FEATURE SELECTION ALGORITHMS FOR CORTICOMUSCULAR COHERENCE-BASED BRAIN-COMPUTER INTERFACES

DEPARTMENT OF COMPUTER, CONTROL, AND MANAGEMENT ENGINEERING ANTONIO RUBERTI

SAPIENZA UNIVERSITÀ DI ROMA

CANDIDATE: **GIOELE MIGNO 1795826**

ADVISOR: **PROF. FEBO CINCOTTI**
CO-ADVISOR: **PROF. EMMA COLAMARINO**

FEATURE SELECTION ALGORITHMS FOR CORTICOMUSCULAR COHERENCE-BASED BRAIN-COMPUTER INTERFACES

FIELD OF APPLICATION:
NEUROENGINEERING

ARTIFICIAL INTELLIGENCE
TASK

FEATURE SELECTION ALGORITHMS FOR CORTICOMUSCULAR COHERENCE-BASED BRAIN-COMPUTER INTERFACES

FIELD OF APPLICATION:
NEUROENGINEERING

ARTIFICIAL INTELLIGENCE
TASK

FEATURE SELECTION ALGORITHMS FOR CORTICOMUSCULAR COHERENCE-BASED BRAIN-COMPUTER INTERFACES

FIELD OF APPLICATION:
NEUROENGINEERING

TYPE OF FEATURE

Introduction

Brain-Computer Interfaces for post-stroke rehabilitation

Stroke

Stroke is a condition in which **cells** in the **brain** suddenly **die** because of lack of oxygen caused by an obstruction in the blood flow or by a hemorrhage.

> In 2021, stroke was responsible for **over 7 millions deaths**, approximately **10%** of the total **global** deaths. It is one of the main reasons for abnormal human death.

Stroke: **Post-stroke patients**

In recent years there has been a noticeable **increase** in the **survival rate** of **post-stroke patients**.

Who survive is usually left with **disabilities**, affecting especially **upper limb movement**.

Patients lose their ability to **control** their **muscles** in the **physiological manner**.

Post-stroke rehabilitation

Rehabilitation makes the difference in **overcoming** or not motor **disabilities**.

Among the recent advanced **rehabilitation techniques,** there are the ones based on **Brain-Computer Interfaces (BCIs).**

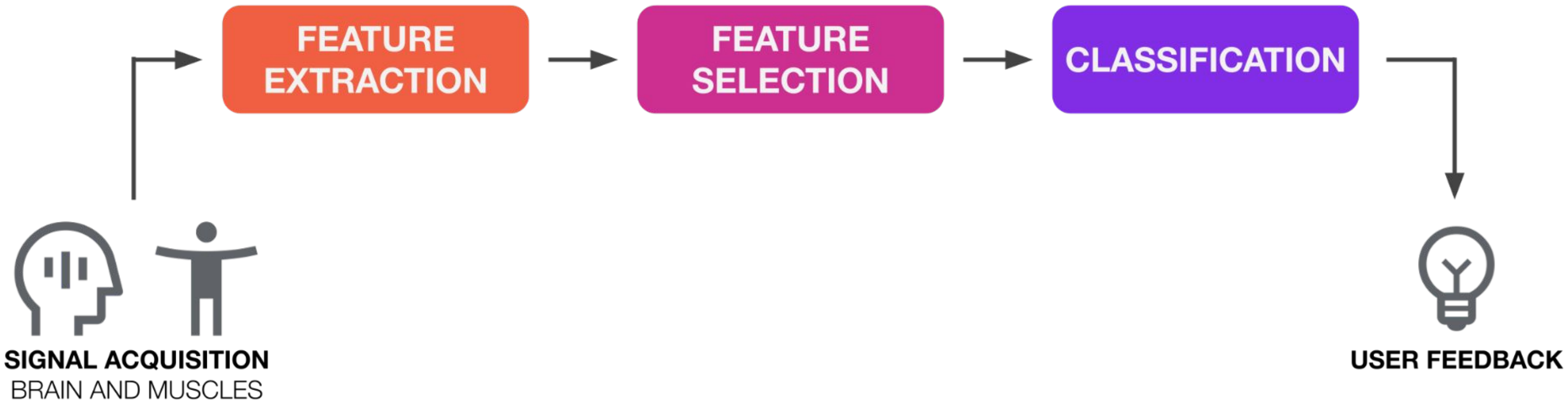
Pichiorri et al. 2023

Post-stroke rehabilitation: **Brain-Computer Interface (BCI)**

BCI provides a way to develop **interaction** between a **brain** and a **computer**. This interaction is achieved by using control signals arising due to the **brain activity**.

In this work, the type of BCI used is called **Hybrid BCI** which also acquires **muscle activity** as peripheral signal.

Post-stroke rehabilitation: **Pipeline**



Pipeline: **Feature selection**

Feature selection is a crucial step usually performed by a **high specialized neurophysiologist**.

It is important to select features that take into account **neurophysiological** and **rehabilitative** principles.

FEATURE EXTRACTION → FEATURE SELECTION → CLASSIFICATION

Pipeline: **Automated feature selection**

The focus of my research is on **automated** feature selection step which does **not** require **high specialized** neurophysiological competence.

The aims are:

- expand BCI use to users with **limited**:
  - **expertise** on BCI;
  - **knowledge** of the **neurophysiological principles** that guide the selection of the features.
- improve the choice by also considering the **classification performance.**

Dataset and problem

Feature selection in a binary classification task

Datasets: Data acquisition

This study is focused on **healthy subjects** who execute movements in a repeatable way intra/inter individuals.

The **processed** data used in this work was **provided by** IRCCS Santa Lucia Foundation, Rome, Italy.

SANTA LUCIA

Data acquisition: **Movements**

The data was collected from 14 healthy subjects performing four different movements:

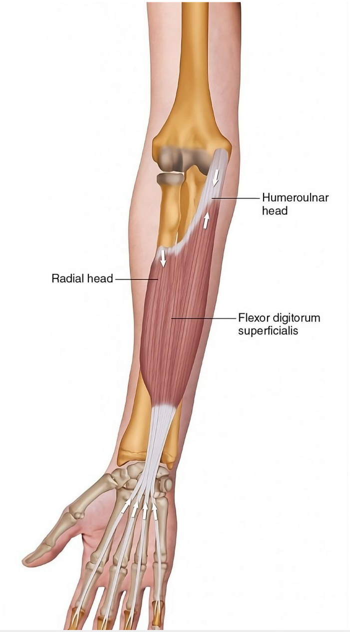
- **GraspL:** Finger grasping with the left hand;
- **GraspR:** Finger grasping with the right hand;
- **ExtL:** Finger extension with the left hand;
- **ExtR:** Finger extension with the right hand;

Each movement was studied **independently** from the others

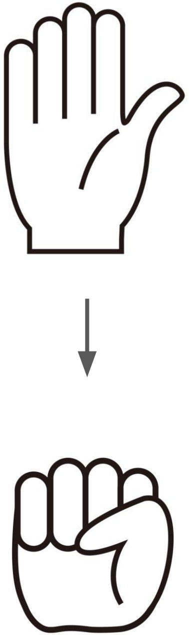
Data acquisition: **GraspL**

For the sake of brevity, during this presentation, the explanation is **focus** on only **finger grasping** `GraspL`.

The same procedure explained in the following was applied also to the other movements.



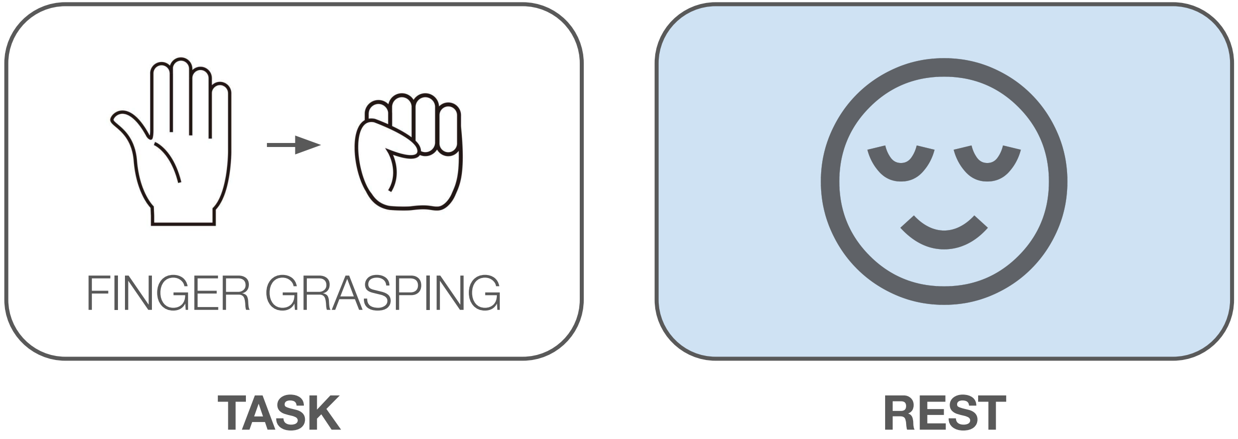
FLEXOR DIGITORUM SUPERFICIALIS



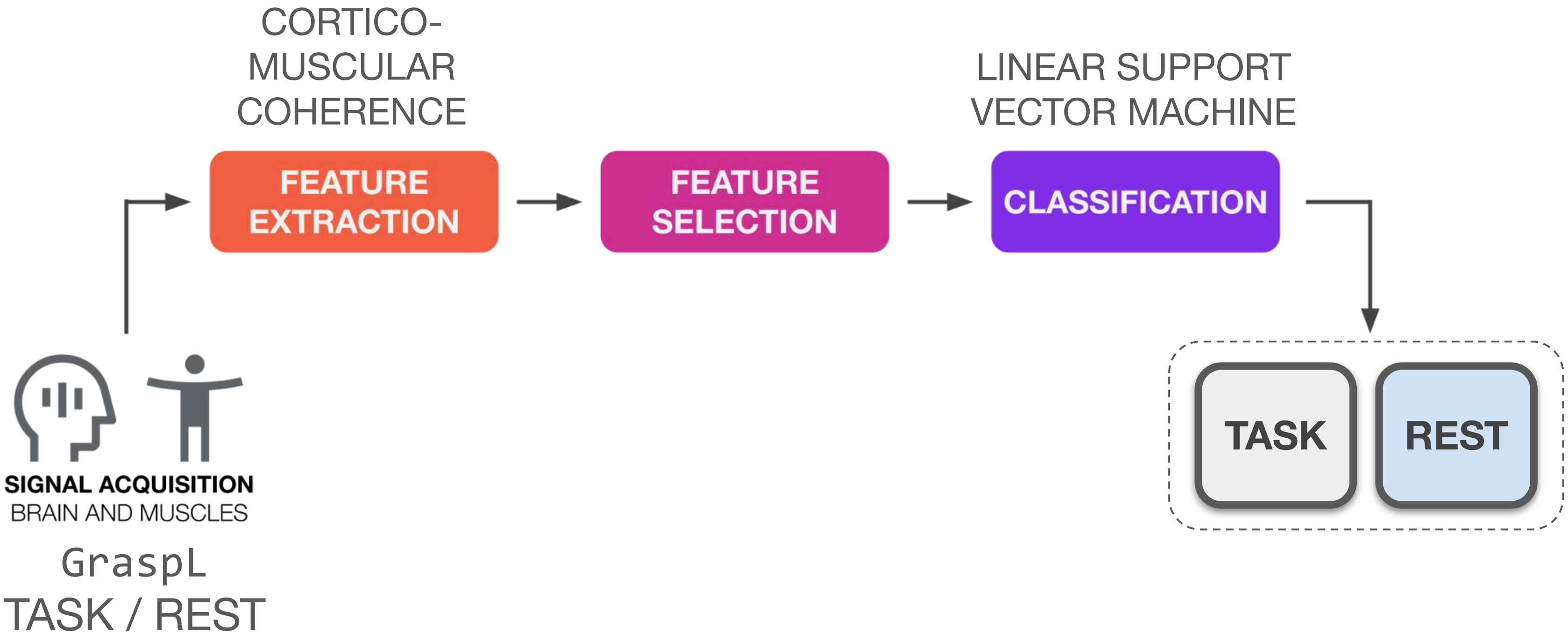
GRASPING
GraspL

GraspL: **Rest and Task**

During **data acquisition**, the subject **alternates** moments where he performs the **movement** and moments where he is at **rest**:

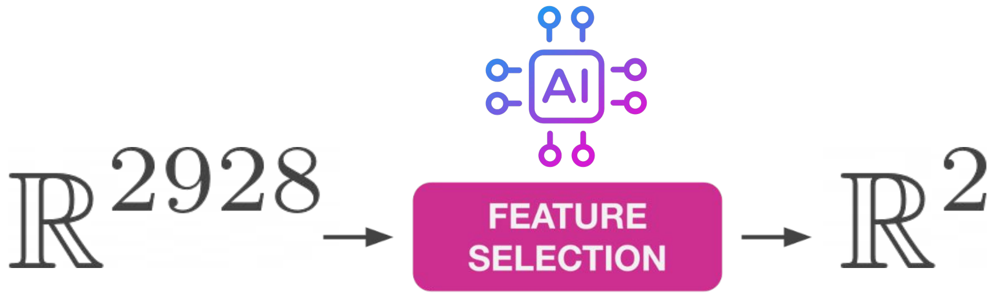


GraspL: **Pipeline**



Pipeline: **Automated feature selection**

From the initial 2928 features, the **two** most **relevant features** must be selected to optimize the classification performance.



Different feature selection algorithms were **compared**.

Proposed solution

State-Of-The-Art

and Random Planet

State-Of-The-Art: Feature selection algorithms

In this field, some of the State-Of-The-Art feature selection methods are:

- **Mutual Information:** Statistical measure-based method;
- **Stepwise Regression:** Statistical search oriented approach;
- **Genetic Algorithm:** Metaheuristic search technique;
- **Decision Tree:** Common method in Machine Learning;
- **Random Forest:** Ensemble of Decision Tree models.

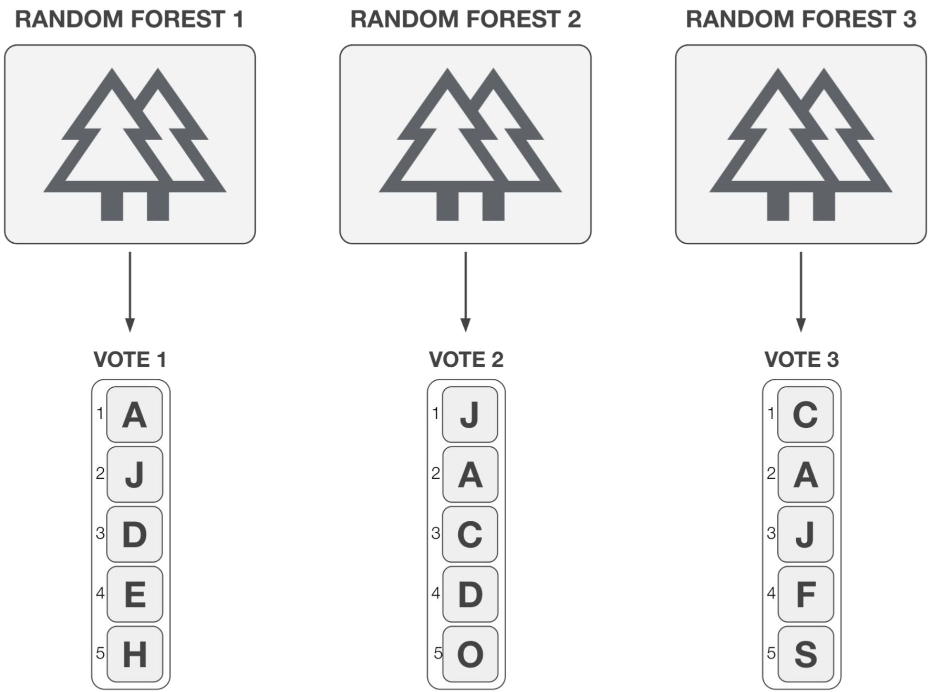
> Both **Decision Tree** and **Random Forest** have **low stability**, namely the features selected are highly **influenced** by the initial conditions (i.e., **random seed** adopted).

Random Planet

Random Planet is a **novel method** introduced in this work to **improve stability** of Decision Tree and Random Forest.

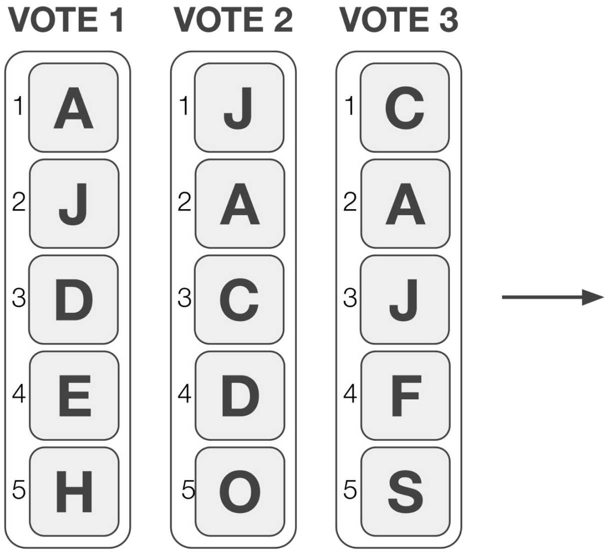
Random Planet is an **ensemble** of Random Forest models:

Random Planet: **Example - Collection of votes**



PROPOSED SOLUTION

Random Planet: **Example - Most voted features**



A = 3
J = 3
C = 2
D = 2
E = 1
F = 1
H = 1
S = 1
O = 1

Random Planet: **Example - Tie-breaker**

A = 3
J = 3
C = 2
D = 2
E = 1
F = 1
H = 1
S = 1
O = 1

Classification performance obtained by a linear SVM using **only** feature A

A = 3 + 0.82
J = 3 + 0.89

Classification performance obtained by a linear SVM using **only** feature J

FINAL RANKING

1 **J**
2 **A**

PROPOSED SOLUTION

Feature selection algorithms compared

The feature selection algorithms compared are:

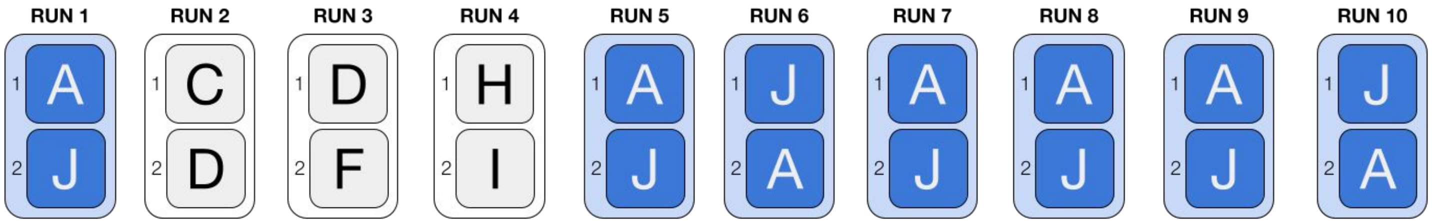
- **Decision Tree** `(DTree);`
- **Genetic Algorithm** `(GAAM):` **9** hyperparameter combinations tested;
- **Mutual Information** `(MInfo);`
- **Random Planet** `(RPlanet):` **45** hyperparameter combinations tested;
- **Stepwise Regression** `(StepwiseR).`

> **Hyperparameter tuning** procedures based on **statistical analysis (ANOVA)**, were performed for **Genetic Algorithm** and **Random Planet**.

Feature selection algorithms compared: **Metrics**

The metrics adopted to compared the feature selection algorithms are:

- Area Under Curve (**AUC**) of the Receiver Operating Characteristic curve;
- **F1 Score**;
- **Soft Stability (SoftS):** Metric defined in this study to measure the stability.



$$\text{SoftS} = 7/10 = 0.7$$

Result

Algorithms comparison

GraspL: Performance comparison

Mean **Soft Stability** (SoftS) of each algorithm and mean **classification** performance obtained using the features selected by each method.

Means computed over the 14 subjects. (Mean ± std)

Method	AUC Train	AUC Test	F1 Train	F1 Test	SoftS
DTree	0.959 ± 0.02	0.957 ± 0.02	0.957 ± 0.02	0.955 ± 0.03	0.343 ± 0.28
GAAM	0.984 ± 0.02	0.975 ± 0.02	0.984 ± 0.02	0.975 ± 0.02	0.443 ± 0.17
MInfo	0.957 ± 0.03	0.952 ± 0.04	0.954 ± 0.04	0.948 ± 0.05	1.0 ± 0.0
RPlanet	0.960 ± 0.02	0.954 ± 0.03	0.958 ± 0.02	0.951 ± 0.03	0.671 ± 0.27
StepwiseR	0.966 ± 0.02	0.964 ± 0.02	0.964 ± 0.03	0.962 ± 0.03	1.0 ± 0.0

Performance comparison: **Statistical significance**

AUC Test performances were compared using **statistical** methods (ANOVA).

Significant **differences** were found only between Mutual Information and Genetic Algorithm with this last having **higher** mean AUC performance.

Method	AUC Test
DTree	0.957±0.02
GAAM	0.975±0.02
MInfo	0.952±0.04
RPlanet	0.954±0.03
StepwiseR	0.964±0.02

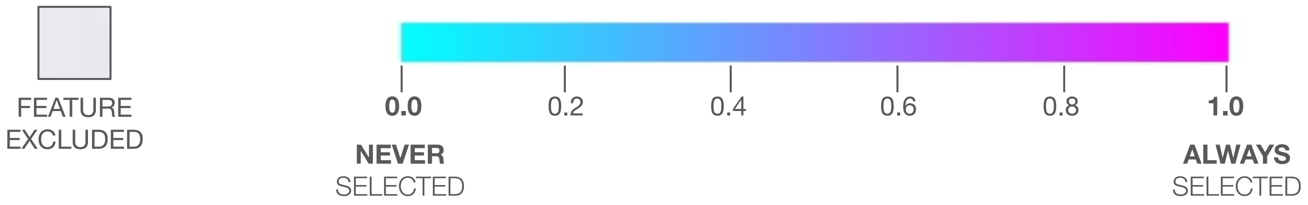
GAAM > MInfo

GraspL: **Selected Features**

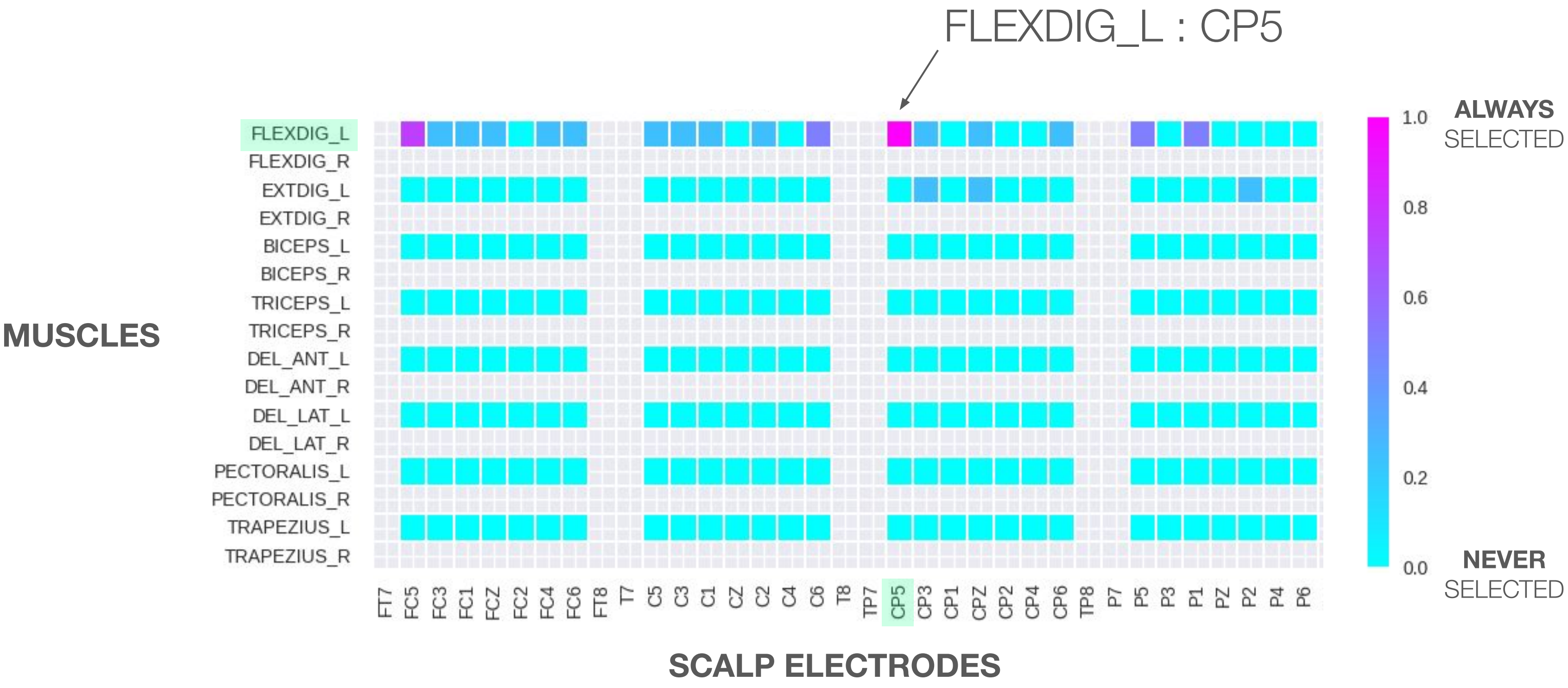
Restricting the initial feature space taking into account **neurophysiological constraints**, it is interesting to see the different features chosen by each method.

This information is represented as an **heatmap** with values ranging from 0 to 1:

- **Value ≃ 0:** the feature was **never** or rarely selected in one of the subjects.
- **Value ≃ 1:** the feature was selected as the best for **most** of the subjects.



Selected Features: **Random Planet**



GraspL, GraspR, ExtL, ExtR

Similar results were obtained for the **other movements** (`GraspR`, `ExtL`, `ExtR`) and using **different features domains** (spatial filtering, brain/muscles channels).

All the feature selection algorithms tested allow to obtain **high classification** performance (AUC-F1 Score Test > 0.9) without **statistical significant** differences most of the time.

GraspL, GraspR, ExtL, ExtR: **Overall perspective**

From an **overall perspective**:

- Genetic Algorithm has the **same** AUC performance of Stepwise Regression
    - Genetic Algorithm has a **lower stability**

- Decision Tree and Random Planet have the **same** AUC performance
    - As desired, Random Planet has always **higher stability**

Future works

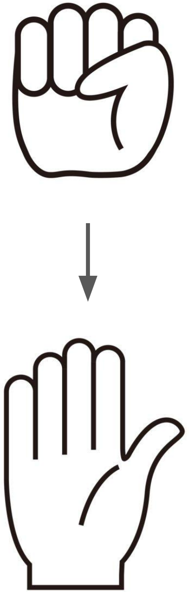
Neurophysiology expert consulting of the heatmaps is necessary to further considerations on what algorithms select the most **relevant** features from a **physiological point of view**.

The comparison should be **extended** to **post-stroke patients** considering also the features a **neurophysiologist** would select as **reference**.

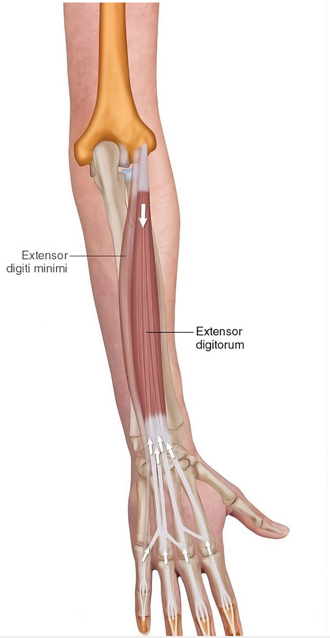
Thanks for your attention

End of the presentation

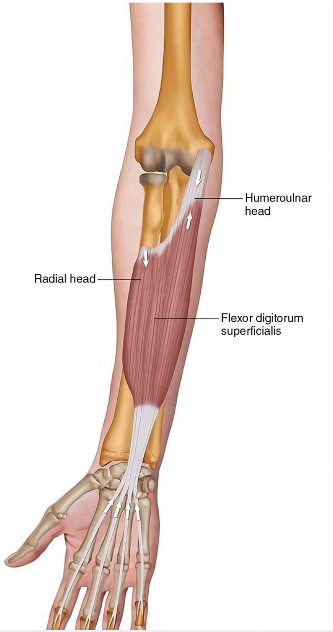
Finger Extension and Grasping



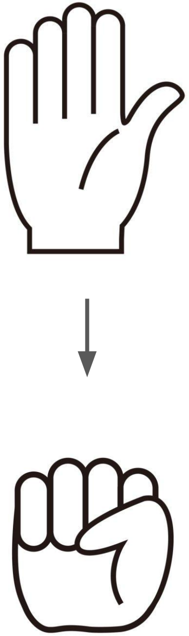
EXTENSION
Ext



EXTENSOR DIGITORUM



FLEXOR DIGITORUM
SUPERFICIALIS



GRASPING
Grasp

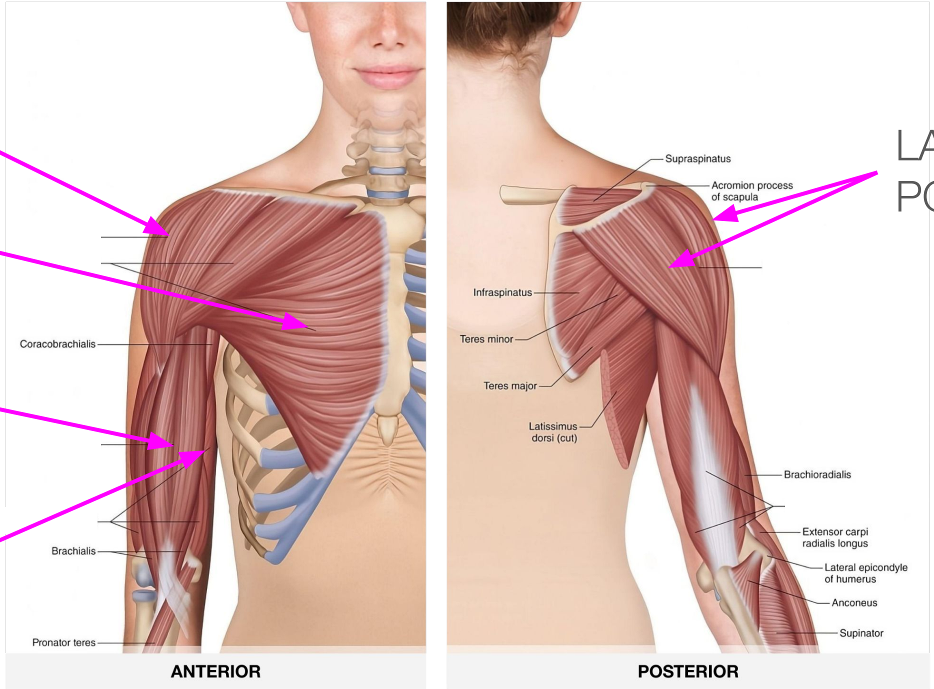
Upper limb muscles

ANTERIOR DELTOID

PECTORALIS MAJOR

BICEPS BRACHII

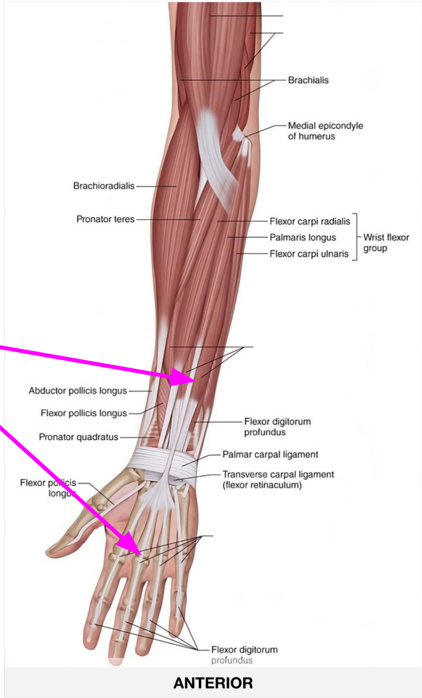
TRICEPS BRACHII

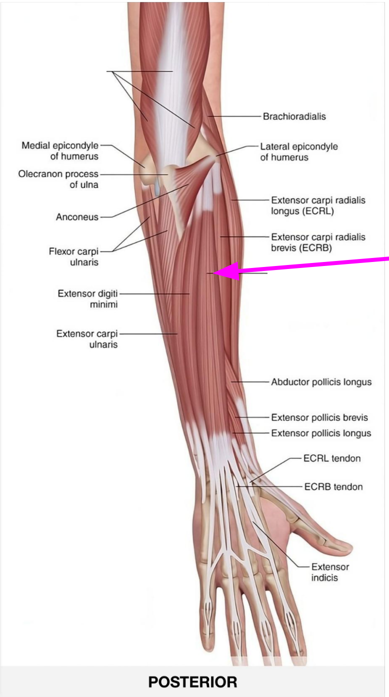


LATERAL AND POSTERIOR DELTOID

Upper limb muscles

FLEXOR
DIGITORUM
SUPERFICIALIS





EXTENSOR
DIGITORUM

APPENDIX

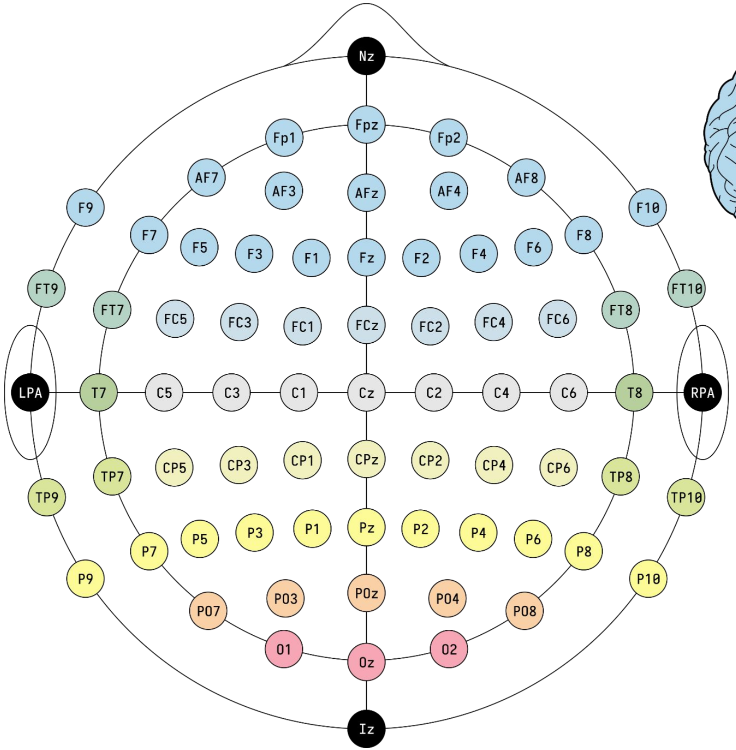
Brain signal: Scalp electroencephalography (EEG)

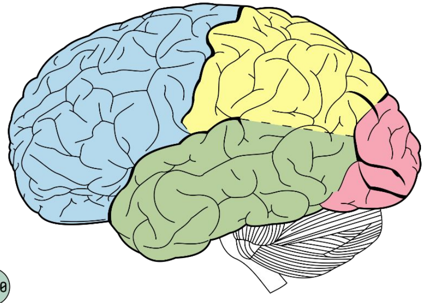
EEG is the **most common** technique to acquire brain signal:

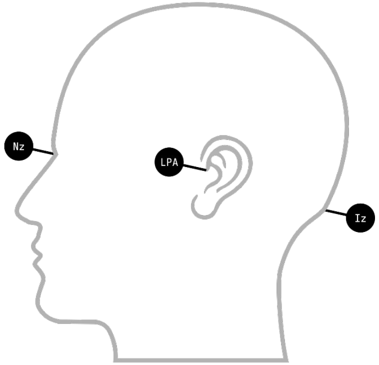
- non-invasive
- high temporal resolution
- relative cheap
- relative easy to use
- scalable

EEG: **10-10 International Electrodes system**

The placement of the electrodes on the scalp follows specific patterns called electrode systems

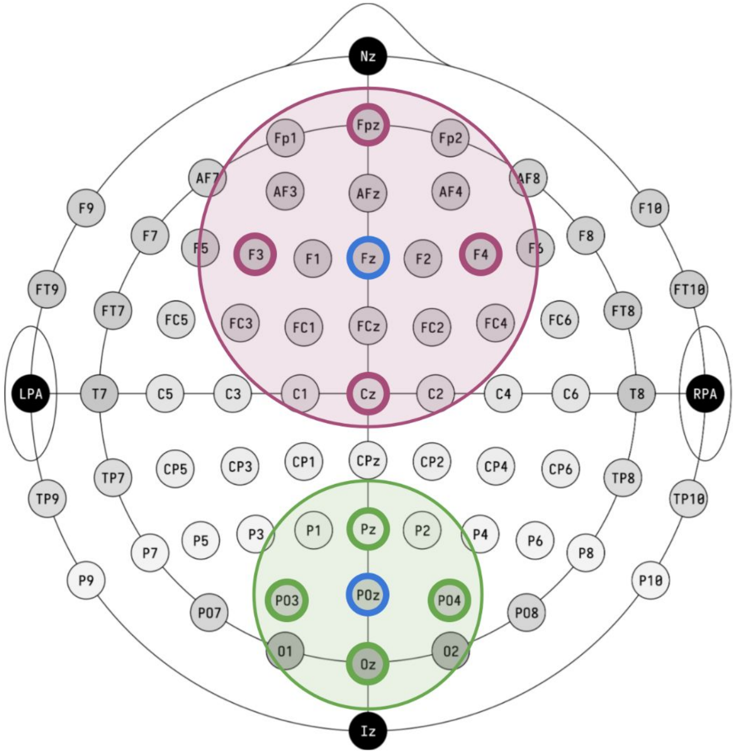






APPENDIX

EEG: **Laplacian filtering**



LARGE LAPLACIAN

$$\overline{F_Z} = F_Z - \frac{1}{4}(F_{PZ} + C_Z + F3 + F4)$$

SMALL LAPLACIAN

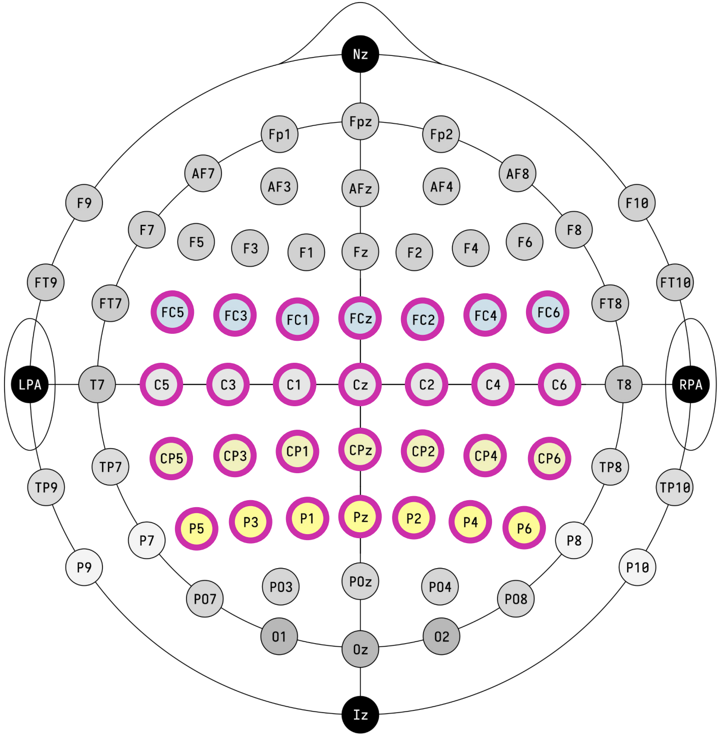
$$\overline{P_{OZ}} = P_{OZ} - \frac{1}{4}(P_Z + O_Z + PO3 + PO4)$$

EEG: **Sensorimotor area**

The brain region of **interest** during limb **rehabilitation** is **sensorimotor area**.

It is responsible for:

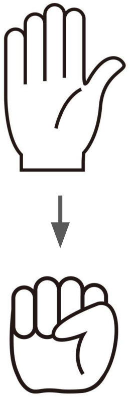
- movement planning
- movement execution
- proprioception
- touch and pain

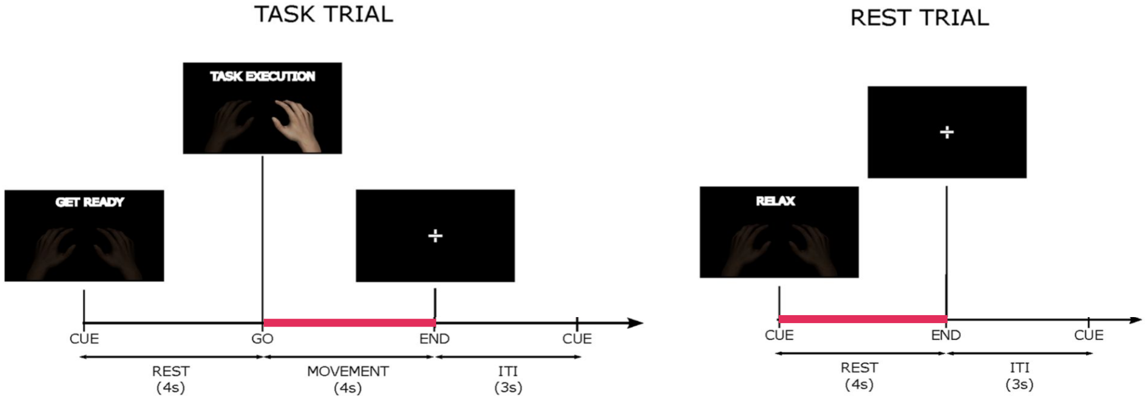


GraspL: **Rest and Task**

For each subject, the recording was repeated randomly for 40 times (i.e., 40 trials):

- **20 task trials:** subject performs `GraspL` movement;
- **20 rest trials:** subject at rest;





APPENDIX

CorticoMuscularCoherence (CMC)

CMC features combines a **EEG** channel `x` and a **EMG** channel `y`:

$$CMC_{xy}(f) = |S_{xy}(f)|^2$$

where `Sxy` is the Cross Power Spectral Density

CMC corresponds to the **linear correlation** between a EEG channel `x` and a EMG channel `y` at a given **frequency** `f`

End of the presentation