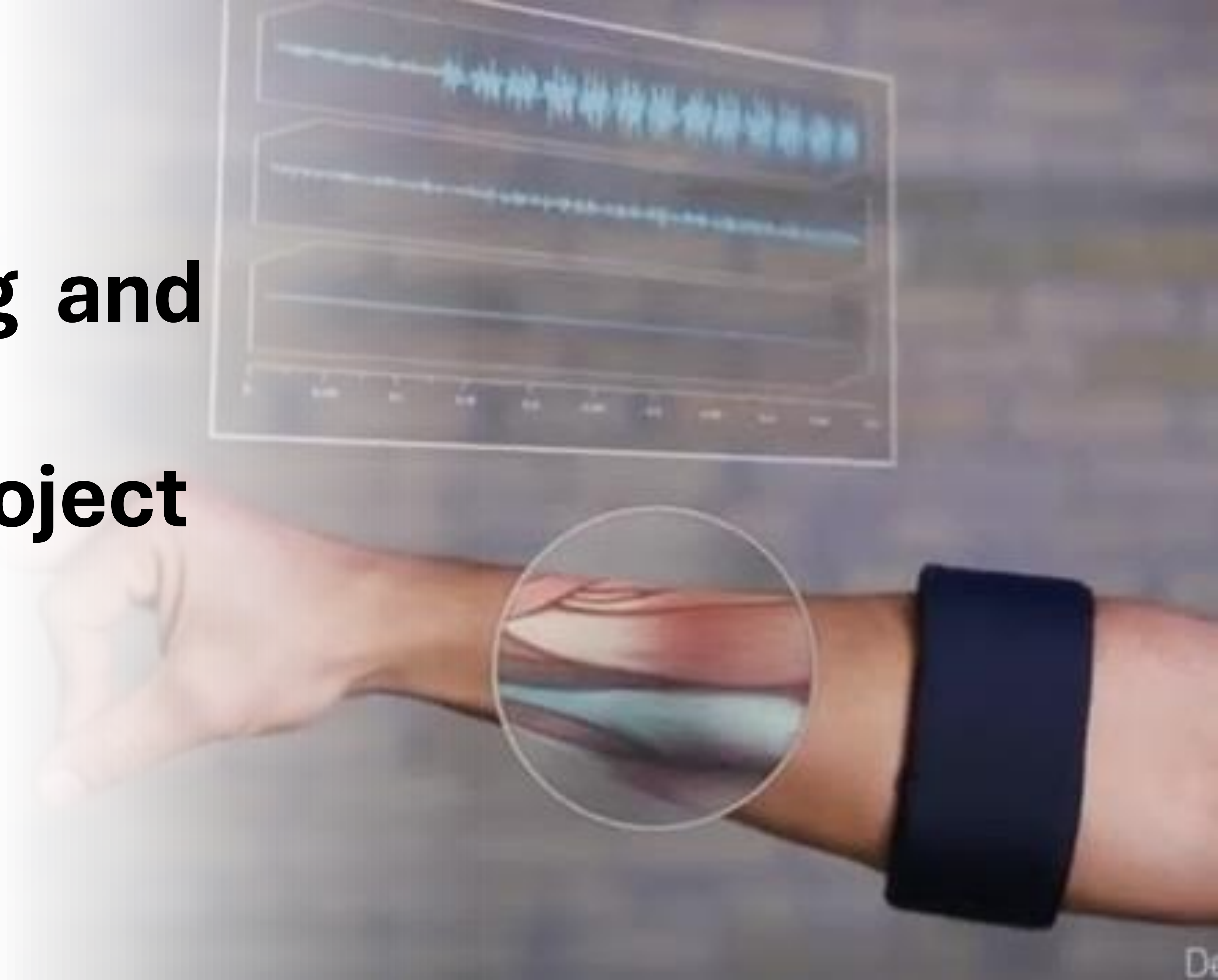
Data Mining and Machine learning project

Alessio Radaelli
Antonella Leoci



Core Problem

The core problem is designing a decoder capable of extracting meaningful commands from physiological data. Specifically, the task is to transform raw Electromyography (EMG) signals—the electrical activity produced by skeletal muscles—into discrete hand-gesture representations.

Using the EMG-EPN-612 dataset, the goal is to build a robust decoding system that can distinguish six states (five active gestures plus a relaxed state) recorded with a Myo armband. The primary difficulty stems from the substantial biological variability among 612 individuals, which makes it challenging to develop a decoder that generalizes reliably to new users without extensive recalibration.

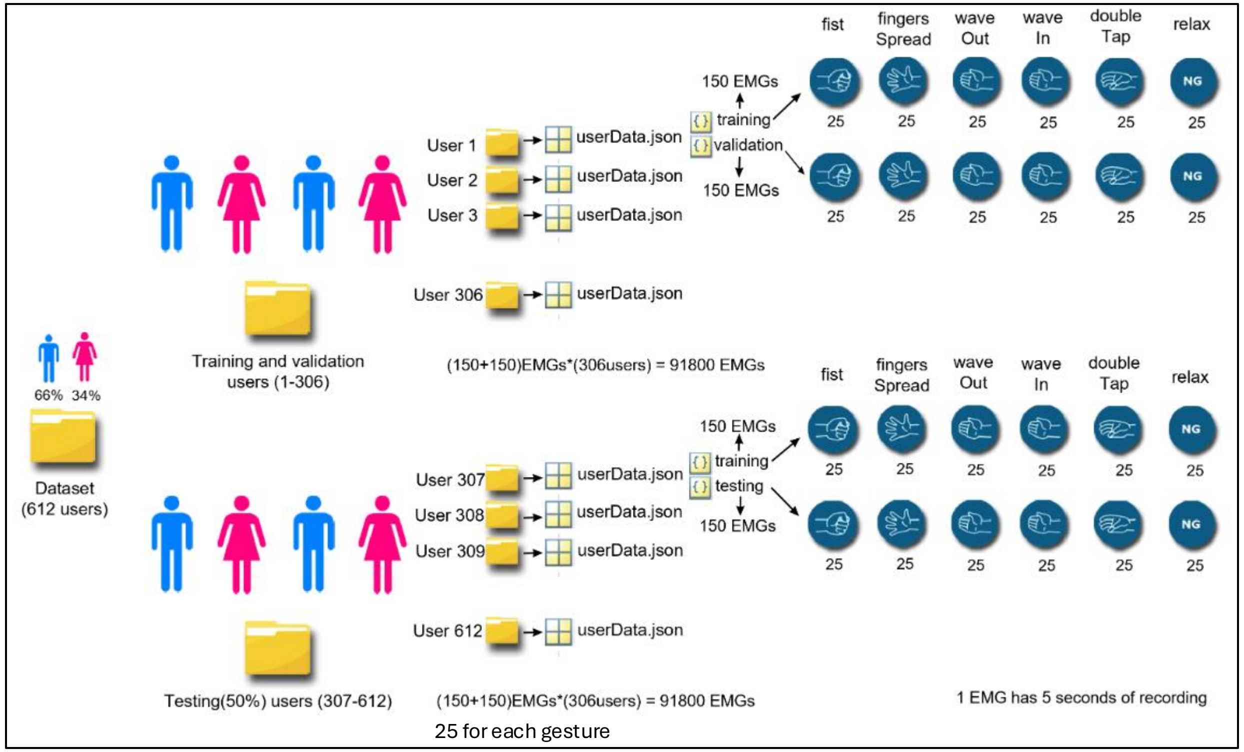
Project final aim

The final goal is to develops a data-driven decoder that classifies six hand-gesture states from EMG signals recorded with a Myo armband. Using the EMG-EPN-612 dataset, the focus is on handling large inter-subject variability and applying DMML techniques to build a gesture recognizer that generalizes to new users with minimal calibration.

Relevance of the DMML techniques :

- Feature Extraction Techniques (time and frequency domain)
- Feature selection (to deal with high-dimensional EMG data)
- Supervised Classification Techniques (rf,knn,svm investigation of many models)
- Post-hoc techniques to explain opaque model's prediction(e.g. SHAP)

DATASET



EMG-EPN-612 DATASET :

- contains EMG signals of **612** people for benchmarking of hand gesture recognition systems.
- The data recorded with the 8-channel Myo armband EMG signals on the forearm at 200 hz
- **6 gesture :** five hand gestures**(** wave-in, wave-out, pinch, open and fist) and hand relaxed
- The dataset is divided into two groups of 306 people each. One group is for training or designing hand gesture recognition models and the other is intended for testing the classification and recognition accuracy of hand gesture recognition models.
- each person has **50 EMGs for each of the 5 gesture** recorded and **also 50 EMGs for the hand relaxed**. (instead are 25 EMGs for the testing set)

Description of input – output format

Input format :

- Numeric(Integer and Discrete) for the EMG (ch1 through ch8): Electromyography data measuring muscle electrical activity.

-**Numeric (Floating-Point**)

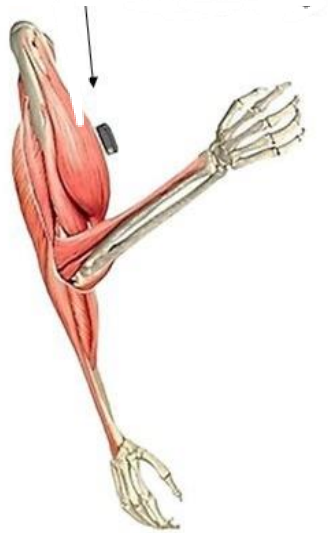
- Accelerometer (x, y, z) measures acceleration.
- Gyroscope (x, y, z), measures rotation rate.
- Quaternion (w, x, y, z) measures orientation in 3D space

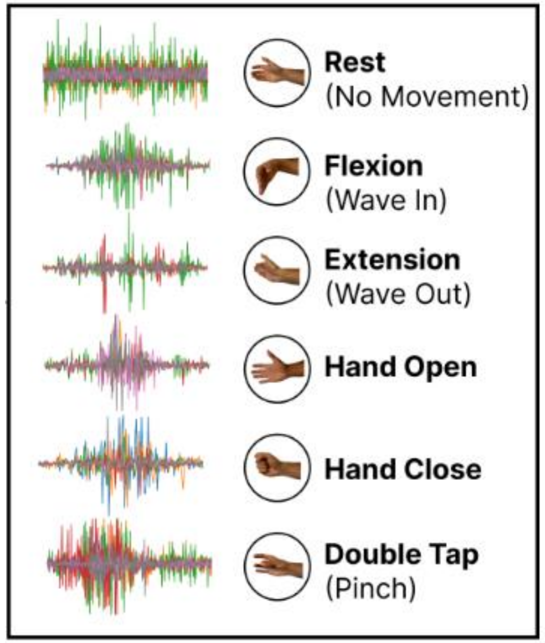
-**Categorical,** in `generalInfo` section defines the classes

-Metadata (Mixed) also present in the demographic info and user info


```
"generalInfo": {
  "deviceModel": "Myo Armband",
  "samplingFrequencyInHertz": 200,
  "recordingTimeInSeconds": 5,
  "repetitionsForSynchronizationGesture": 5,
  "myoPredictionLabel": {
    "noGesture": 0,
    "fist": 1,
    "waveIn": 2,
    "waveOut": 3,
    "open": 4,
    "pinch": 5
  }
```


Output format:Categorical






```
"userInfo": {
  "name": "user1",
  "age": 19,
  "gender": "man",
  "occupation": "student",
  "ethnicGroup": "latin",
  "handedness": "right",
  "ArmDamage": "False",
  "distanceFromElbowToMyoInCm": 6,
  "distanceFromElbowToUlnaInCm": 24,
  "armPerimeterInCm": 23,
  "date": "30-Oct-2019 14:34:28"
},
```

References

https://zenodo.org/records/4421500

Frontiers | Big data in myoelectric control: large multi-user models enable robust zero-shot EMG-based discrete gesture recognition

[2409.07484] FORS-EMG: A Novel sEMG Dataset for Hand Gesture Recognition Across Multiple Forearm Orientations