

**DI TECNOLOGIA** 

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



 $d_L(i,j)$ 

 $\max(d_L)$ 

S(i,j) = 1 -

Muhammad Abubakar Yamin<sup>1,2</sup>, Paola Valsasina<sup>9</sup>, Michael Dayan<sup>3</sup>, Sebastiano Vascon<sup>4,5</sup>,

Jacopo Tessadori<sup>1</sup>, Massimo Filippi<sup>9,10,11,12</sup>, Vittorio Murino<sup>1,6,7</sup>, Maria A. Rocca<sup>9,10,12</sup>, and Diego Sona<sup>1,8</sup>

<sup>1</sup>Pattern Analysis and Computer Vision, Istituto Italiano di Tecnologia, Genova, Italy <sup>2</sup>Dipartimento di Ingegneria Navale, Elettronica e delle Telecomunicazioni, University of Genova, Italy. <sup>3</sup>Human Neuroscience Platform, Fondation Campus Biotech Geneva, Geneva, Switzerland. <sup>4</sup>Department of Environmental Sciences, Informatics and Statistics, Ca' Foscari University, Venice, Italy. <sup>5</sup>European Center for Living Technology, Ca' Foscari University, Venice, Italy. <sup>6</sup>Dipartimento di Informatica, University of Verona, Italy. <sup>7</sup>Ireland Research Center, Huawei Technologies Co. Ltd., Dublin, Ireland. <sup>8</sup>Neuroinformatics Laboratory, Fondazione Bruno Kessler, Trento, Italy. <sup>9</sup>Neuroimaging Research Unit, Institute of Experimental Neurology, Division of Neuroscience, IRCCS San Raffaele Scientific Institute, Milan, Italy. <sup>10</sup>Neuroimaging of CNS WM Unit, Institute of Experimental Neurology, Division of Neuroscience; Neurology Unit, IRCCS San Raffaele Scientific Institute, Milan, Italy. <sup>11</sup>Neurophysiology Unit, IRCCS San Raffaele Scientific Institute, Milan, Italy. <sup>12</sup>Vita Salute San Raffaele University, Milan, Italy.

## Outline

- Aim: Encoding of brain functional connectivity (FC) data to discriminate between healthy controls (HC) and multiple-sclerosis (MS) patients
- Use of Euclidean distance (ED) and Euclidean Mean (EM) is suboptimal because it does not capture the real geometry of Riemannian Manifold (RM) of symmetric positive definite (SPD) matrices
- Covariance-based static FC matrices were clustered using geodesic distance (GD) and geodesic mean (GM) [1,2].
- Clusters were used to encode the data for the classification task using geodesic methods on manifold [3]
- Tested on a private dataset of 33 HC and 72 MS-patients (37) Relapsing-Remitting (RRMS), and 35 Progressive (PMS)).
- Neuroscientific result: Alteration in FC is helpful in discriminating HC and patients affected with different phenotype of MS.
- Computational result: Using the proper data representation allow an effective exploitation of metrics defined on the manifold



### **Riemannian Manifold of SPD matrices and Geodesic Clustering**

- Covariance matrices are always symmetric positive semi-definite, and a small regularization makes them SPD forming a RM.
- Euclidean distance between matrices is suboptimal and it is better to consider a metric defined on the Riemannian Manifold.
- We adopt the Log-Euclidean distance, which is simple, and fast to compute[4], formula to compute GD between two matrices  $\Sigma_i \& \Sigma_i$  is  $d_L(\Sigma_i, \Sigma_j) = ||\log(\Sigma_i) - \log \Sigma_j)||$

**Geodesic Clustering** 

**Aim:** Alterations in brain connections can be grasped by clusters, so clustering FC matrices into homogenous groups of subjects

- 1. k-means clustering was implemented using geodesic distance and geodesic mean.
- 2. Dominant-Set, a graph theoretic clustering algorithm, which sequentially computes well separated and compact subset of nodes called dominant-sets (DS)
- Geodesic mean of multiple covariance matrices can be computed in a closed form[5]:

$$\Sigma_L = \exp\left\{\arg \inf_{\Sigma} \sum_{i=1}^{\infty} ||\log \Sigma_i - \log \Sigma||^2\right\} = \exp\left\{\frac{1}{n} \sum_{i=1}^{\infty} \log \Sigma_i\right\}$$

#### **Data Encoding and Classification Experiments**

- To avoid double dipping we are using 5-fold cross validation

helpful in ÎS

- $\mathbf{x}^T A \mathbf{x}$ Solved using game dynamic e.g., replicator dynamics [6] maximize
- Data is represented in the form of similarity matrix
- No prior information on number of clusters required
- Leaves clutter elements unassigned



#### **Performance of Linear-SVM Classifier**

25<sup>th</sup> International Conference on Pattern Recognition (ICPR 2020). Milan, Italy 10-15 January 2021.