Machine learning applicato ad un esperimento fMRI

Simone Romano[1]

[1] Università degli studi di Salerno

Supervisori:
Prof. Roberto Tagliaferri
Prof. Fabrizio Esposito
Dott. Giancarlo Valente
Dott. Angela Serra

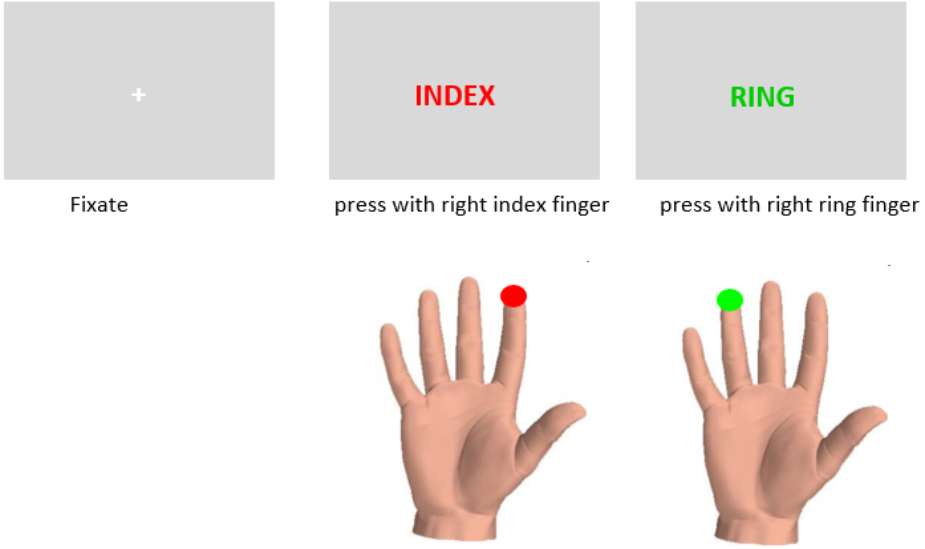
Tesi magistrale Informatica

Outline

1. **Introduzione**
2. **Metodologie**
3. **Risultati**
   - Errori cross-validation
   - Errori one shot
4. **Conclusioni**

Dati

- Un campione ogni 2 secondi
- Dimensione spaziale ~10^3 voxel
- Il segnale misurato per una singola pressione è rappresentato con una matrice **t**x**v**:
  - t = time, 8 samples
  - v = voxels, approx 10^3
- Abbiamo 48 risposte per ogni dito per un totale di 96 risposte



Obiettivi

Obiettivi:

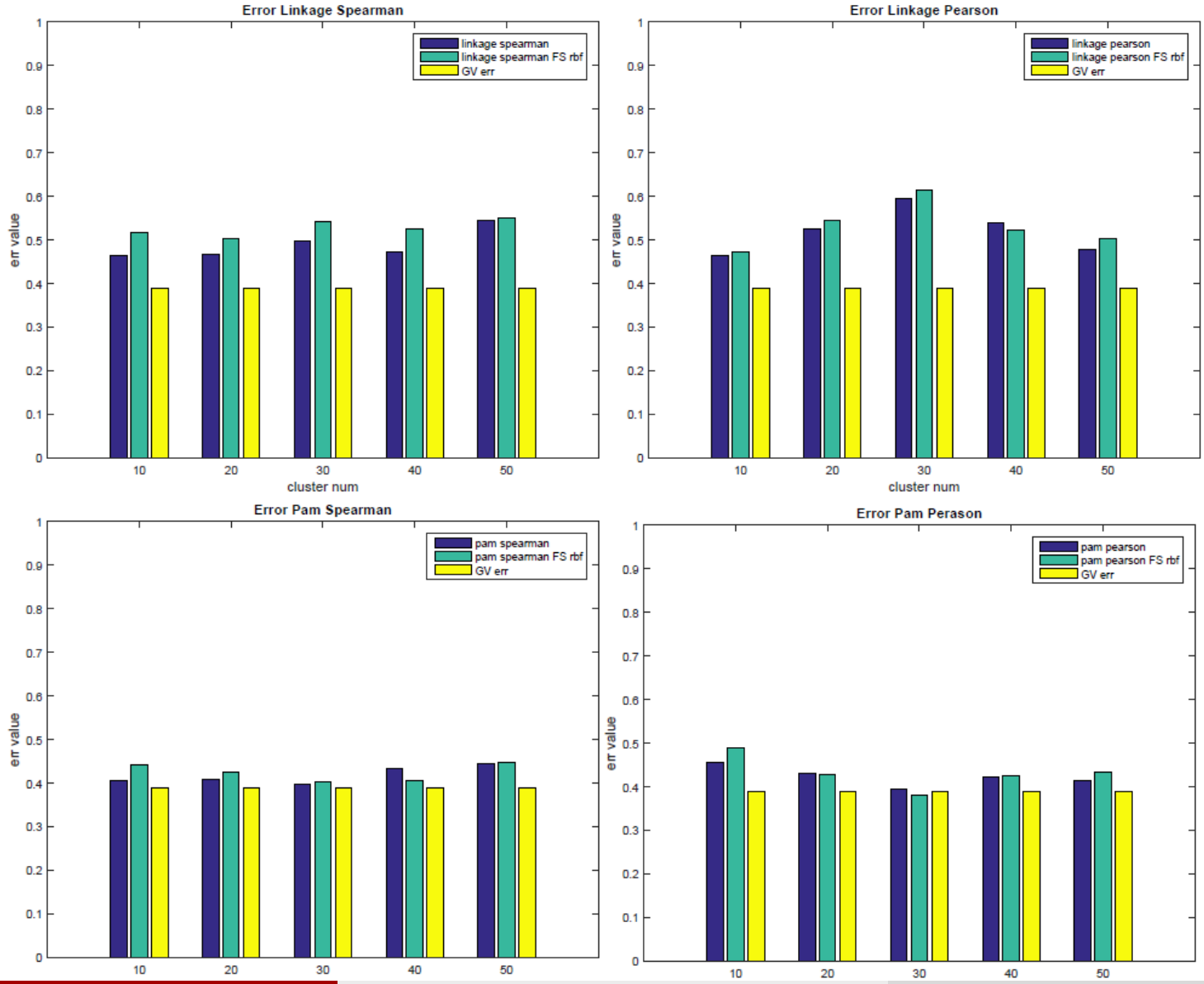
- clusterizzazione voxel
- generazione mappe
- stima risposta emodinamica
- classificazione trial (2 diverse classi)

Lavoro precedente

Pipeline:

- Compressione serie temporali utilizzando la **media** dei valori (2-7)
- Clusterizzazione voxel
  - **Pam** e **Linkage** con dissimilarità di Pearson e Spearman
- 10 fold cross-validation (lineare/non lineare)

Risultati lavoro precedente



Outline

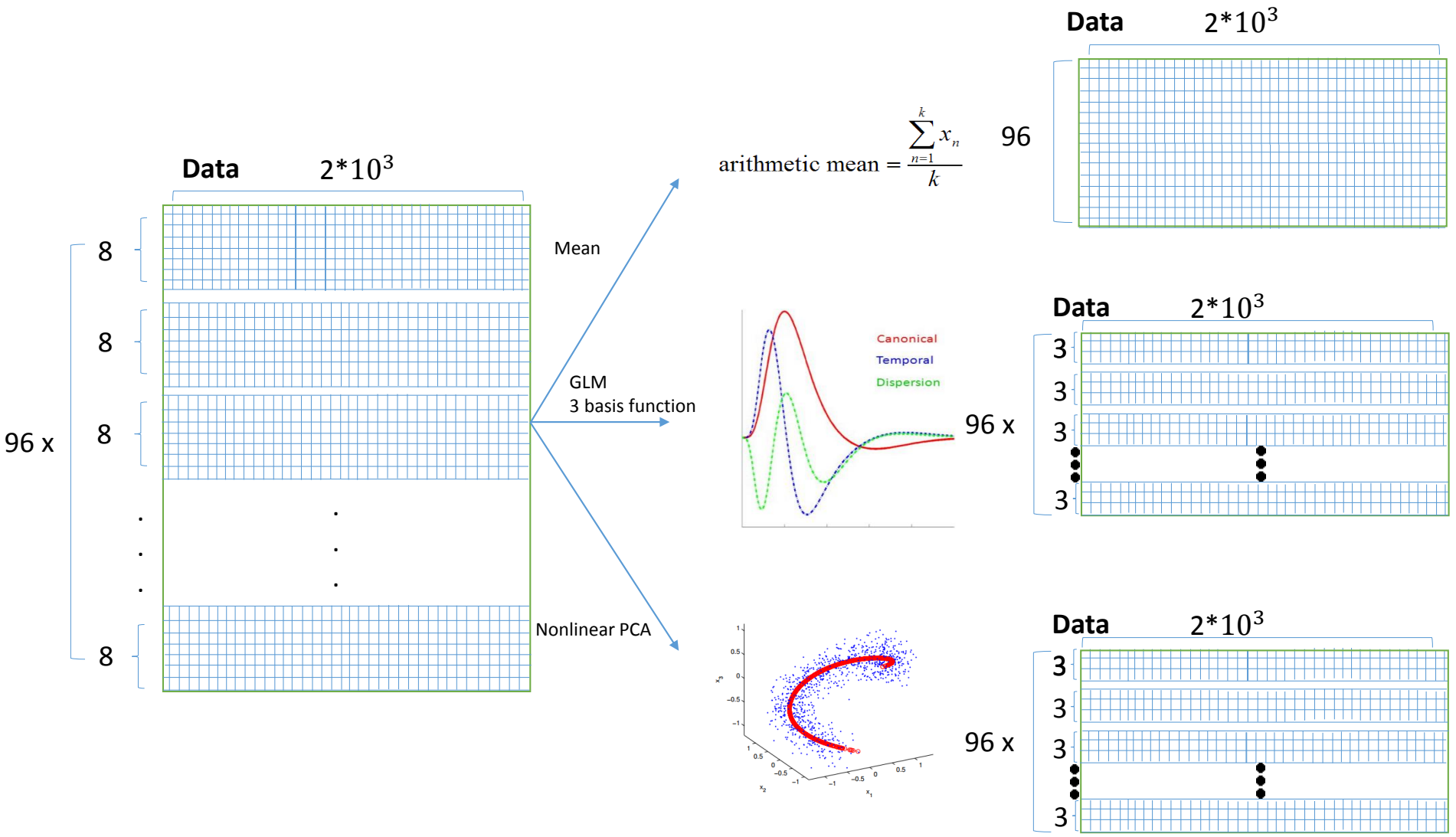
1. Introduzione
2. **Metodologie**
3. Risultati
   - Errori cross-validation
   - Errori one shot
4. Conclusioni

Approcci utilizzati

Nuovi approcci:

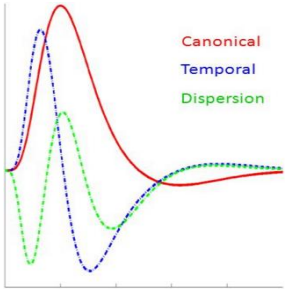
- Ridurre il numero di cluster
- Compressione serie temporali utilizzando
  - GLM su 3 basi [1]
  - PCA non lineare [2]
- Nuovo schema di cross-validation

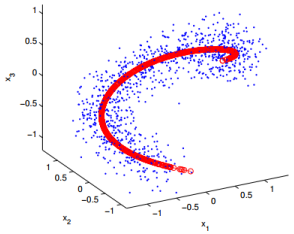
Compressione serie temporali

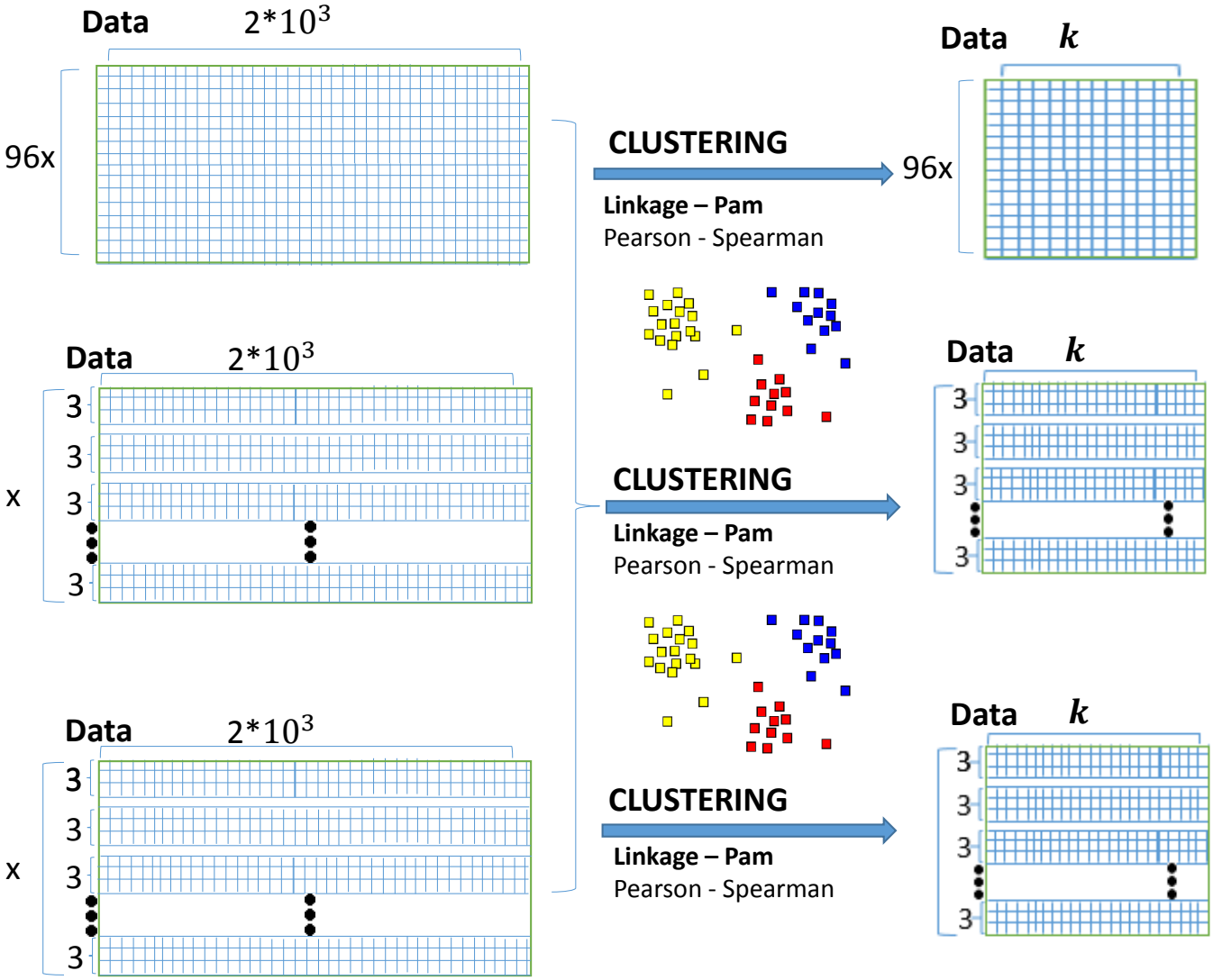


Clustering voxel

$$\text{arithmetic mean} = \frac{\sum_{n=1}^{k} x_n}{k}$$





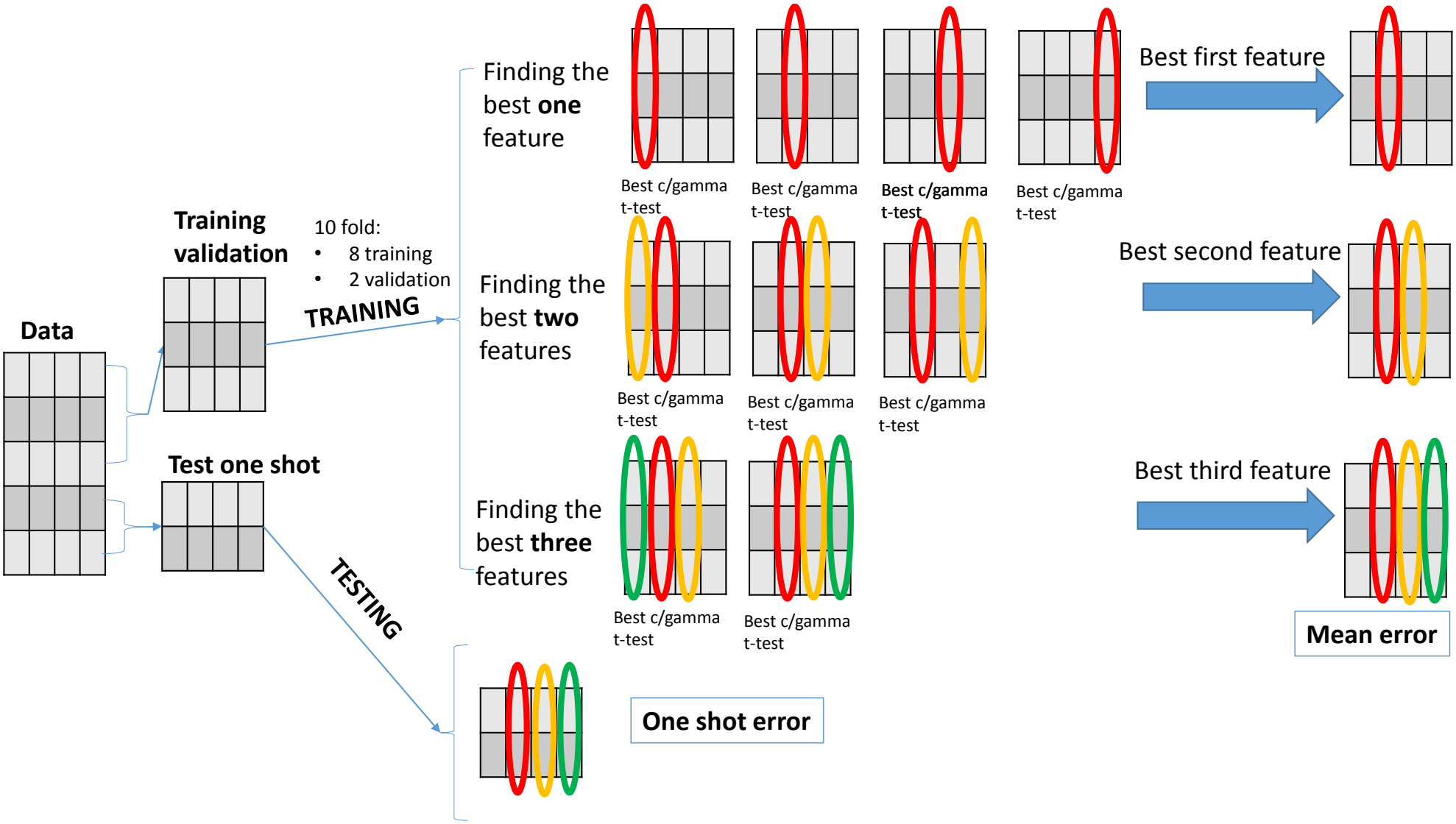


Cross-validation

Lo schema di cross-validation utilizzato è descritto di seguito:

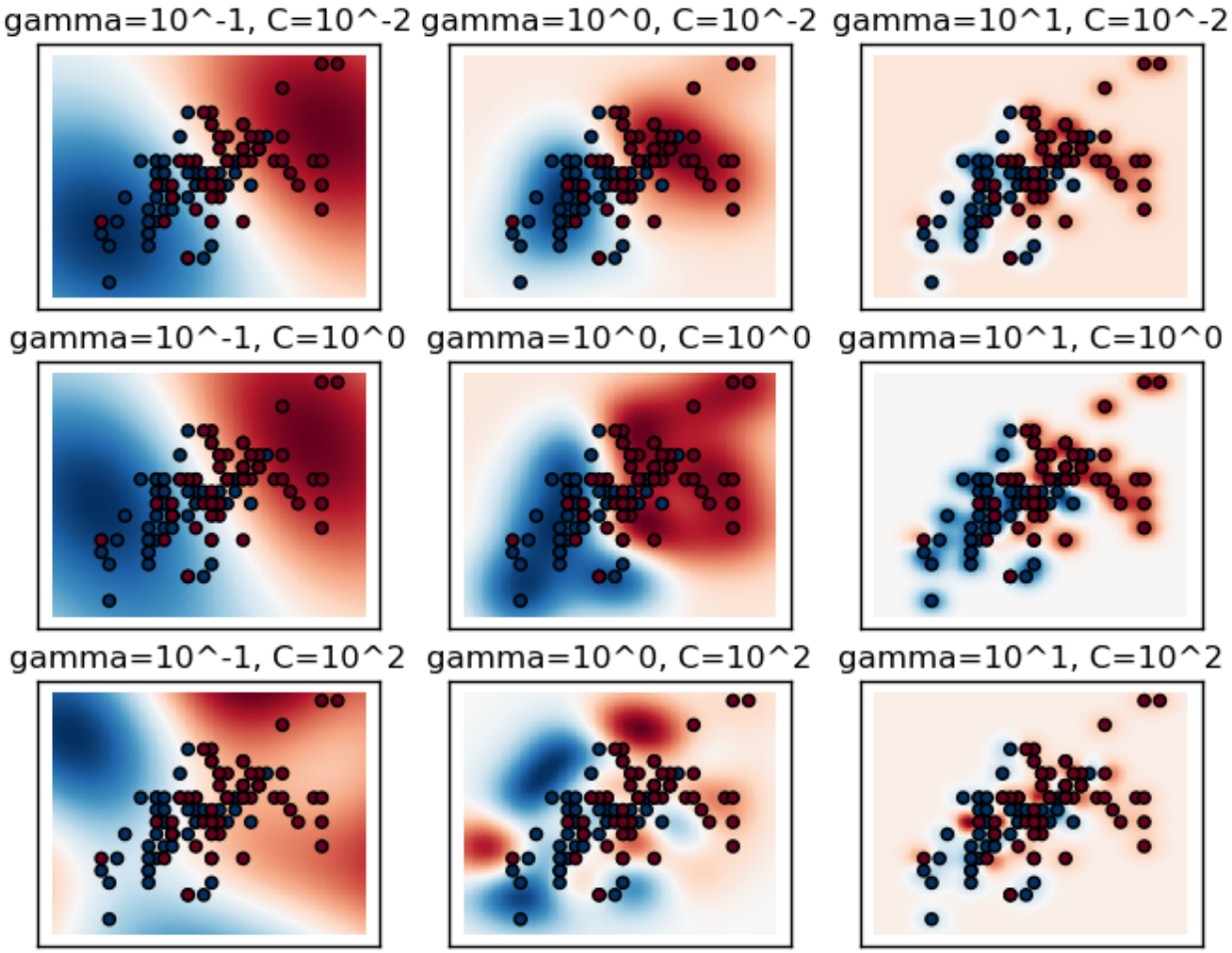
- Divisione dataset in:
  - **training/validation** - 80 elementi
  - **testing** - 16 elementi (one shot)
- Training:
  - Divisione in 10 fold
    - 8 training
    - 2 validation
  - Features-selection/scelta parametri c e γ in cross-validation (t-test)
    - $c \in \{10^-3, 10^-2, 10^-1, 10^0, 10^1\}$
    - $\gamma \in \{10^-3, 10^-2, 10^-1, 10^0\}$
  - **Output:** errore medio
- Test:
  - Training svm su 80 elementi di training/validation con parametri/features trovate in precedenza
  - Testing one shot sui 16 elementi lasciati fuori nel primo step
  - **Output**: errore one-shot

Cross-validation - 1



Parametri svm

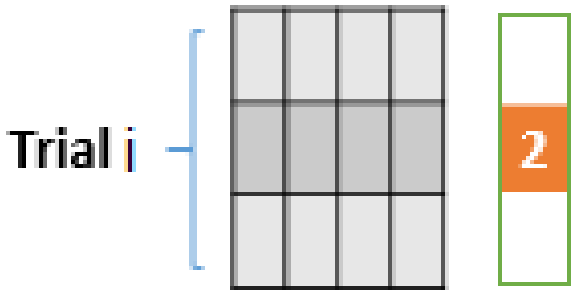
Influenza dei parametri c e γ sul classificatore svm [3]:



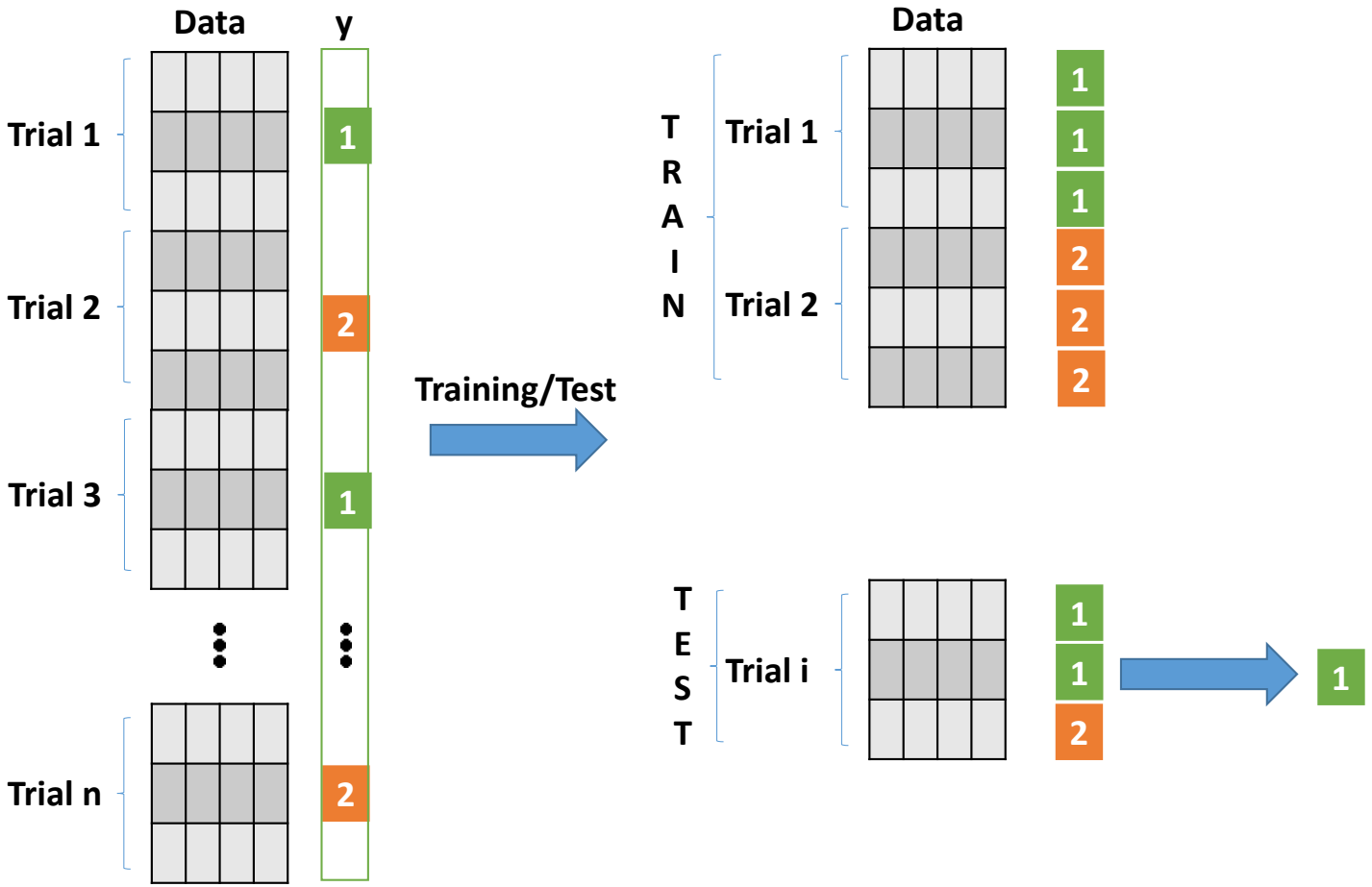
Cross-validation - Multi-pattern

Per i casi in cui le serie temporali sono state compresse con GLM o con PCA non lineare, la classificazione è stata fatta a votazione:

- Per ogni trial *i*
  - utilizzo del modello su ognuno dei 3 pattern del trial
  - il trial è assegnato alla classe *vincente* 2 volte su 3



Cross-validation - Multi-pattern - 1



Outline

1. Introduzione
2. Metodologie
3. **Risultati**
   - Errori cross-validation
   - Errori one shot
4. Conclusioni

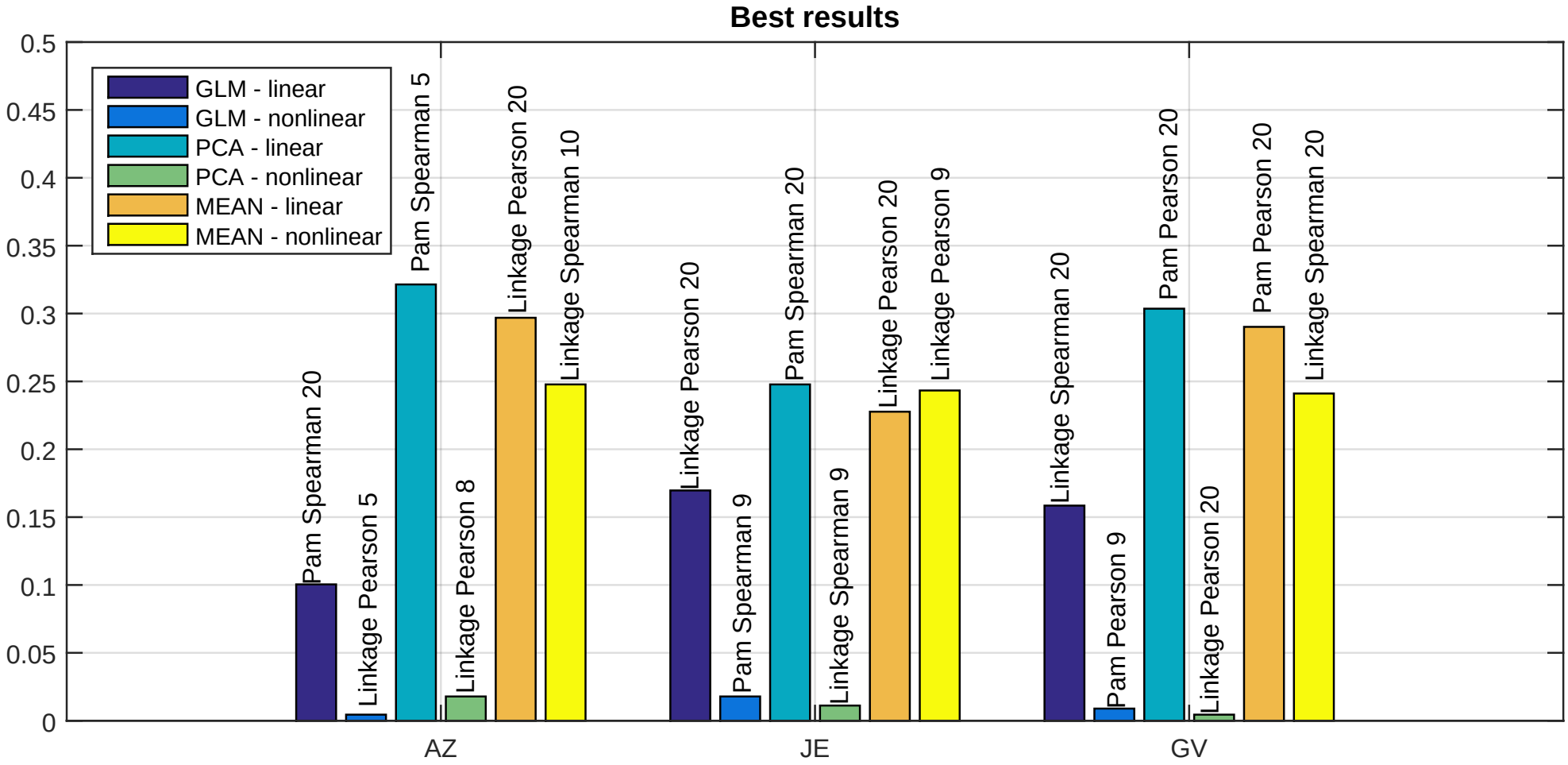
1 Introduzione

2 Metodologie

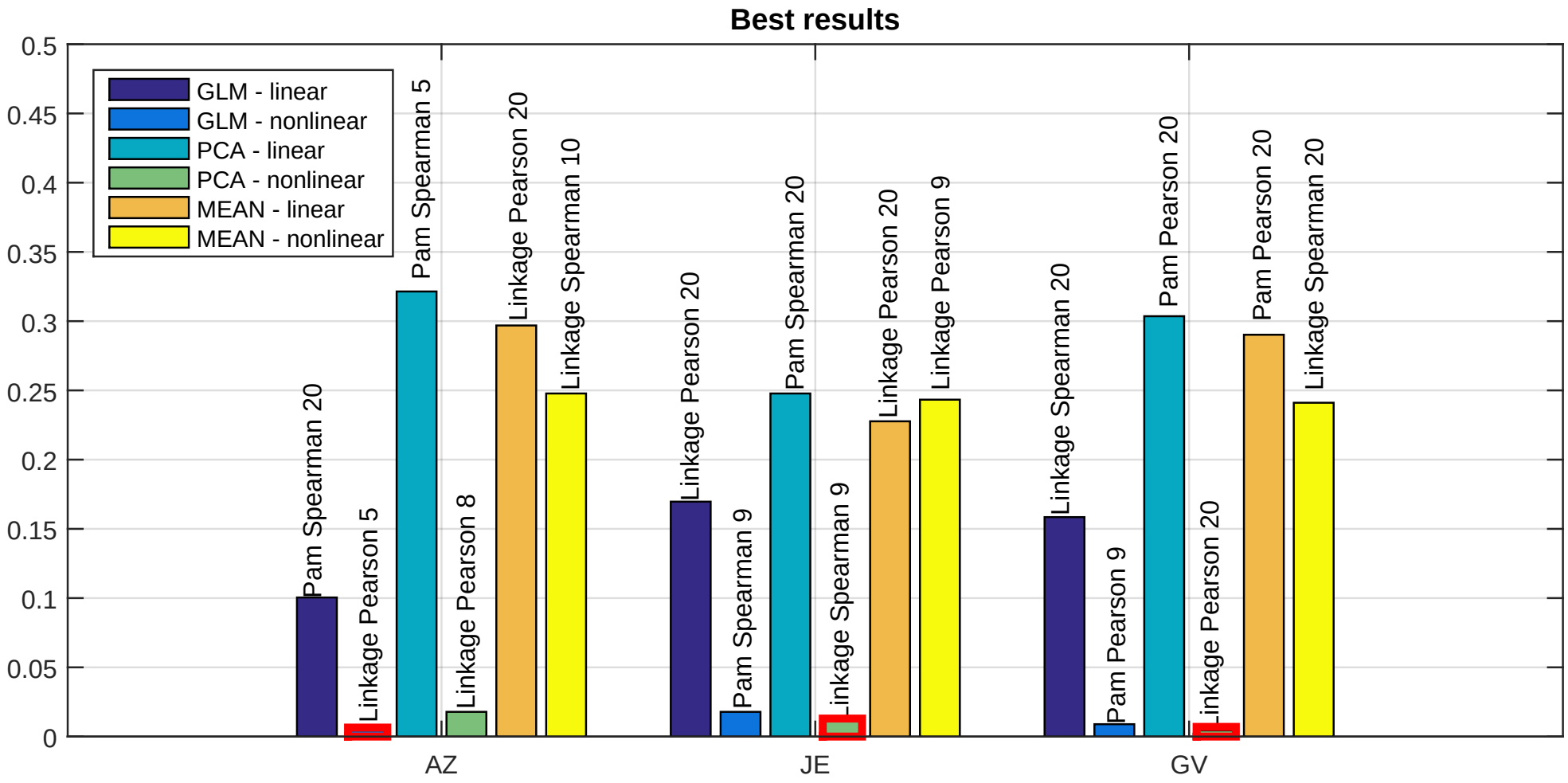
3 **Risultati**
- Errori cross-validation
- Errori one shot

4 Conclusioni

Risultati migliori



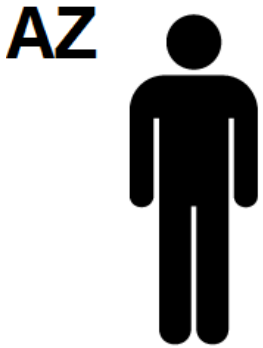
Risultati migliori - 1



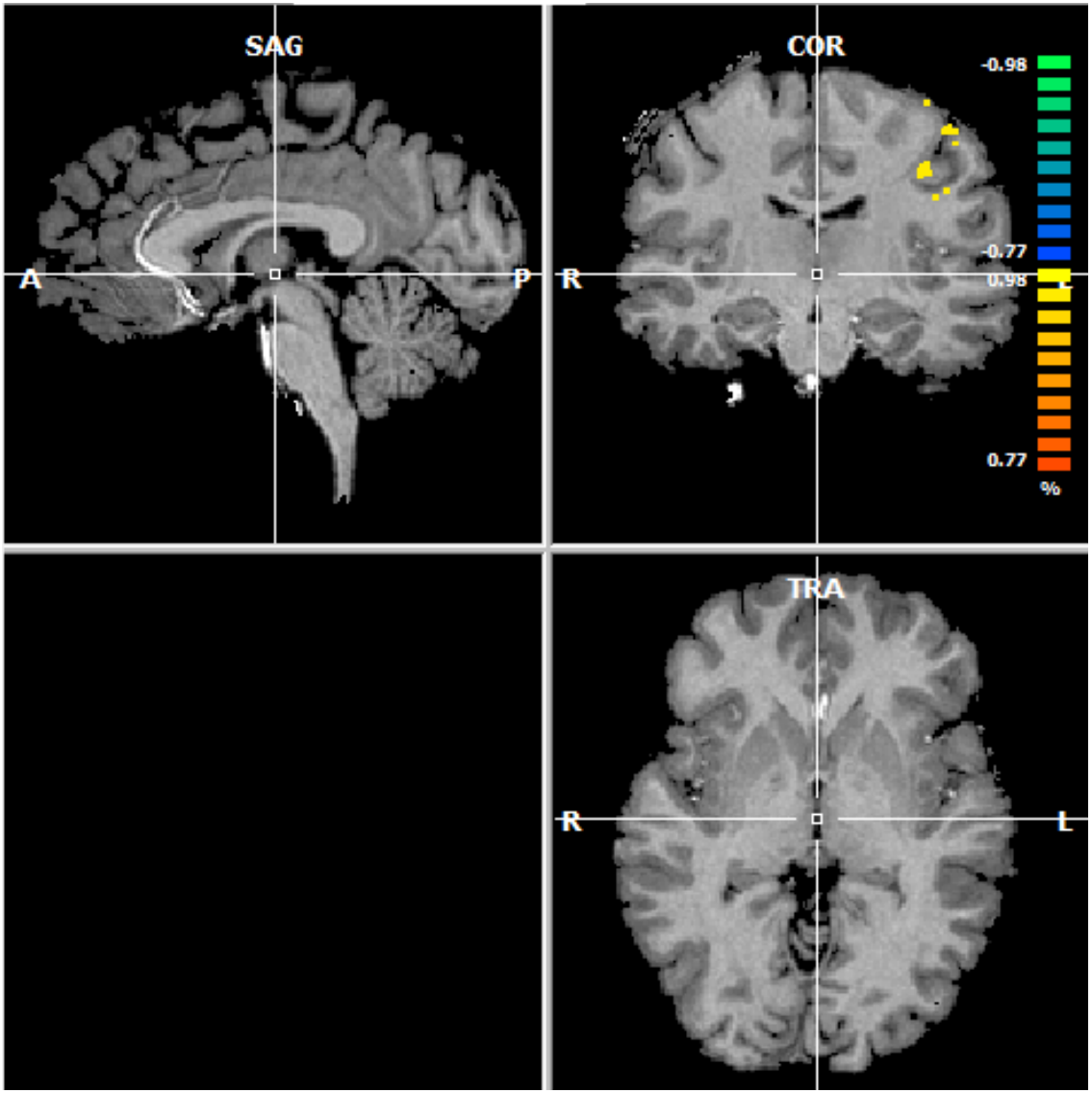
Soggetto AZ - Miglior risultato

Il miglior risultato è stato ottenuto con

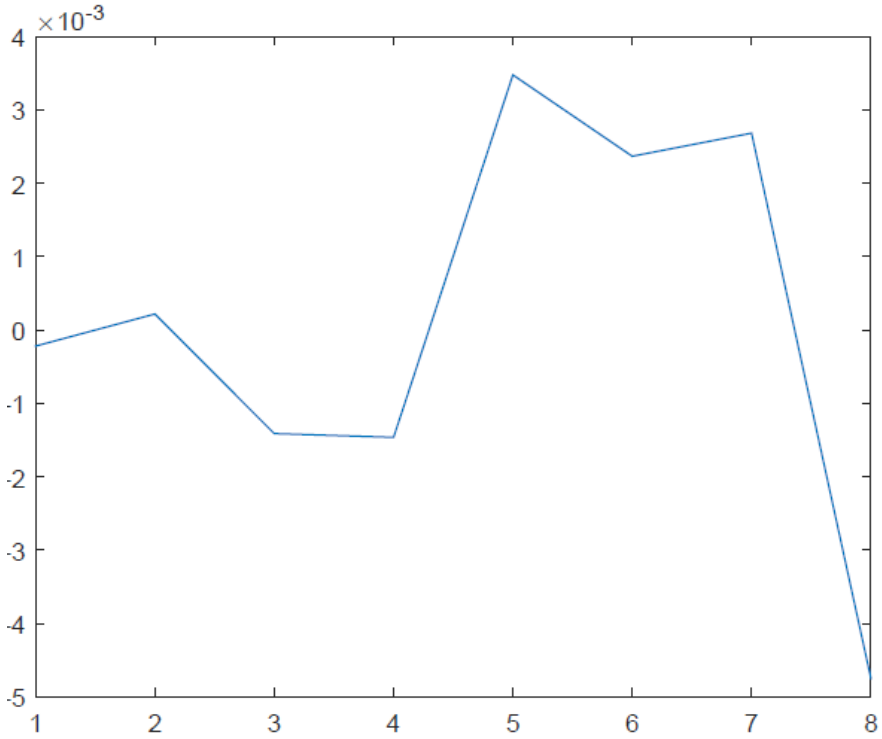
- **Compressione** serie temporale con GLM su 3 basi
- **Clustering** dei voxel con algoritmo Linkage - Pearson
- **Numero di cluster** pari a 5
- **SVM** non lineare
  - c = 10
  - $gamma = 0.001$
- **Miglior feature** = centroide cluster 5
- **#voxel** = 567

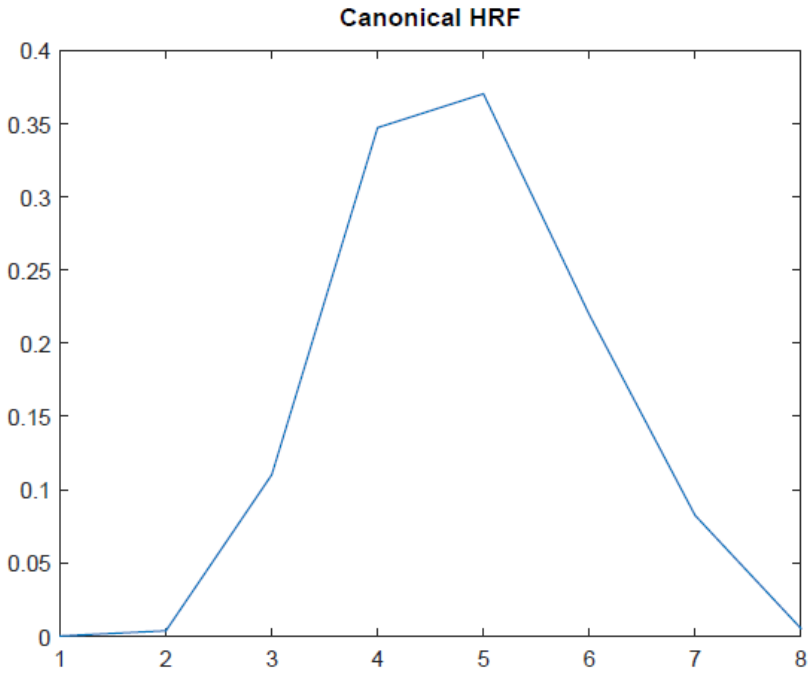


Soggetto AZ - Mappe



Soggetto AZ - Emodinamica

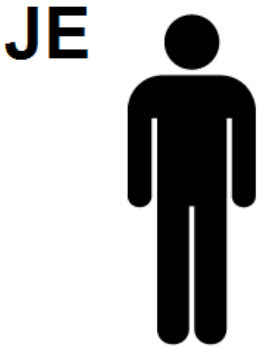




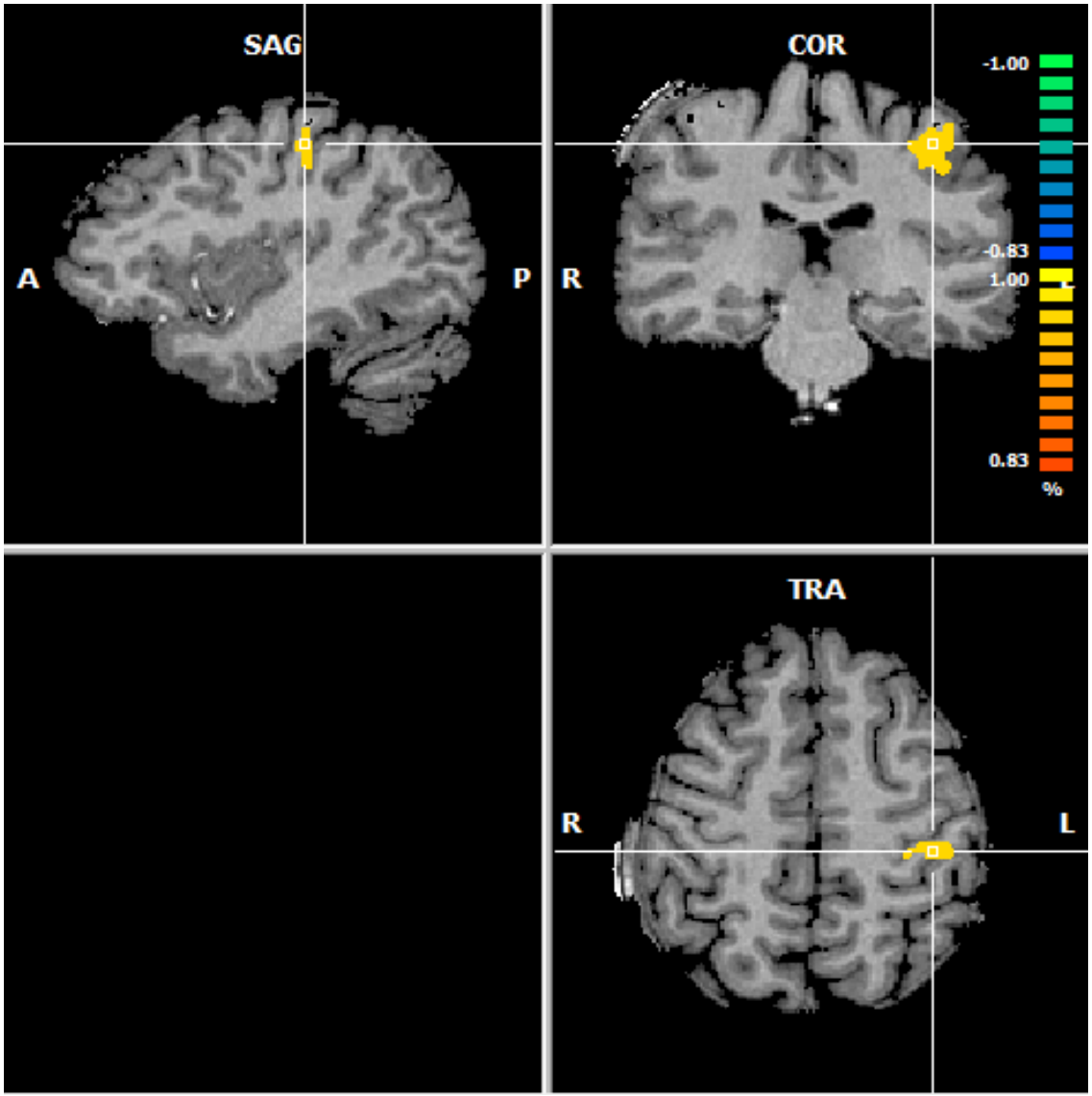
Soggetto JE - Miglior risultato

Il miglior risultato è stato ottenuto con

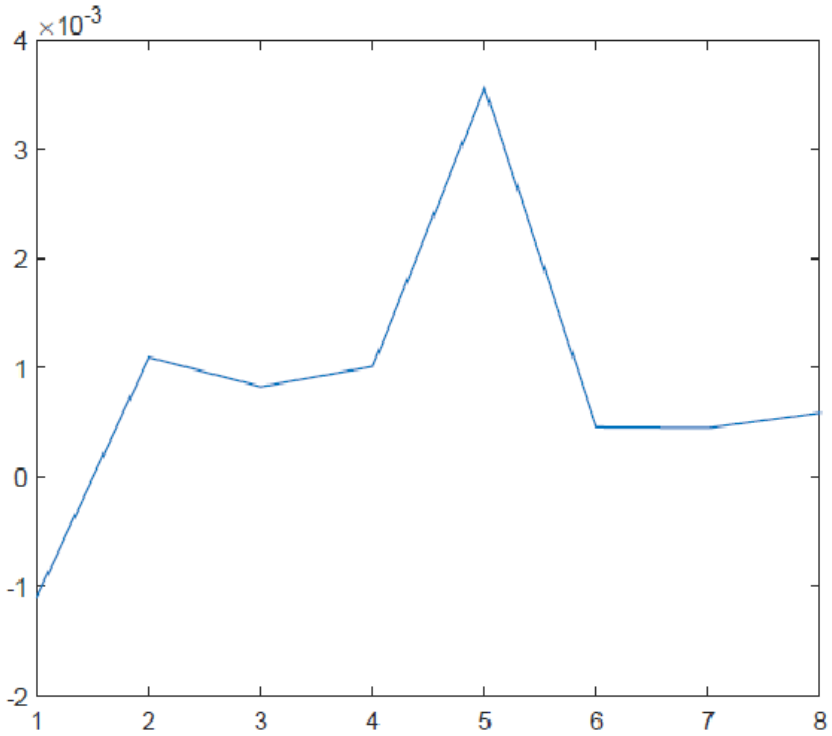
- **Compressione** serie temporale con PCA non lineare
- **Clustering** dei voxel con algoritmo Linkage - Spearman
- **Numero di cluster** pari a 9
- **SVM** non lineare
  - c = 10
  - $gamma = 0.001$
- **Miglior feature** = centroide cluster 4
- **#voxel** = 100

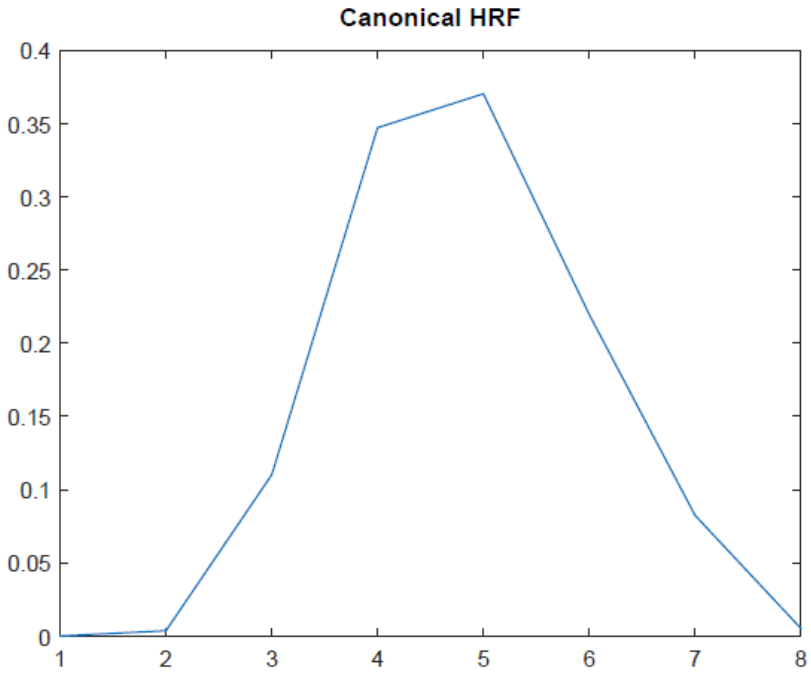


Soggetto JE - Mappe



Soggetto JE - Emodinamica

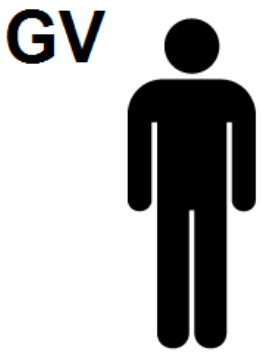




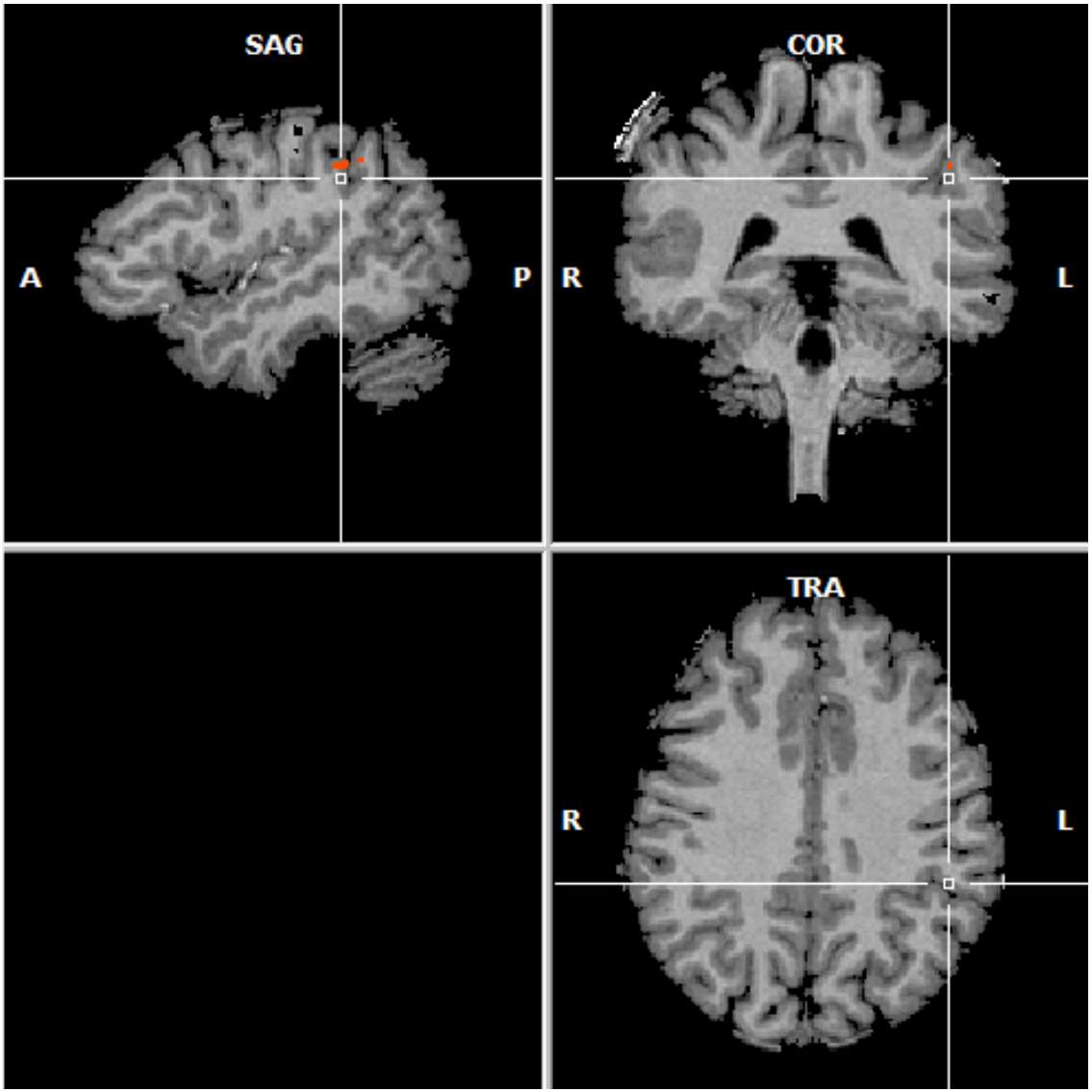
Soggetto GV - Miglior risultato

Il miglior risultato è stato ottenuto con

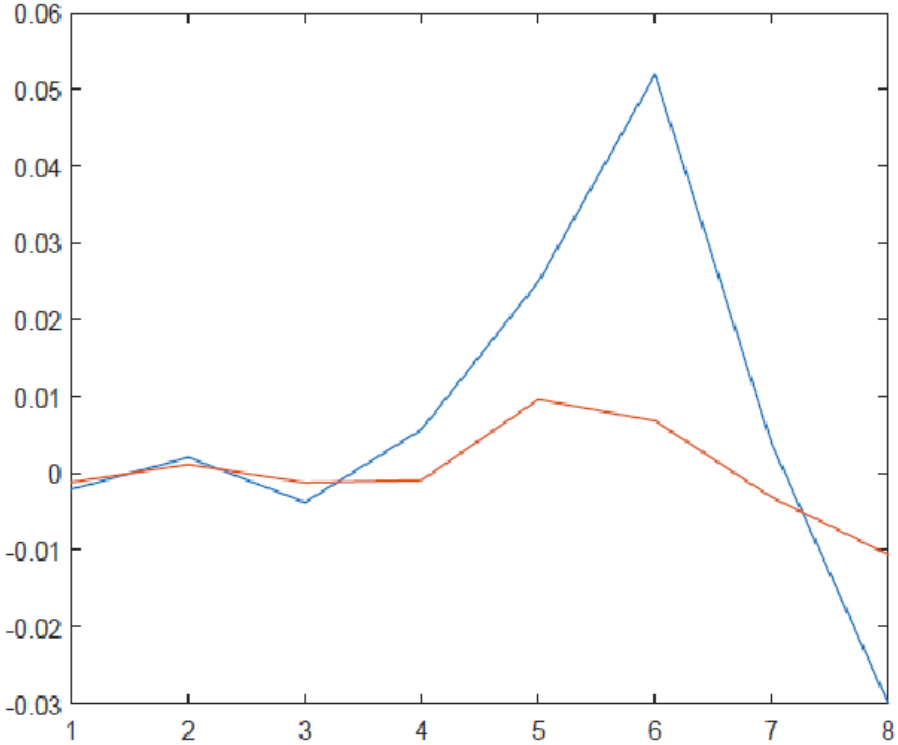
- **Compressione** serie temporale con PCA non lineare
- **Clustering** dei voxel con algoritmo Linkage - Pearson
- **Numero di cluster** pari a 20
- **SVM** non lineare
  - c = 10
  - $gamma = 0.1$
- **Miglior feature** = centroidi cluster 6 e 16
- **#voxel** = 19 + 59

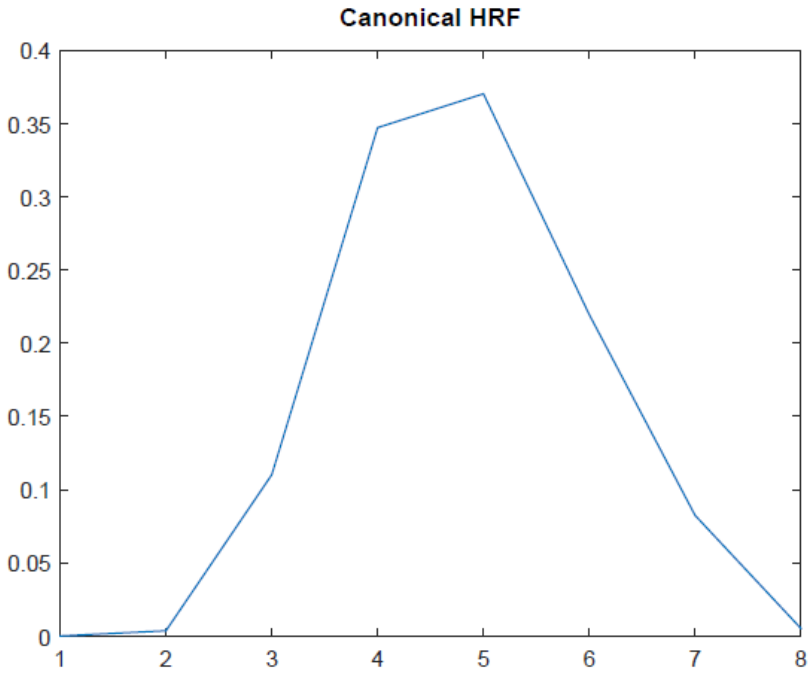


Soggetto GV - Mappe



Soggetto GV - Emodinamica





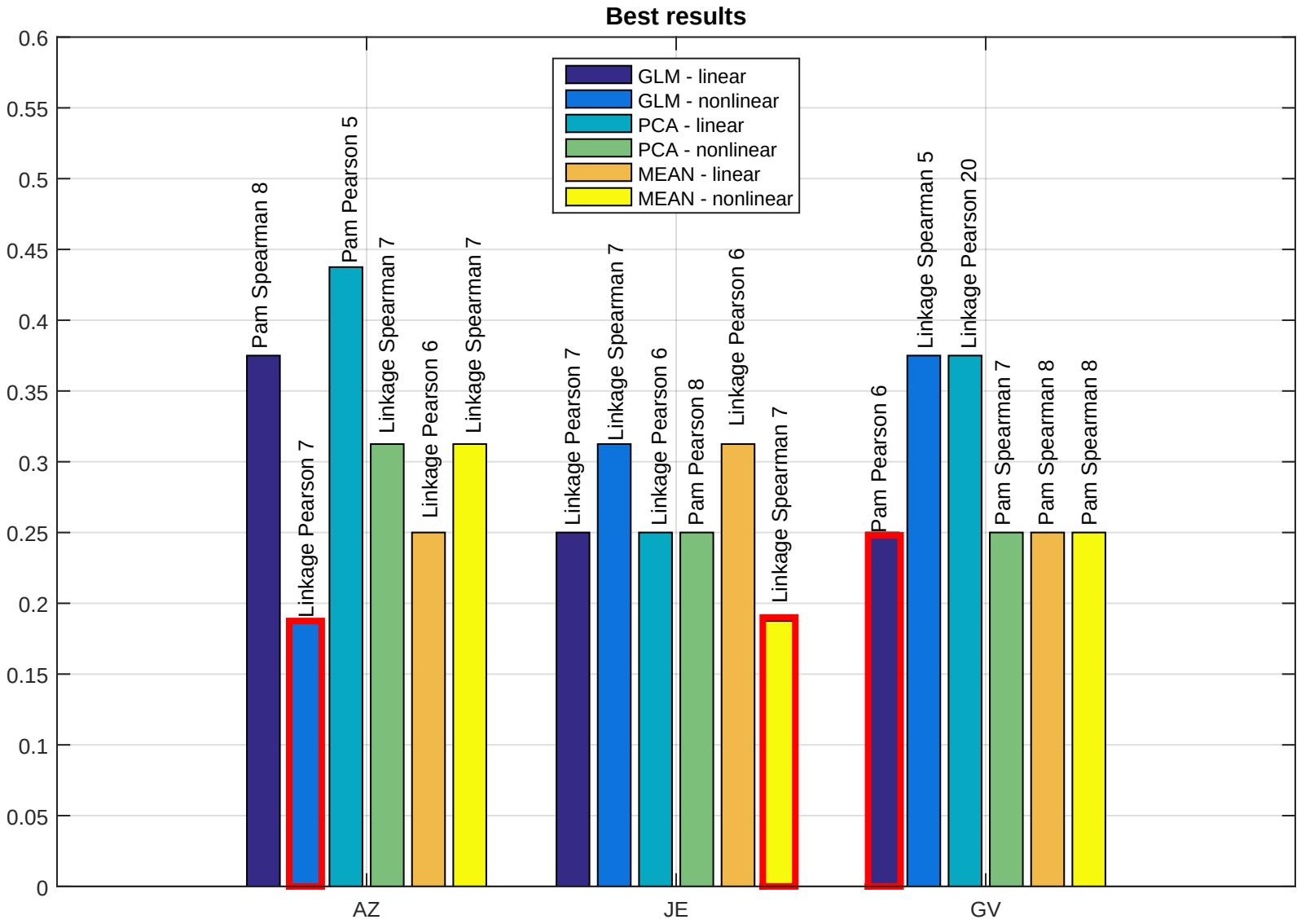
1 Introduzione

2 Metodologie

3 **Risultati**
- Errori cross-validation
- Errori one shot

4 Conclusioni

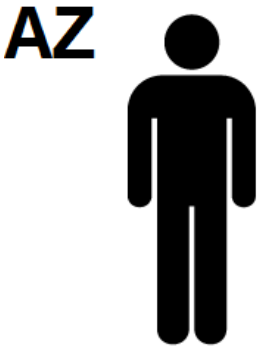
Risultati migliori



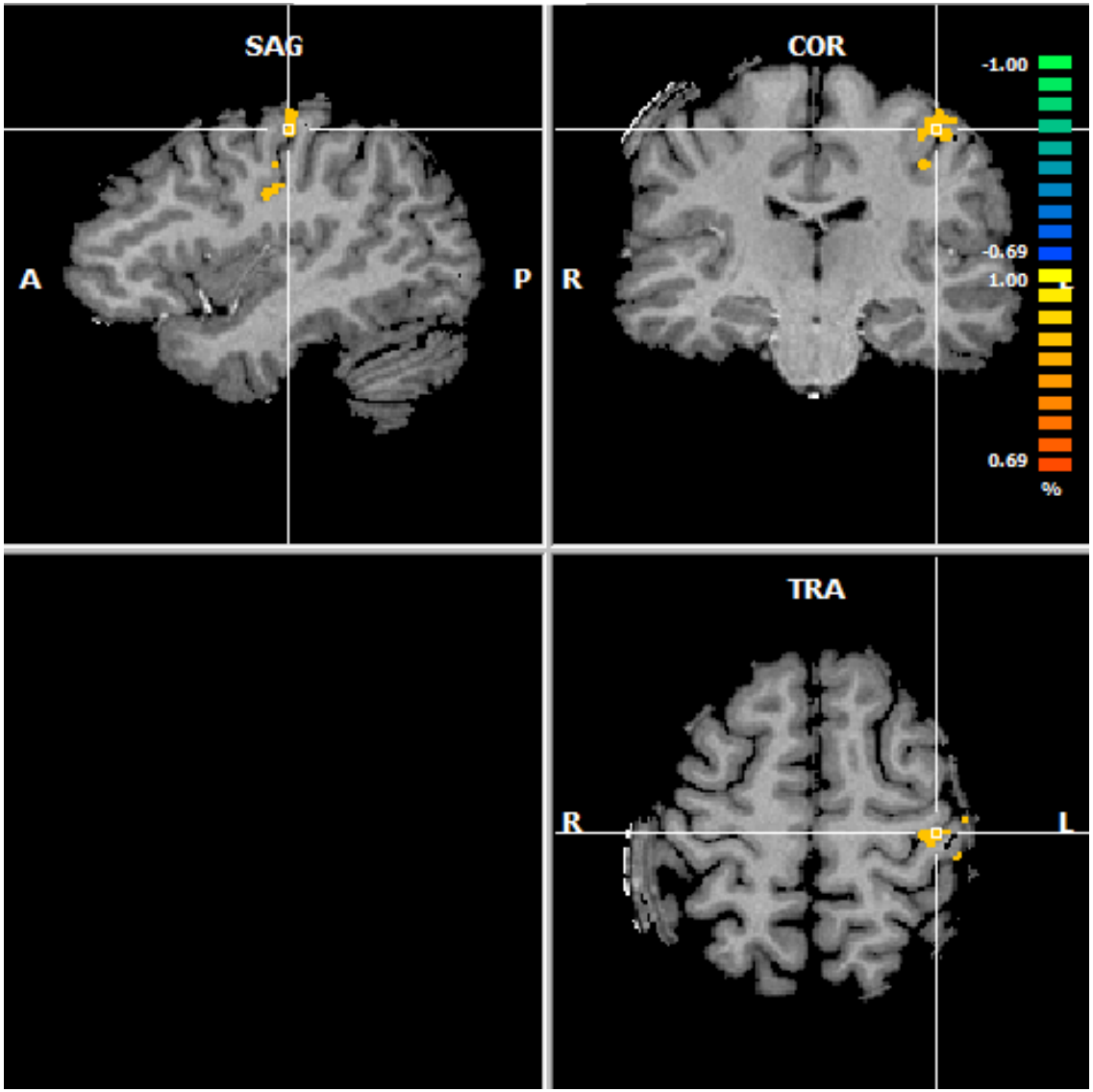
Soggetto AZ - Miglior risultato

Il miglior risultato è stato ottenuto con

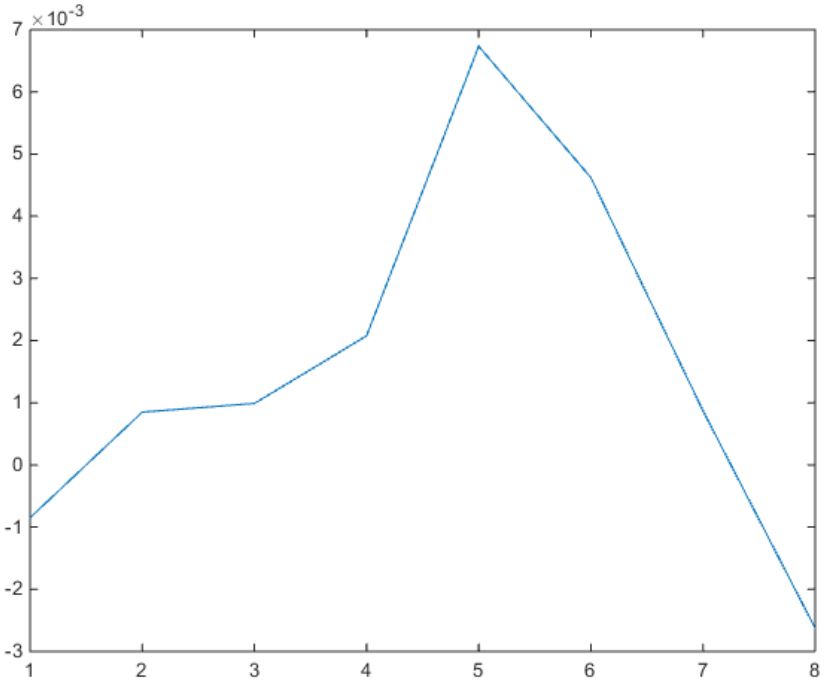
- **Compressione** serie temporale con GLM su 3 basi
- **Clustering** dei voxel con algoritmo Linkage - Pearson
- **Numero di cluster** pari a 7
- **SVM** non lineare
  - c = 1
  - $gamma = 0.01$
- **Miglior feature** = centroide cluster 5
- **#voxel** = 93

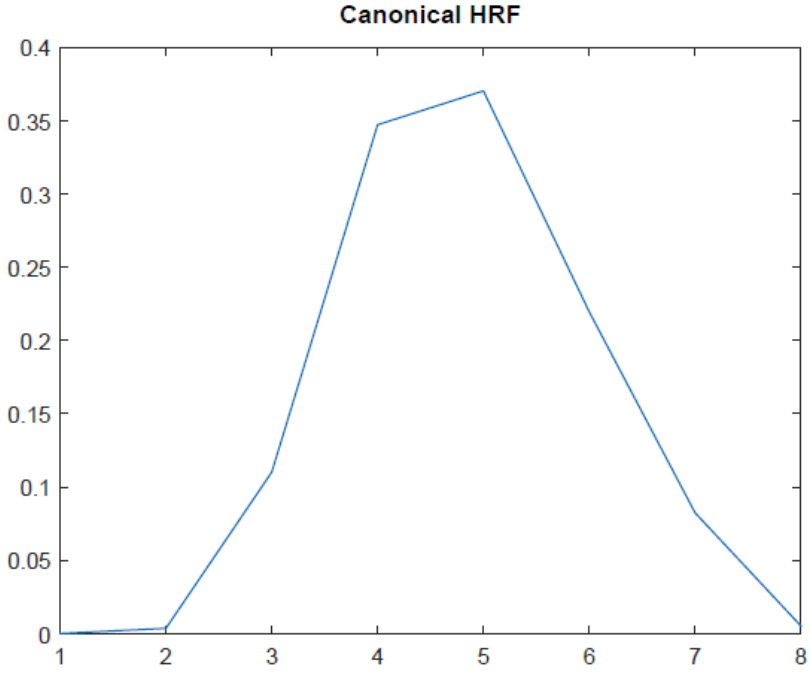


Soggetto AZ - Mappe



Soggetto AZ - Emodinamica

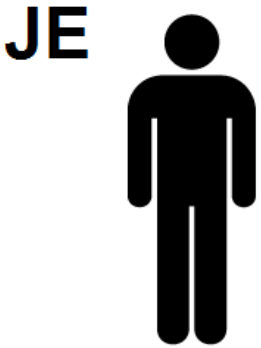




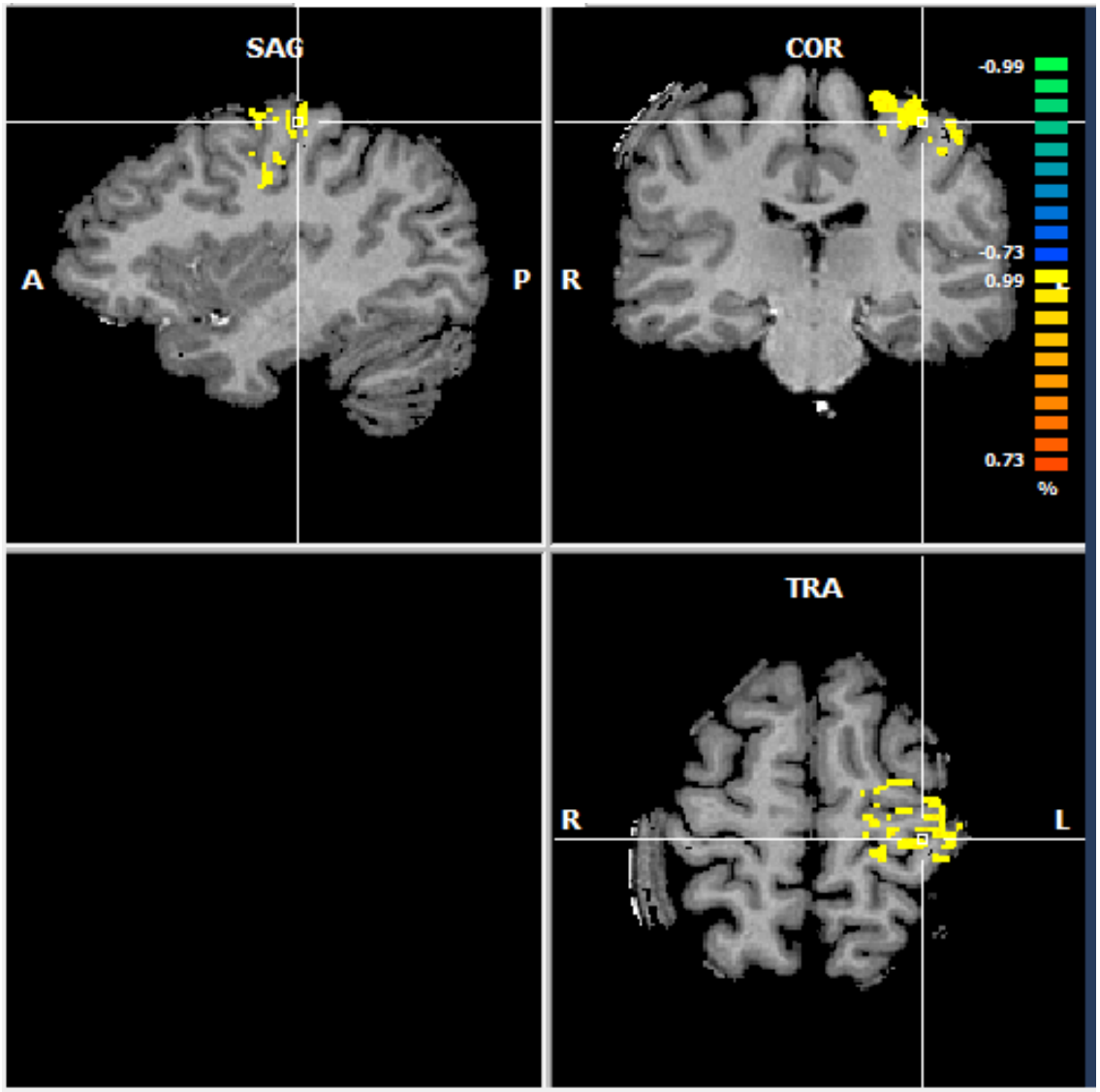
Soggetto JE - Miglior risultato

Il miglior risultato è stato ottenuto con

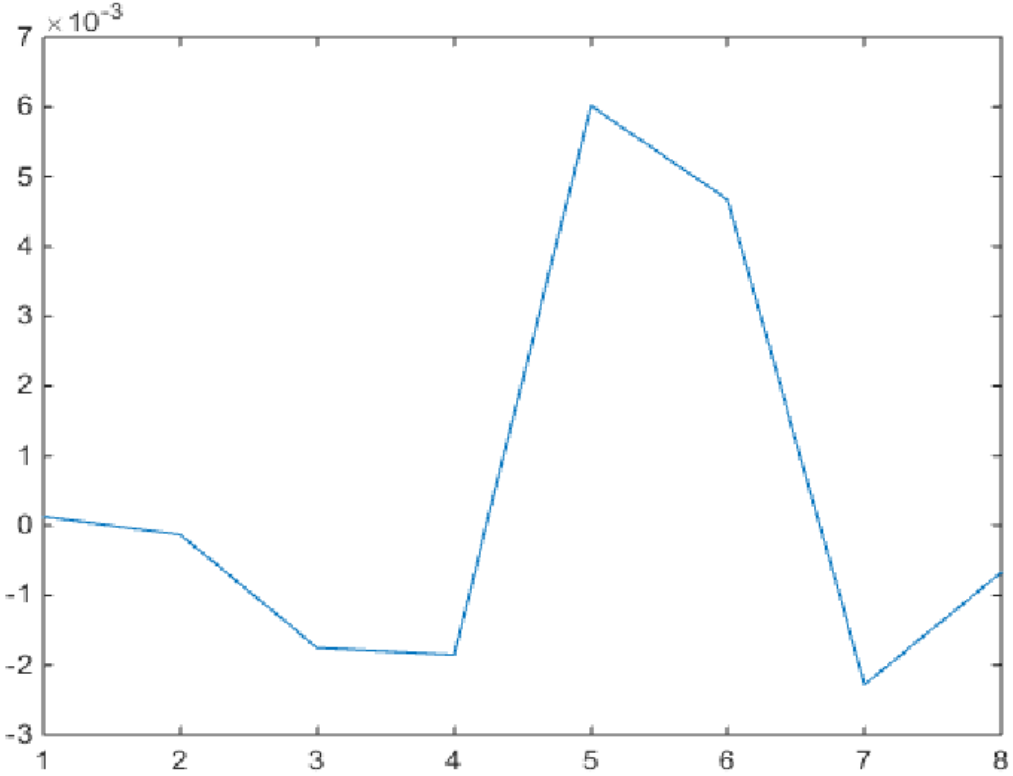
- **Compressione** serie temporale con media
- **Clustering** dei voxel con algoritmo Linkage - Spearman
- **Numero di cluster** pari a 7
- **SVM** non lineare
  - c = 1
  - gamma = 0.1
- **Miglior feature** = centroide cluster 6
- **#voxel** = 673

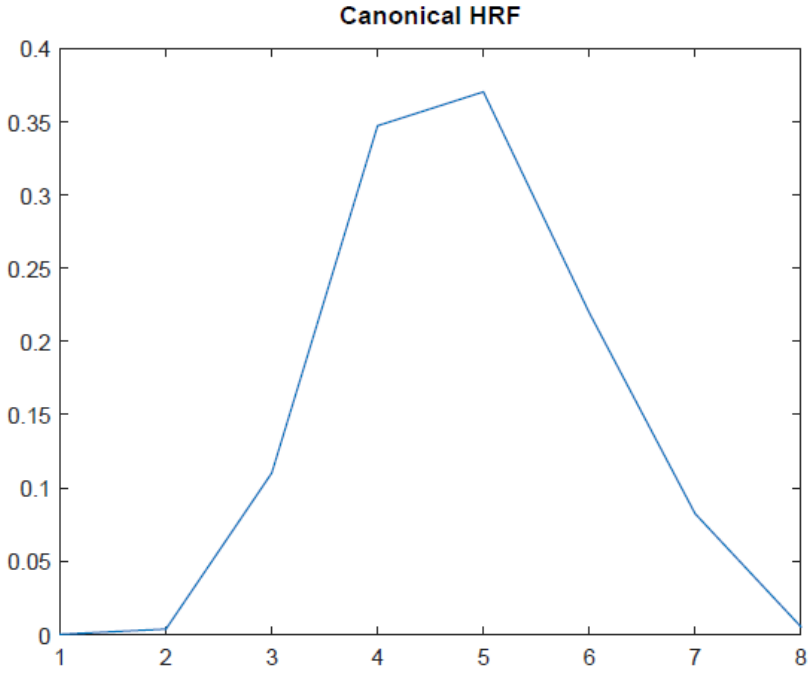


Soggetto JE - Mappe



Soggetto JE - Emodinamica

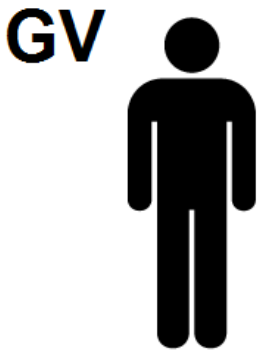




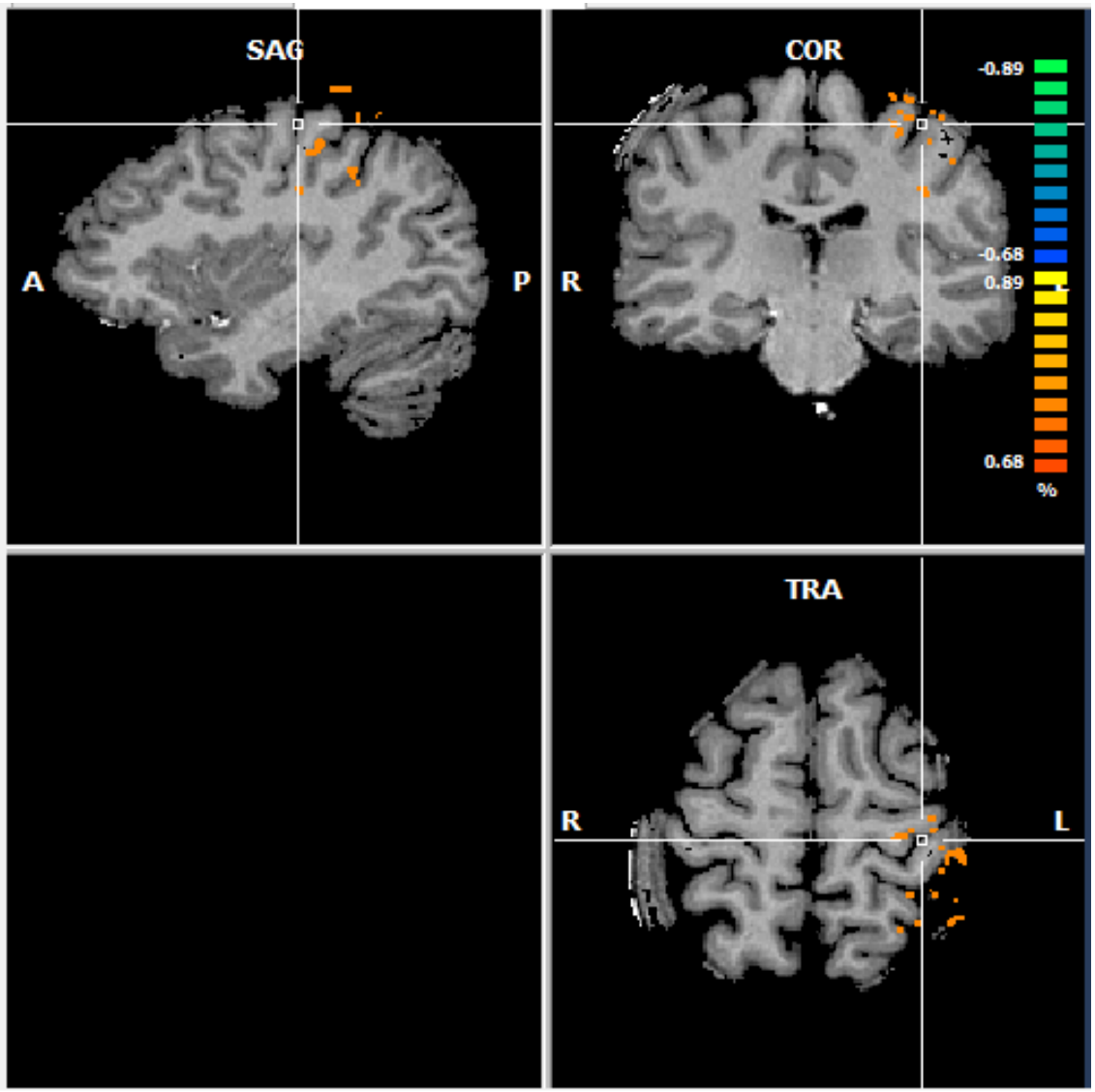
Soggetto GV - Miglior risultato

Il miglior risultato è stato ottenuto con

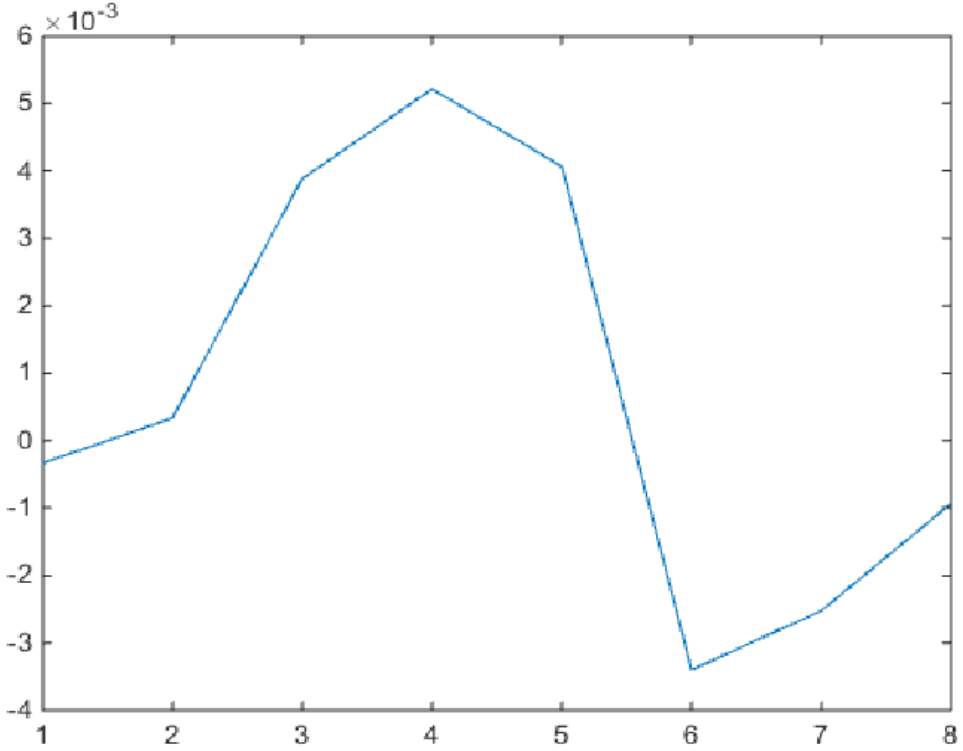
- **Compressione** serie temporale con GLM
- **Clustering** dei voxel con algoritmo Pam - Pearson
- **Numero di cluster** pari a 6
- **SVM** lineare
  - c = 0.1
- **Miglior feature** = centroide cluster 3
- **#voxel** = 407

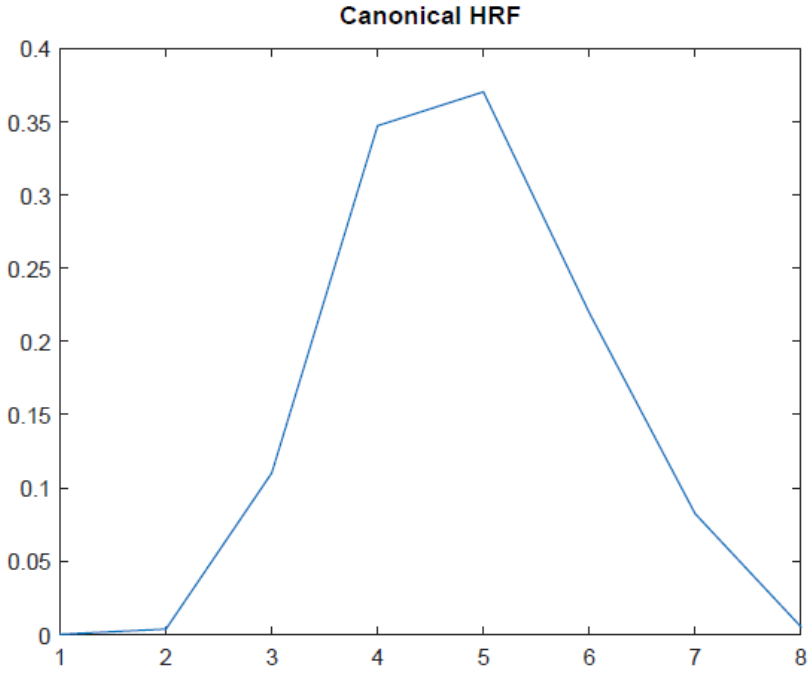


Soggetto GV - Mappe



Soggetto GV - Emodinamica





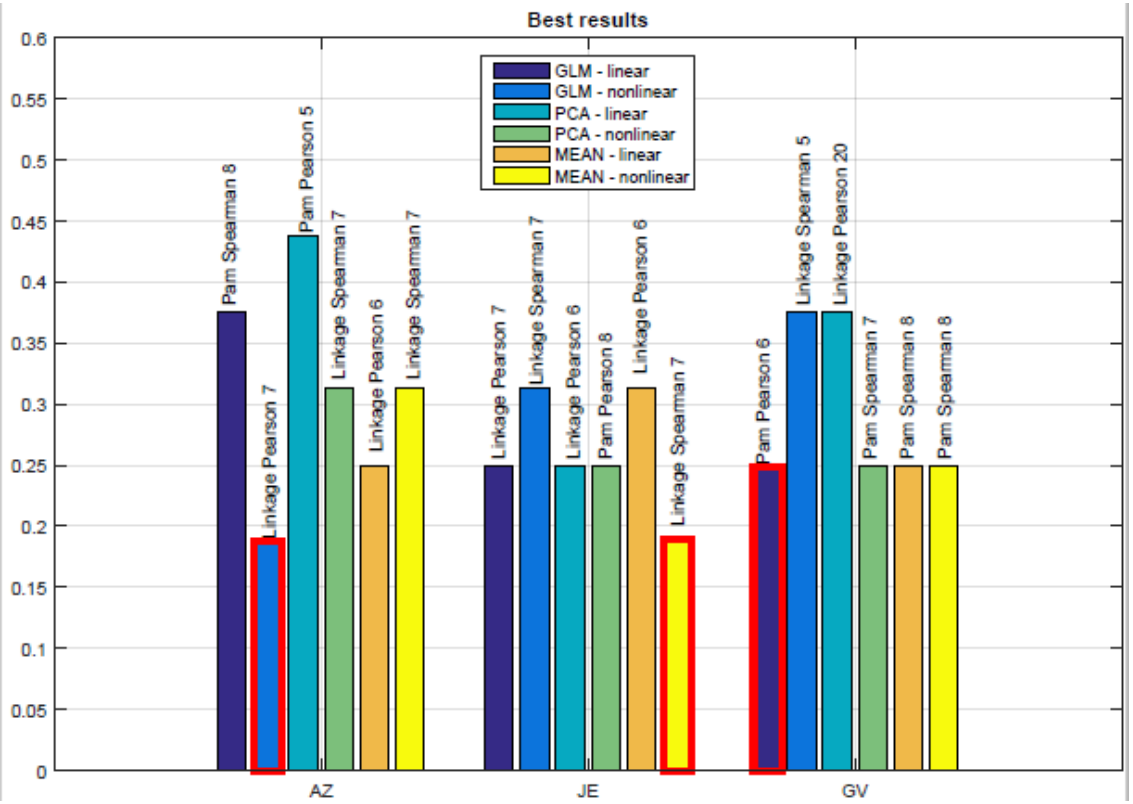
Outline

1. Introduzione
2. Metodologie
3. Risultati
   - Errori cross-validation
   - Errori one shot
4. **Conclusioni**

Conclusioni

Dal lavoro presentato è possibile concludere che:

- il **nuovo schema di cross-validation** si è rivelato molto efficace
  - applicato alla vecchia pipeline ha prodotto risultati migliori del 20% circa



Riferimenti I

- R. Henson, "Analysis of fmri timeseries." http://www.fil.ion.ucl.ac.uk/spm/doc/books/hbf2/pdfs/Ch10.pdf.
- "Nonlinear pca toolbox for matlab." http://www.nlpca.org/matlab.html.
- "Rbf svm parameters." http://scikit-learn.org/stable/auto_examples/svm/plot_rbf_parameters.html, 2014.

Grazie per l'attenzione