Encoding Brain Networks Through Geodesic Clustering of Functional Connectivity for Multiple Sclerosis Classification

Muhammad Abubakar Yamin, Paola Valsasina, Michael Dayan, Sebastiano Vascon, Jacopo Tessadori, Massimo Filippi, Vittorio Murino, Maria A. Rocca and Diego Sona.

Pattern Analysis and Computer Vision, Istituto Italiano di Tecnologia,Genova, Italy.
Department of Electrical, Electronics and Telecommunication Engineering and Naval Architecture, Universit`a degli Studi di Genova.

Department of Environmental Sciences, Informatics and Statistics, Ca' Foscari University, Venice, Italy

Neuroinformatics Laboratory, Fondazione Bruno Kessler, Trento, Italy

Department of Computer Science, Universit`a di Verona, Verona, Italy

Vita-Salute San Raffaele University, Milan, Italy

Outline

- Goal and Motivation
- Method
  - Datasets
  - The Pipeline
  - Background: Manifold Representation Of SPD Matrices
  - Geodesic Clustering
  - Feature Extraction And Classification
  - Experiments
- Results on Datasets
- Conclusions

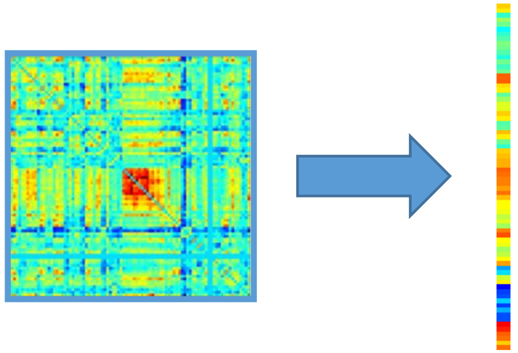
Goal and Motivation

Goal

- Encoding of brain functional connectivity (FC) data to discriminate between healthy controls (HC) and multiple-sclerosis (MS) patients

Motivation

- The FC analysis is based on graphs comparison, which is usually done by Euclidean distance (ED)
- Use of ED is sub-optimal because it does not capture the real geometry of manifold of symmetric positive definite (SPD) matrices
- A Better choice is to exploit the geometrical nature of SPD matrices on Riemannian manifold.

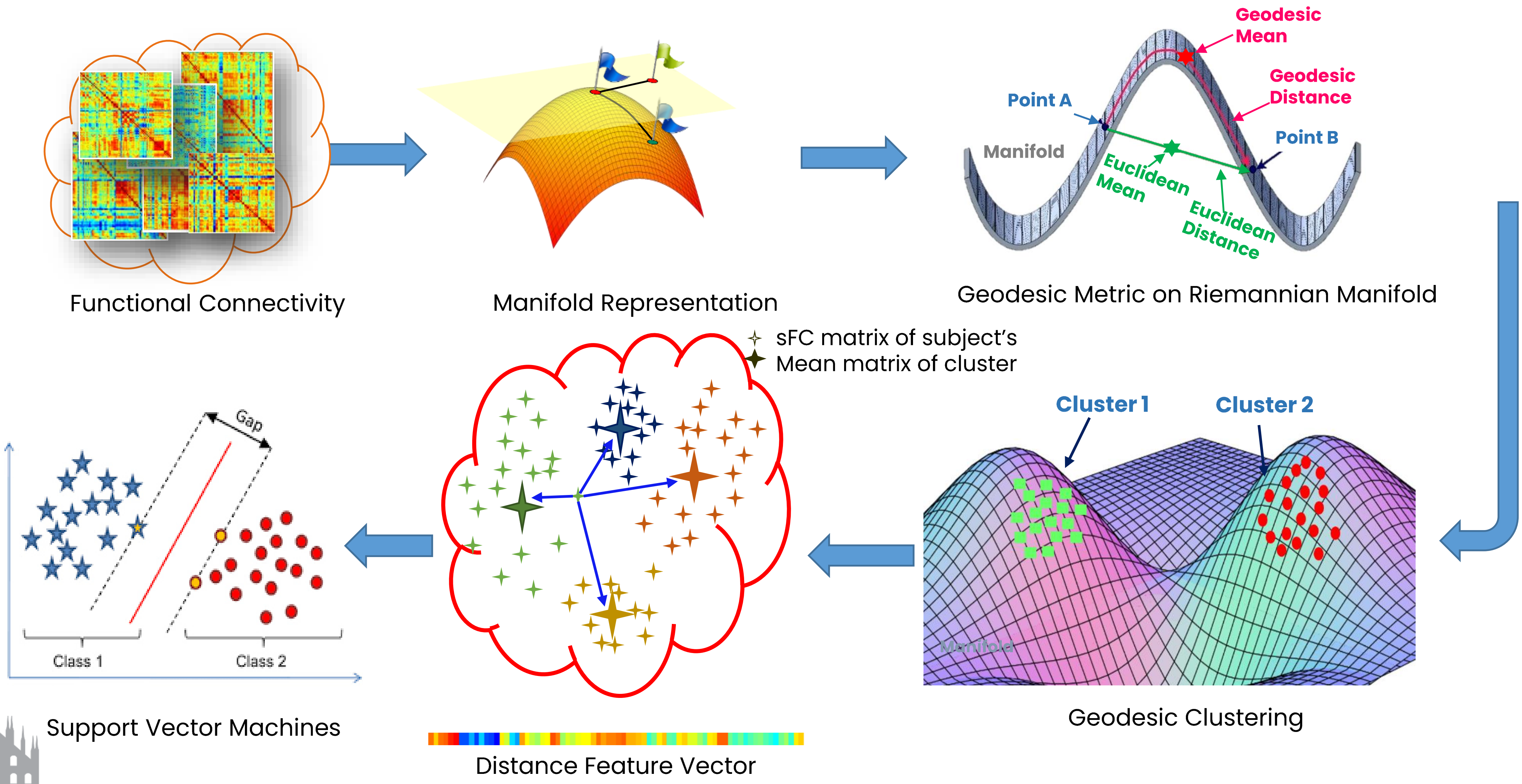


Method (Datasets)

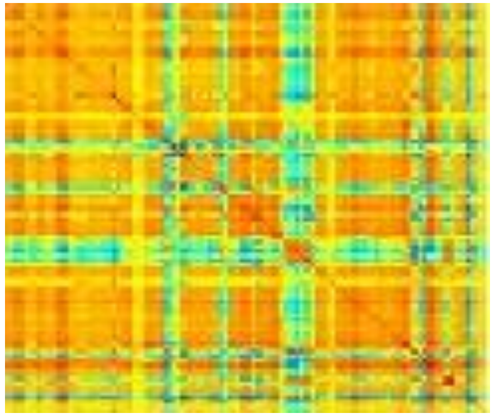
Dataset : Private dataset collected at the Neuroimaging Research Unit (Hospital San Raffaele, Milan, Italy)

- Resting state-functional magnetic resonance imaging (rs-fMRI)
- 33 HC and 72 multiple-sclerosis (MS) patients (age matched)
- 37 relapsing-remitting (RRMS) and 35 Progressive (PMS)
- FC matrix size is 90x90, based on AAL atlas computed using covariance.

The Pipeline



Method (Background: Manifold Representation Of SPD Matrices)

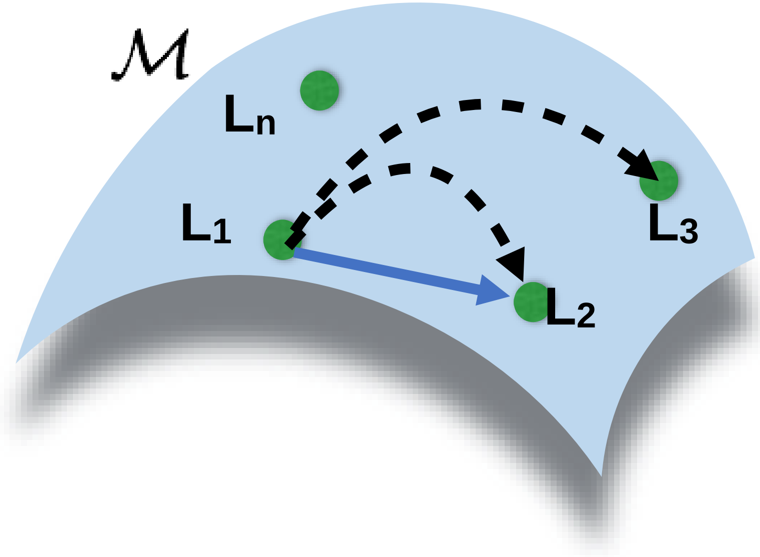


Covariance based connectivity matrices $\{\Sigma_i\}$

- Set of symmetric & positive semi-definite matrices.
- Represent the functional connectomes showing both positively and negatively synchronous connections
- Can be easily made SPD with a small regularization $\hat{\Sigma} = \Sigma + \lambda I$,

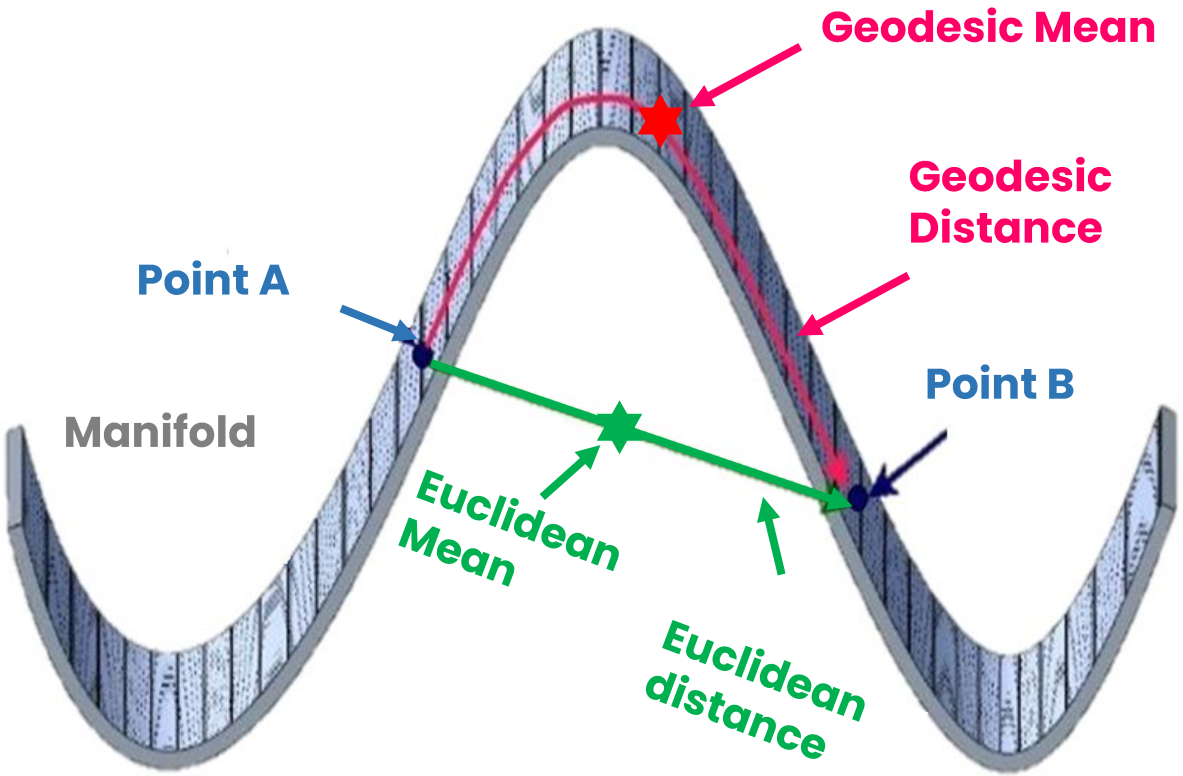
$$x^{\mathrm{T}}\Sigma x \geq 0 \quad \forall x \neq 0 \in \mathbb{R}$$

Symmetric Positive Definite (SPD) Riemannian Manifold



Method (Background: Geodesic Analysis on Manifold)

- Euclidean distance is sub-optimal because it does not capture the real geometry of manifold of SPD matrices
- Use of Geodesic distance is proposed which better define the distance along with manifold.
- Log-E distance eq.(1) [1] and geodesic mean in the closed form eq.(2) [2].



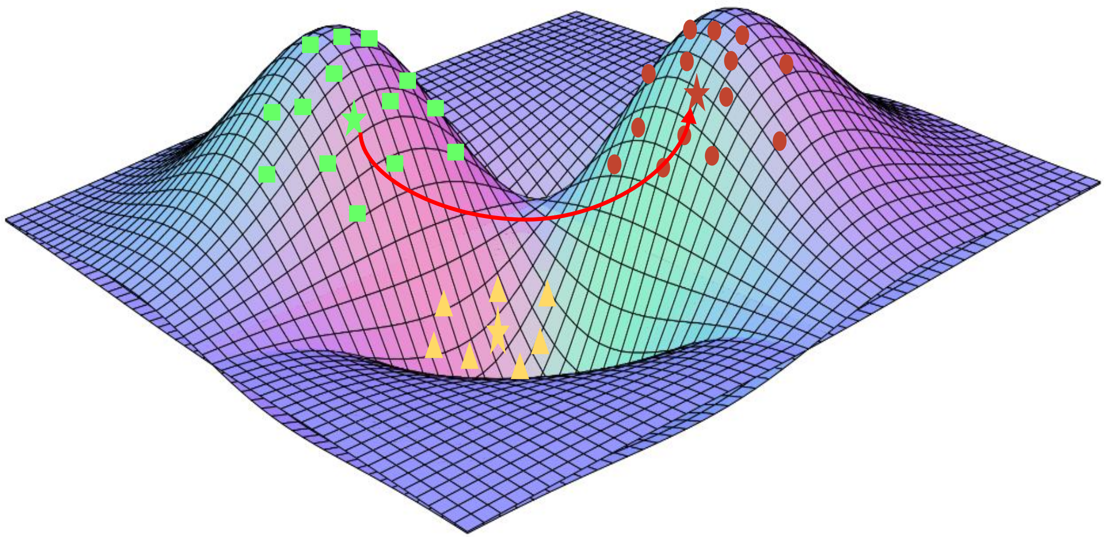
$$Log - Euclidean\ Distance: \quad d_L(\Sigma_i, \Sigma_j) = ||\log \Sigma_i - \log \Sigma_j|| \qquad (1)$$

$$Geodesic\ Mean: \quad \Sigma_L = \exp\{\arg\inf_{\Sigma} \sum_{i=1}^{n} ||\log \Sigma_i - \log \Sigma||^2\} = \exp\left\{\frac{1}{n}\sum_{i=1}^{n} \log \Sigma_i\right\} \qquad (2)$$

Method (Geodesic K-means Clustering of SPD Matrices)

- **Aim:**
  - Cluster FC matrices into homogeneous groups of subjects.

- **Underlying assumption:**
  - Alterations in brain connections grasped by the clusters.



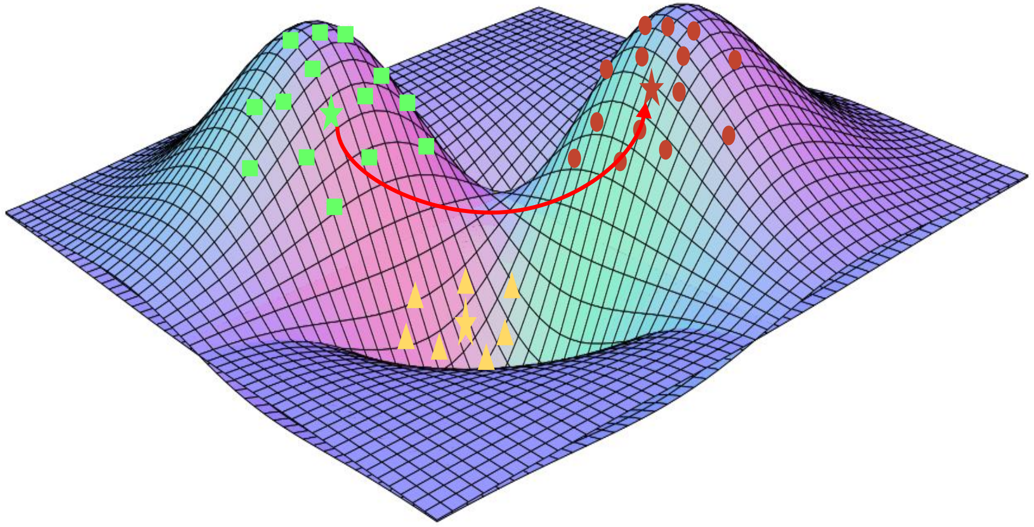
- K-means clustering was implemented using geodesic distance and geodesic mean

- **Drawbacks in traditional K-mean?**
  - Need to pick 'K' ,
  - Sensitive to initialization
  - Sensitive to outliers

Method (Geodesic Dominant Set Clustering of SPD Matrices)

AIM

Cluster FC matrices into homogeneous groups of subjects.



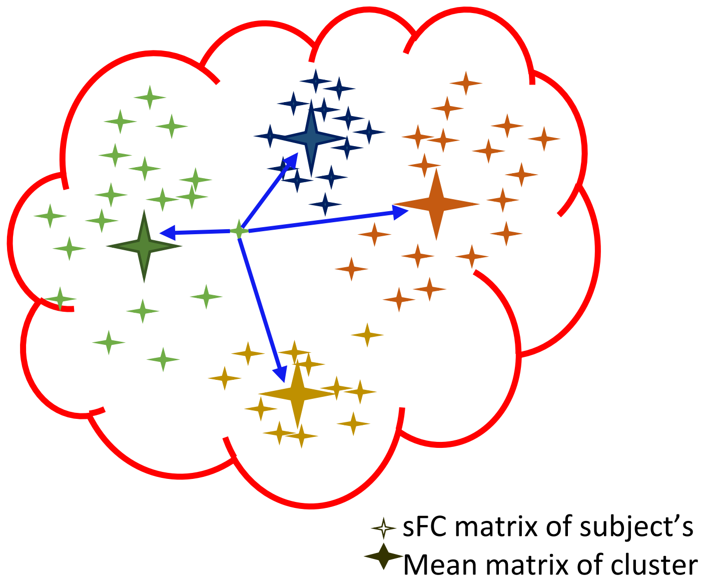
Dominant set clustering :

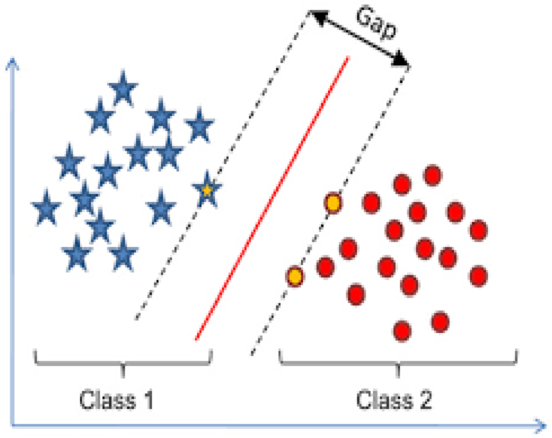
- Graph theoretic concept [3]
- Computes well separated and compact subset of nodes (dominant sets (DS))
- Extraction of DS is sequential (one by one)
- can be solved using game dynamic e.g. replicator dynamics [4]   maximize $\mathbf{x}^T A\mathbf{x}$
- Clusters are more similar inside and less similar to outside. $w_S(i) > 0$, for all $i \in S$ (internal homogeneity)

  $w_{S \cup \{i\}}(i) < 0$, for all $i \notin S$ (external homogeneity)
- Data in form of similarity matrix $S(i, j) = 1 - \frac{d_L(i, j)}{\max(d_L)}$
- No prior information on number of clusters (since we extract them sequentially) .
- Leaves clutter elements unassigned

Method (Feature Extraction and Classification)

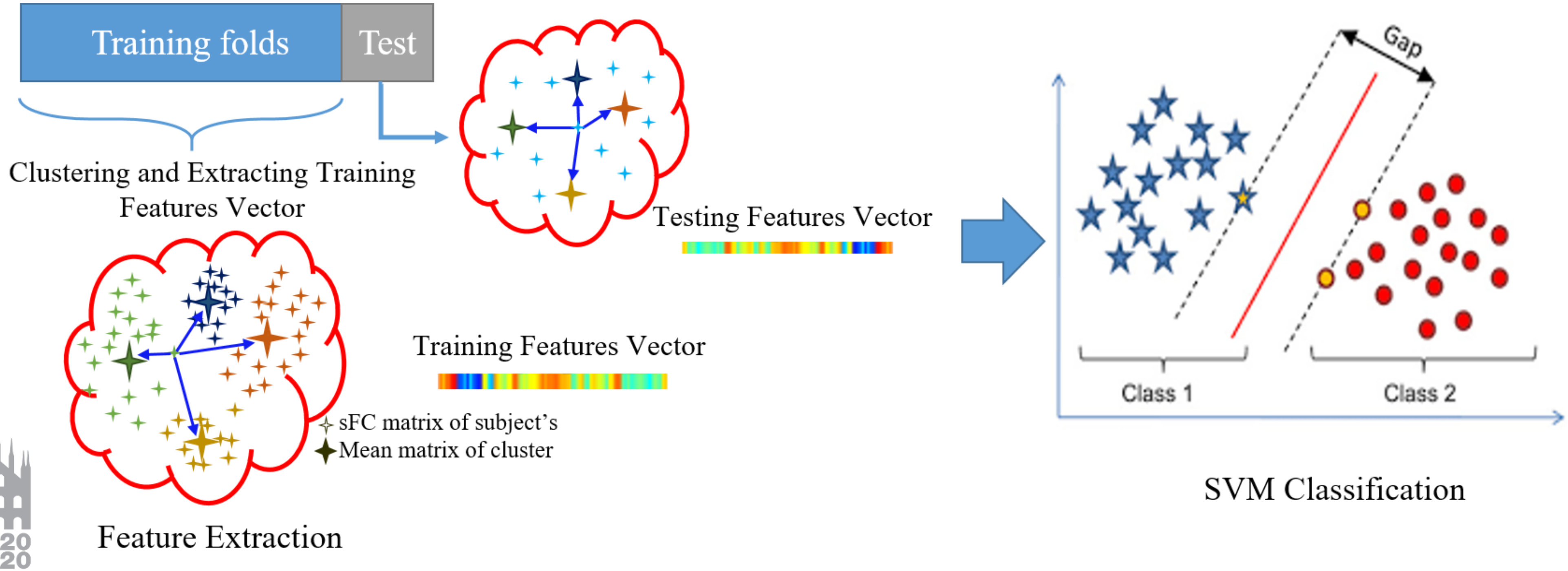
- Due to high dimensionality of FC matrices (90x90), feature encoding is needed.
- Using cluster centroid as dictionary
- Building vector representation by computing geodesic distance from each cluster centroid.
- Using this vector as feature set to classifier (Support Vector Machine, SVM).





Experiments

- For k-means Number of clusters were chosen to be variable between K=2-15.
- To avoid double dipping
  - 5-fold cross validation
  - Training folds for clustering & extracting training features
  - Test fold for computing test feature vector.
- Repeating 5-fold cross validation 100 time and taking mean of accuracies.
- Permutation test on labels (To check the significance of results).
- For Comparison, same analysis is performed using ED

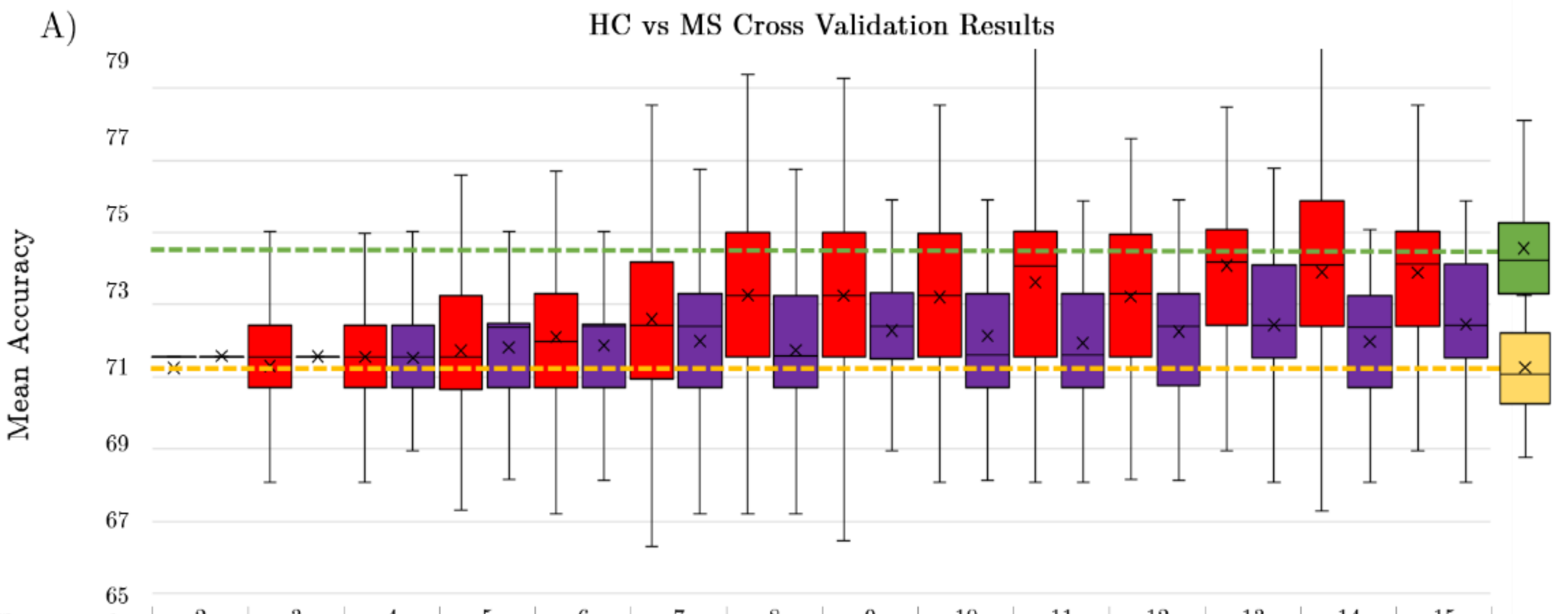


ICPR 2020

iit PAVIS

Results

- Average of 100 iterations of 5-folds cross validations of experiments:
- DS always extract 6 or 7 clusters, so for comparison we implement K-mean for K= 2-15
- HC vs MS: Accuracy 73.94 %



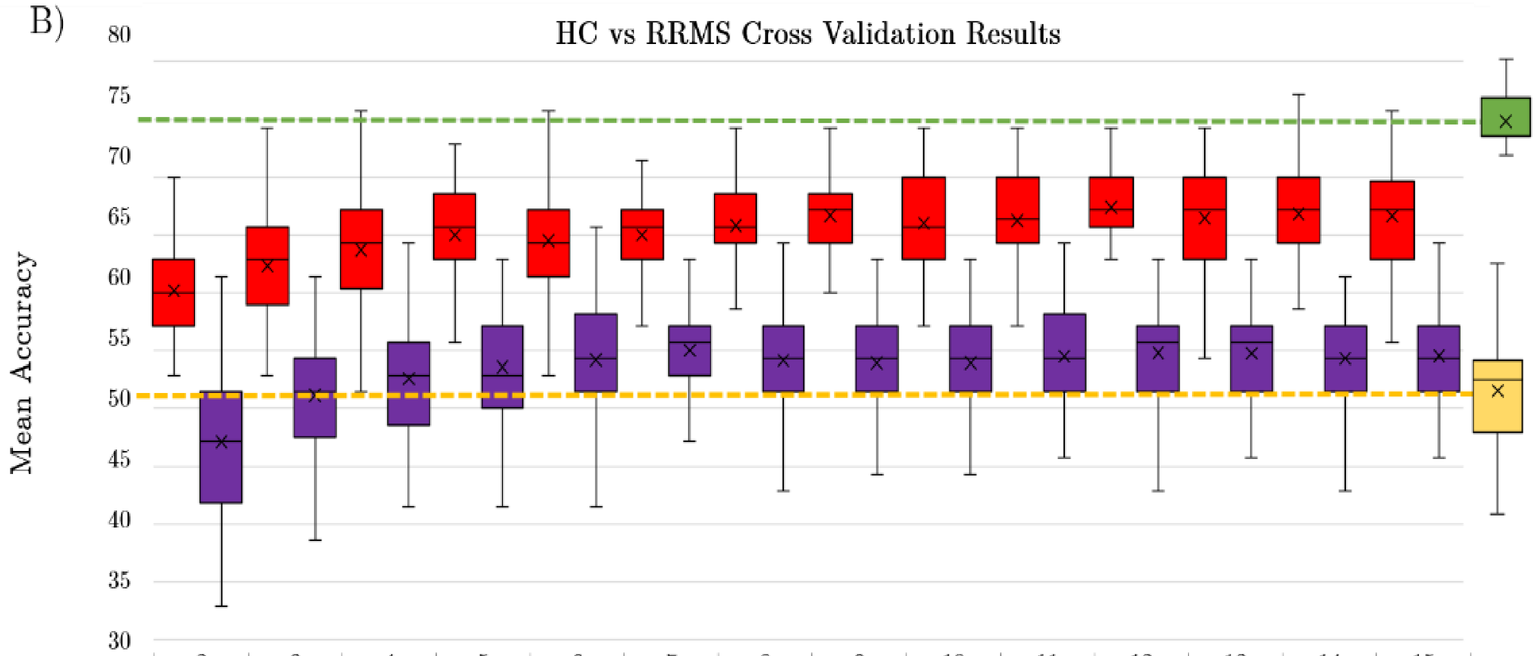
Cluster number	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Geod K-mean	71.24	71.30	71.55	71.72	72.11	72.60	73.26	73.25	73.21	73.62	73.22	73.80	73.63	73.62
Eucl K-mean	71.58	71.57	71.53	71.82	71.86	71.99	71.73	72.28	72.13	71.93	72.26	72.44	72.00	72.45

TABLE I
AVERAGE CONFUSION MATRIX OF CLASSIFICATION RESULTS FOR THE PROPOSED MEAN GEODESIC DS CLUSTERING APPROACH AND BEST OF GEODESIC K-MEANS CLUSTERING FOR HC VS. MS.

Geodesic Dominant-Set				Geodesic k-means		
		Predicted Class			Predicted Class	
		HC	MS		HC	MS
Actual Class	HC	13.98	19.02	HC	14.5	18.5
	MS	8.47	63.53	MS	10.5	61.5

Results

- Average of 100 iterations of 5-folds cross validations of experiments:
- DS always extract 6 or 7 clusters, so for comparison we implement K-mean for K= 2-15
- HC vs MS: Accuracy 73.94 %
- HC vs RRMS: Accuracy 72.51%



Cluster number	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Geod K-mean	60.17	62.34	63.74	65.01	64.51	65.00	65.83	66.69	66.03	66.27	67.39	66.49	66.84	66.67
Eucl K-mean	47.11	51.11	52.59	53.59	54.19	55.04	54.13	53.94	53.91	54.50	54.81	54.77	54.42	54.53

TABLE II

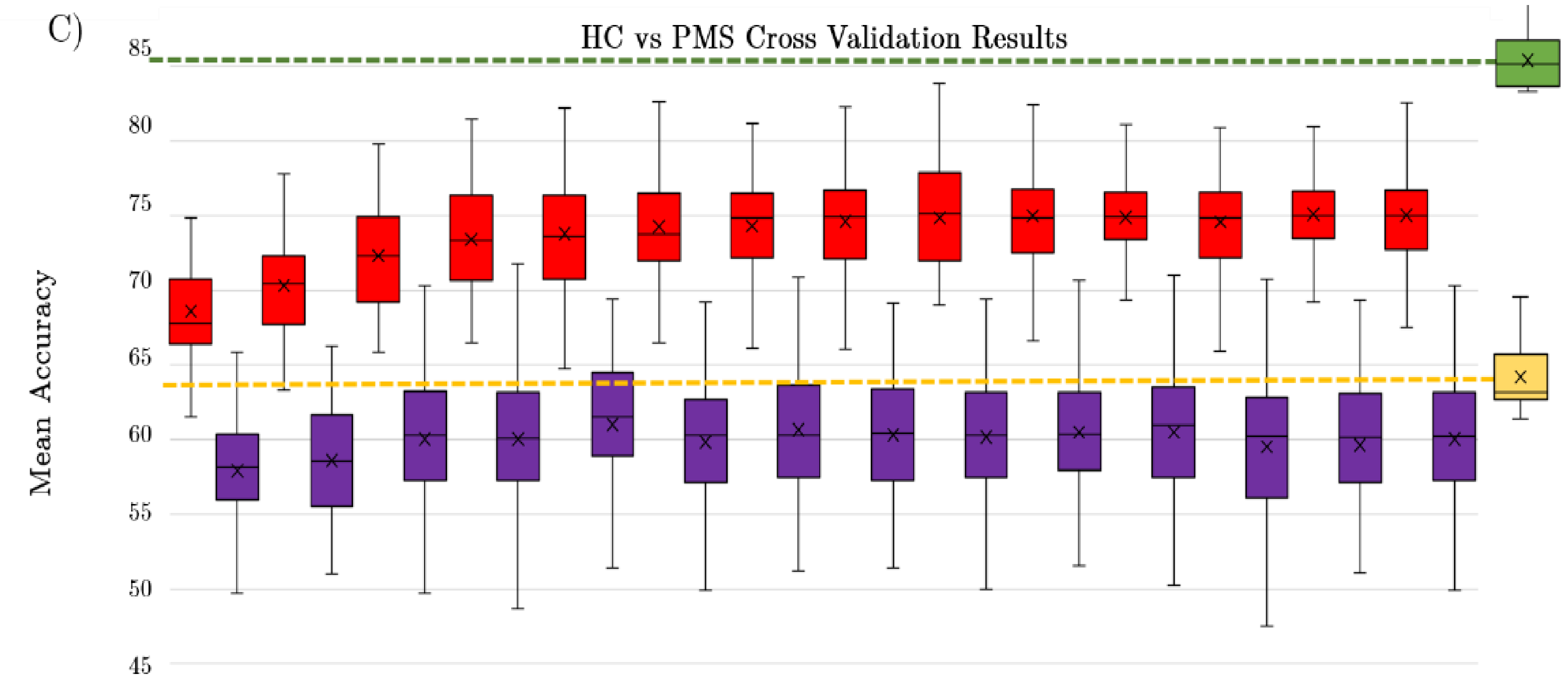
AVERAGE CONFUSION MATRIX OF CLASSIFICATION RESULTS FOR THE PROPOSED MEAN GEODESIC DS CLUSTERING APPROACH AND BEST OF GEODESIC K-MEANS CLUSTERING FOR HC VS. RRMS.

Geodesic Dominant-Set				**Geodesic k-means**		
		Predicted Class			**Predicted Class**	
		HC	RRMS		HC	RRMS
Actual Class	HC	21.68	11.32	HC	19.28	13.72
	RRMS	7.92	29.08	RRMS	10.11	26.89

Results

- Average of 100 iterations of 5-folds cross validations of experiments:
- DS always extract 6 or 7 clusters, so for comparison we implement K-mean for K= 2-15
- HC vs MS: Accuracy 73.94 %
- HC vs RRMS: Accuracy 72.51%
- HC vs PMS: Accuracy 84.06 %
- Geodesic Clustering gives superior results and also Geodesic Dominant-Set clustering is always better in performance as compared to K-Mean clustering.

C)



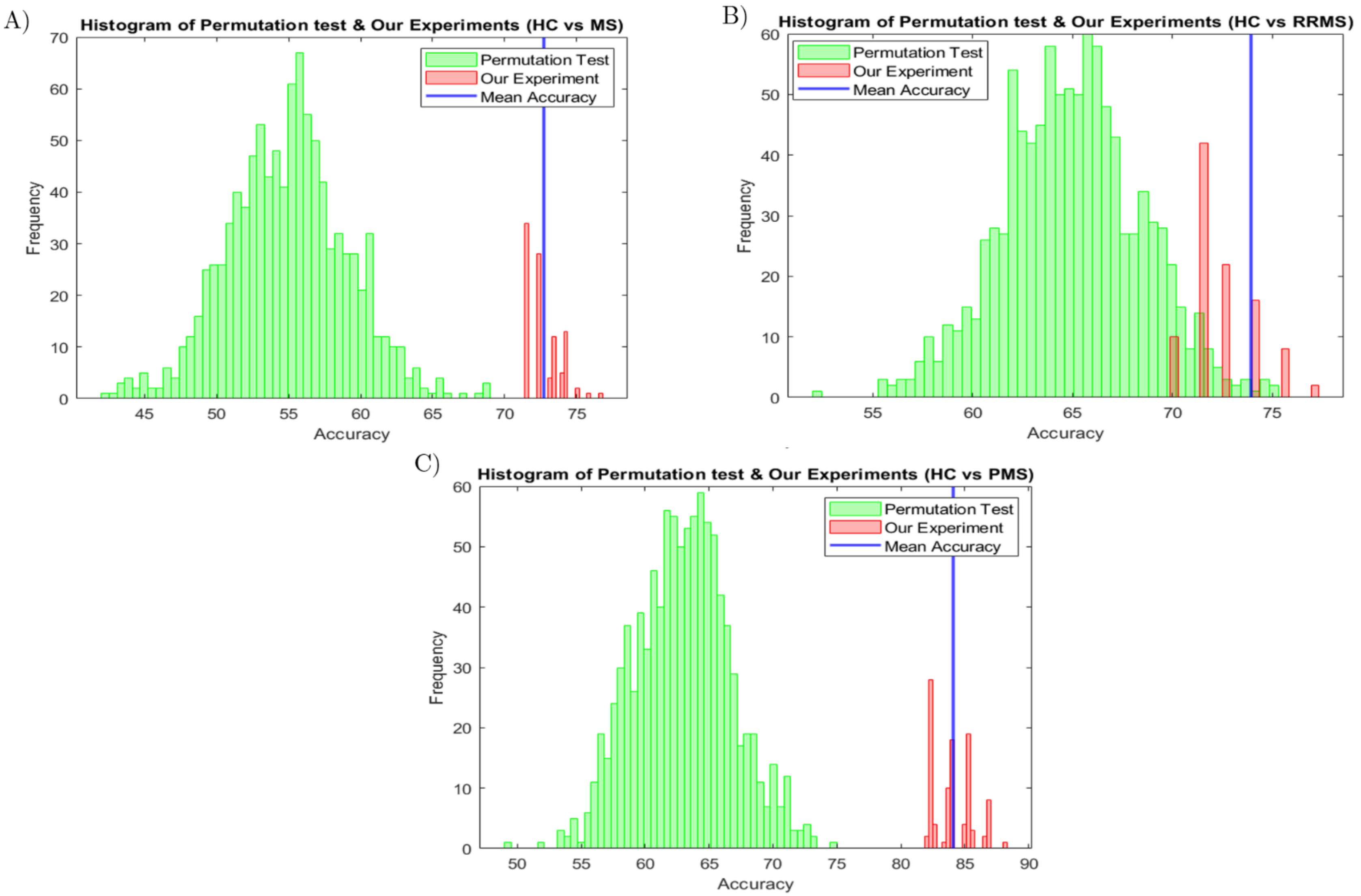
Cluster number	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Geod K-mean	68.60	70.33	72.31	73.41	73.79	74.26	74.30	74.62	74.87	75.00	74.91	74.60	75.09	75.01
Eucl K-mean	57.93	58.61	60.05	60.03	61.01	59.85	60.67	60.31	60.20	60.51	60.50	59.53	59.57	60.05

TABLE III
AVERAGE CONFUSION MATRIX OF CLASSIFICATION RESULTS FOR THE PROPOSED MEAN GEODESIC DS CLUSTERING APPROACH AND BEST OF GEODESIC K-MEANS CLUSTERING FOR HC VS. PMS

Geodesic Dominant-Set				Geodesic k-means		
		Predicted Class			Predicted Class	
		HC	PMS		HC	PMS
Actual Class	HC	26.08	6.92	HC	25.16	7.84
	PMS	3.93	31.07	PMS	9.12	25.88

Results

- Average of 100 iterations of 5-folds cross validations of experiments:
- DS always extract 6 or 7 clusters, so for comparison we implement K-mean for K= 2-15
- HC vs MS: Accuracy 73.94 %
- HC vs RRMS: Accuracy 72.51%
- HC vs PMS: Accuracy 84.06 %
- Geodesic Clustering gives superior results and also Geodesic Dominant-Set clustering is always better in performance as compared to K-Mean clustering.
- Significance of permutation test
    - HC vs MS, P_value <0.0005
    - HC vs RRMS, P_value <0.05
    - HC vs PMS, P_value <0.0005



ICPR 2020

iit PAVIS

Conclusion

- **Neuroscientific**
  - Alteration in brain is helpful in discriminating between HC and patients affected with different phenotype of MS.

- **Computational**
  - Proper data representation allow an effective exploitation on the manifold space.
  - Geodesic method-based clustering gives superior results.
  - Specific encoding of FC matrices leads to good performance in discriminating task.

Reference

1. L. Dodero, H. Q. Minh, M. S. Biagio, V. Murino and D. Sona. “Kernelbased classification for brain connectivity graphs on the Riemannian manifold of positive definite matrices”. ISBI 2015. 16-19 April
2. Dryden, i.l., Koloydenko, a. and Zhou, d., 2009. Non-euclidean statistics for covariance matrices with applications to diffusion tensor imaging. Annals of applied statistics, 3 (3), pp. 1102 - 1123.
3. M. Pavan and M. Pelillo, ”Dominant Sets and Pairwise Clustering,” in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 29, no. 1, pp. 167-172, Jan. 2007.doi: 10.1109/TPAMI.2007.250608
4. J. W. Weibull, Evolutionary game theory. MIT press, 1997

Thank you!

ICPR 2020

25th INTERNATIONAL CONFERENCE ON PATTERN RECOGNITION

Milan, Italy 10 | 15 January 2021