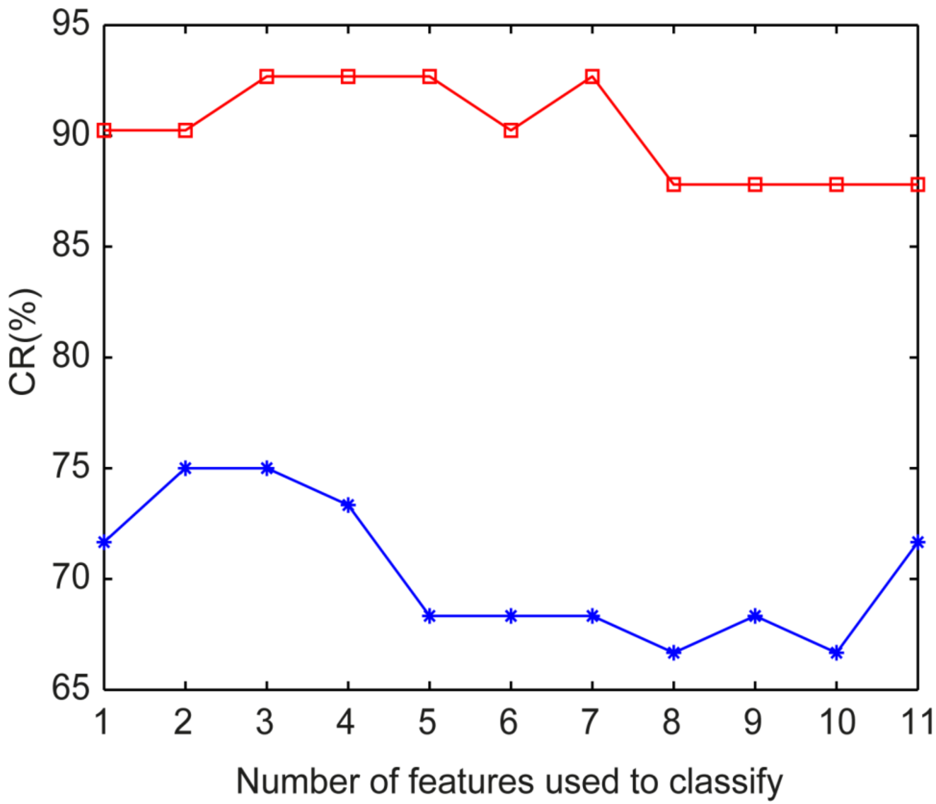
Classification (multiple features)

• In order to standardize the obtained results, we carried out one more experiment:

> both data sets were evaluated using the obtained parameters from the other data set.

•The change of parameters clearly reduces the CR obtained for the MCI data set but only presents a slight decrease for the Mild AD data set in comparison with the results obtained for each data set using its own OFR-selected measures and frequency ranges.

Early detection of AD
Classification (multiple features)



Evolution of the CR obtained, using the different numbers of features.

The line with asterisks shows the MCI data set using OFR-selected features for Mild AD patients.

The line with squares indicates the Mild AD data set using OFR-selected features for MCI patients.

Early detection of AD

Conclusions

- In this study, a group of synchrony measures and a frequency power measure were used to perform an early diagnosis of AD.

- Single features were used to compute CR in order to obtain the optimal frequency range that best discriminates between AD patients and healthy subjects.

- A multiple feature classification approach based on OFR was also presented, with the aim of obtaining a final CR that improves upon state of the art results, was described.

Early detection of AD

Conclusions

- The two data sets analyzed in this study (MCI and Mild AD) were obtained through different EEG recordings in two different hospitals, with different EEG systems and slightly different protocols.
- We can expect significant variations in the experimental conditions.
- We chose to perform an independent study of each database separately.
- Interestingly, with both data sets high classification rates were obtained: 95% for the MCI data set (using 11 features), and 100% for the Mild AD set (using 4 features).

Early detection of AD
Conclusions

- The results in frequency bands differing from the standard ranges were shown to be more discriminant.
- It seems that using a specific configuration and computing neural synchrony in a specific frequency range is more effective than standardizing all configurations.
- We also explored the possibility of using the same features for both data sets:
    - using features optimal for the MCI data set, we obtained promising results for the Mild AD data set maintaining the result for MCI.
- The standardization of features for the MCI and Mild AD data sets is worth future investigation.

UVIC
UNIVERSITAT
DE VIC

Thank you!

Overview

- Signal acquisition
- Signal processing
- Classification
- Application
- Conclusions