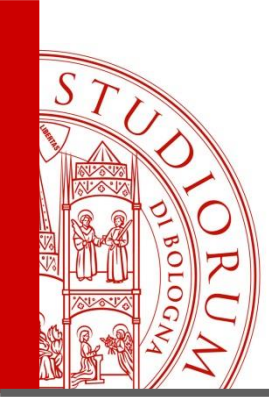


CCS/ITALY

ITALIAN REGIONAL CONFERENCE ON COMPLEX SYSTEMS

Classification of Genome Wide Association data by Belief Propagation Neural network

Daniele Dall'Olio[1], *Nico Curti*[1,2], *Armando Bazzani*[1,2], *Daniel Remondini*[1,2] , *Gastone Castellani*[1,2]

[1]Department of Physics and Astronomy, University of Bologna
[2]INFN Bologna

Presented by:
Daniele Dall'Olio

Trento, Italy
July 1-3, 2019

Overview

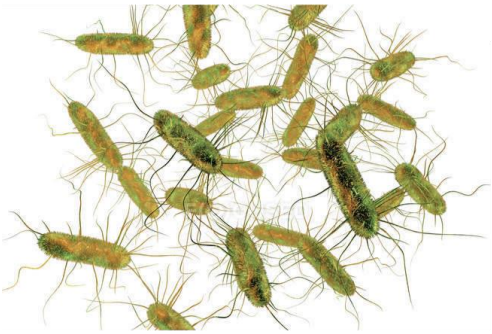
European Commission | Horizon 2020 European Union funding for Research & Innovation

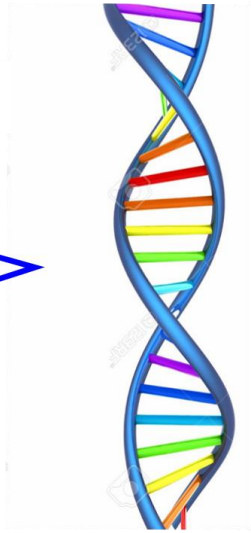
Objective: *Identification, containment and mitigation of emerging infectious diseases and foodborne outbreaks.*

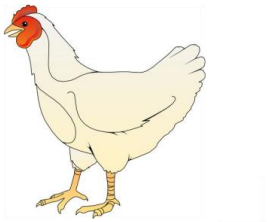
Source Attribution

Genome Wide Association

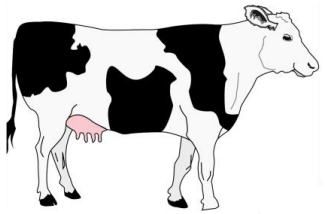
Bacteria



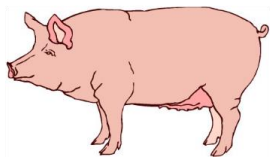




?



?

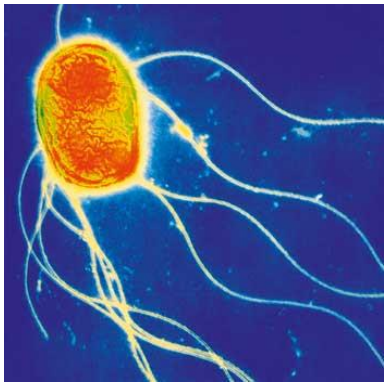


?

Data

European Commission
Horizon 2020
European Union funding
for Research & Innovation

Salmonella Enterica



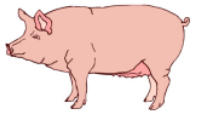
SNPs

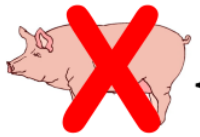
Reference : CCGTTAGAGTTACAATTCGA

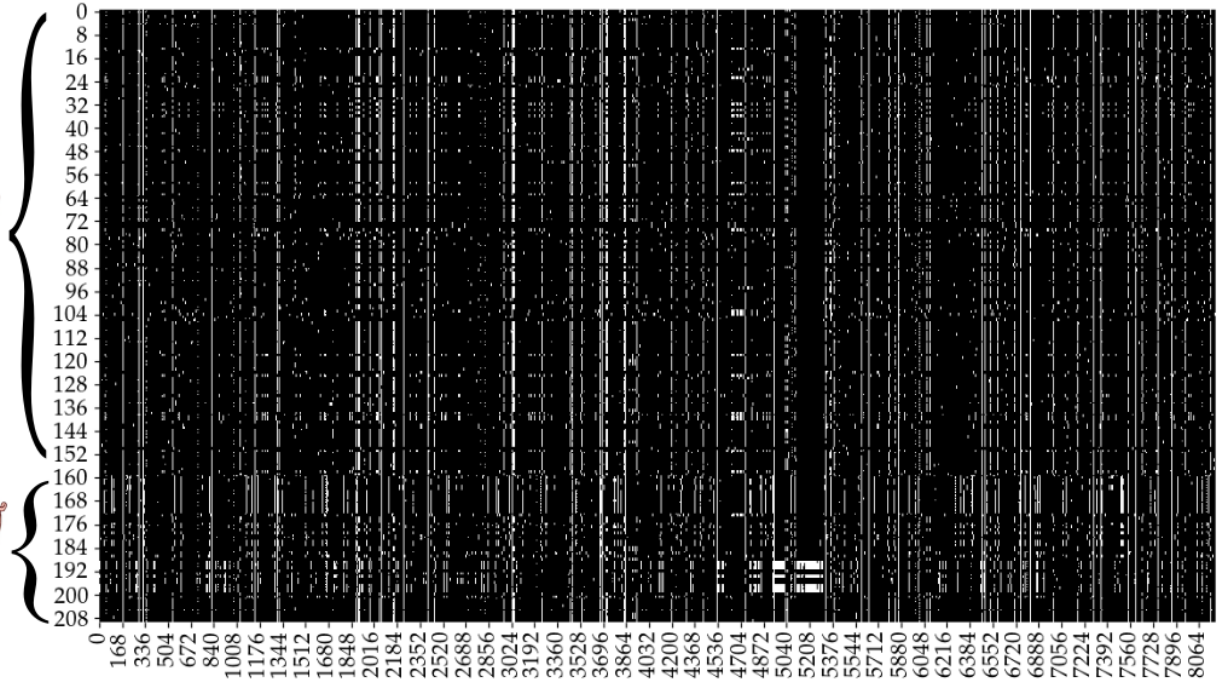
Sample : CGGTTAGAGTAACTATTCCA

Binary SNPs : 01000000001001000010

Samples	**210**
Pigs	**159**
no-Pigs	**51**
Filtered Bases	**8189**

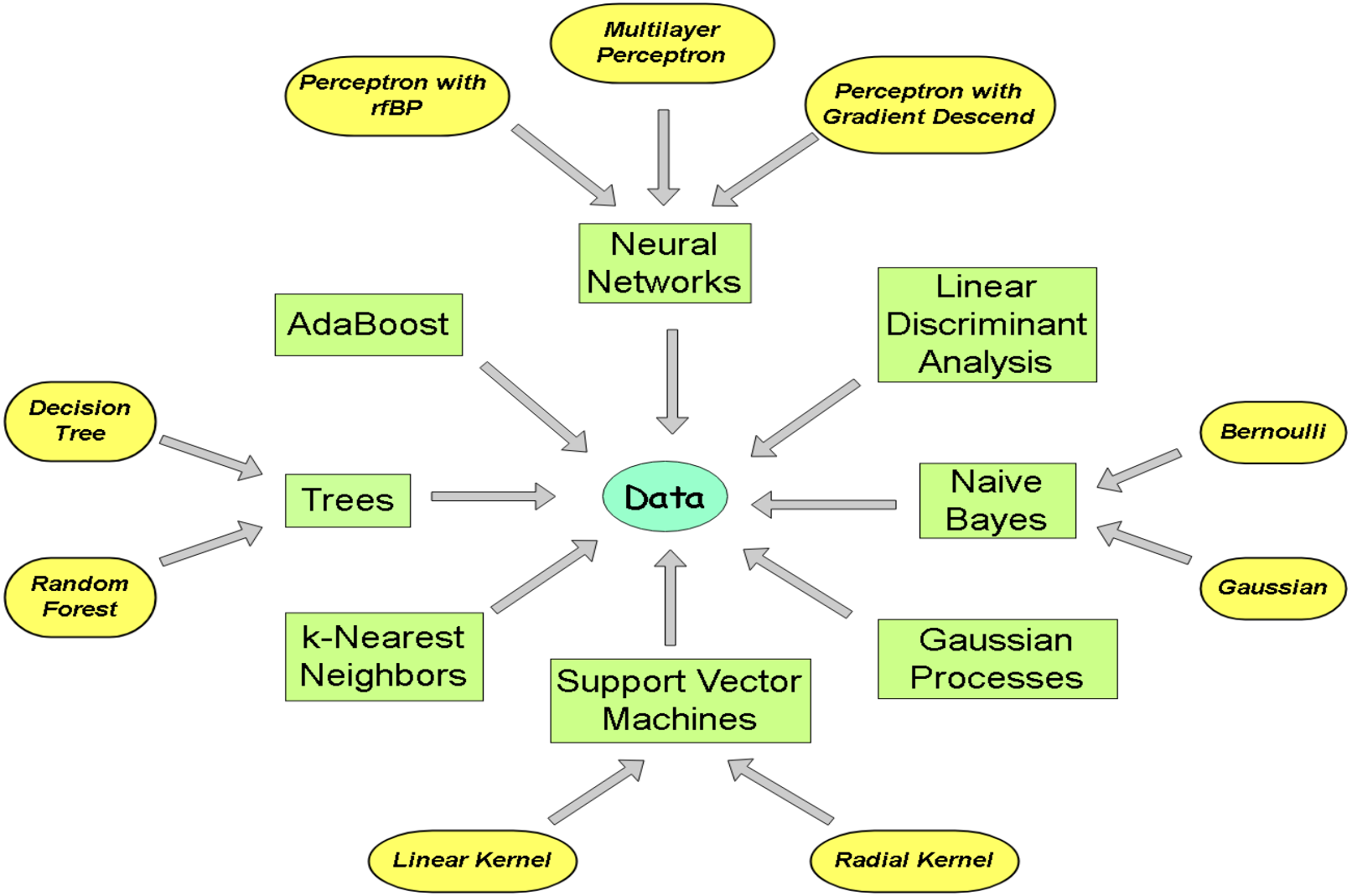






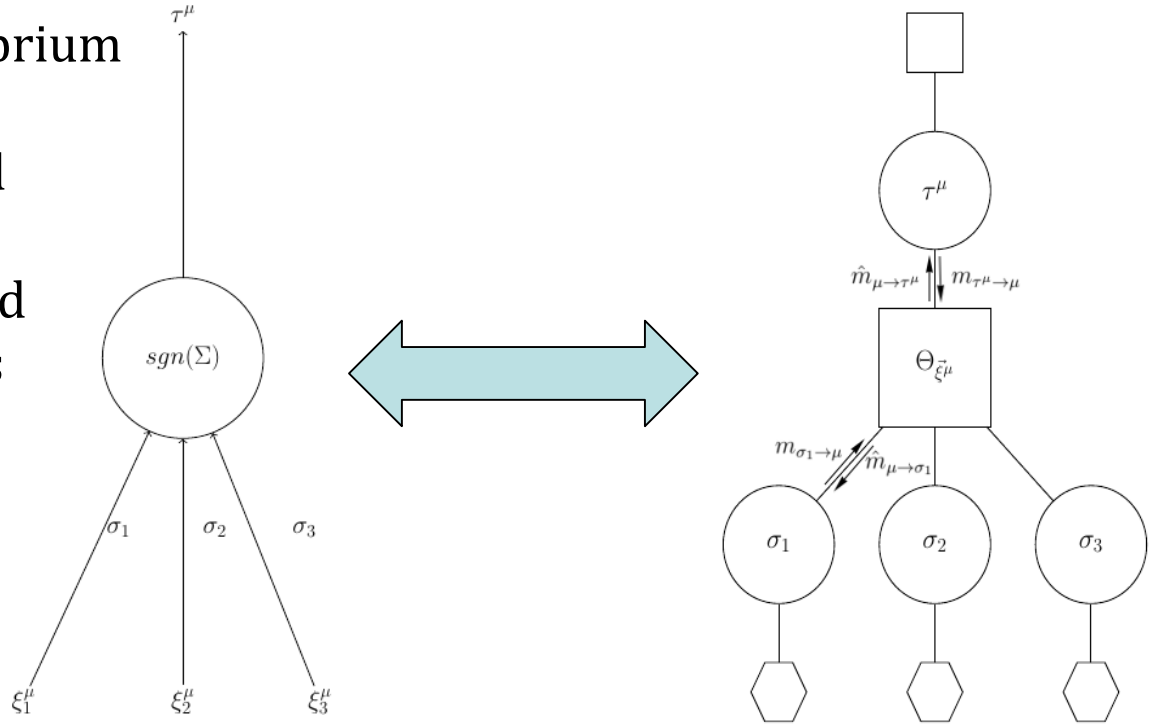
Filtered Bases

Classifiers



replicated focusing Belief Propagation[1]

- Derived from a out-of-equilibrium distribution
- Entropy-maximization based learning rule
- Parameterized and reinforced Belief Propagation equations



Binary Perceptron ⟷ Perceptron graph

New implementation:

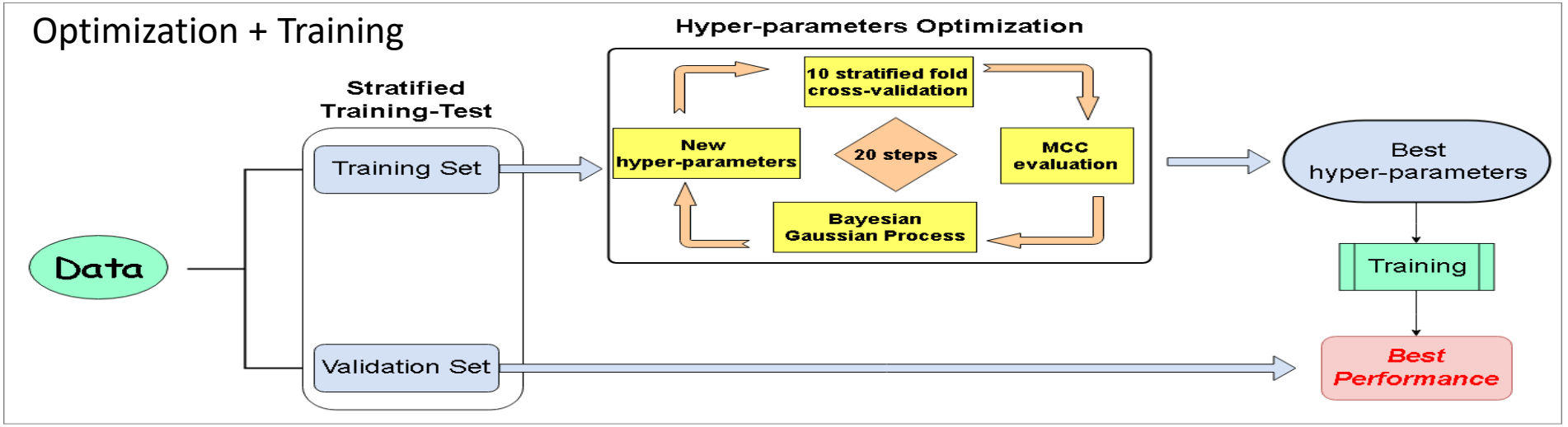
- **C++ library**
- **Python wrapper**
- **Optimized for parallel computing**
- **Integrated with *scorer* library**

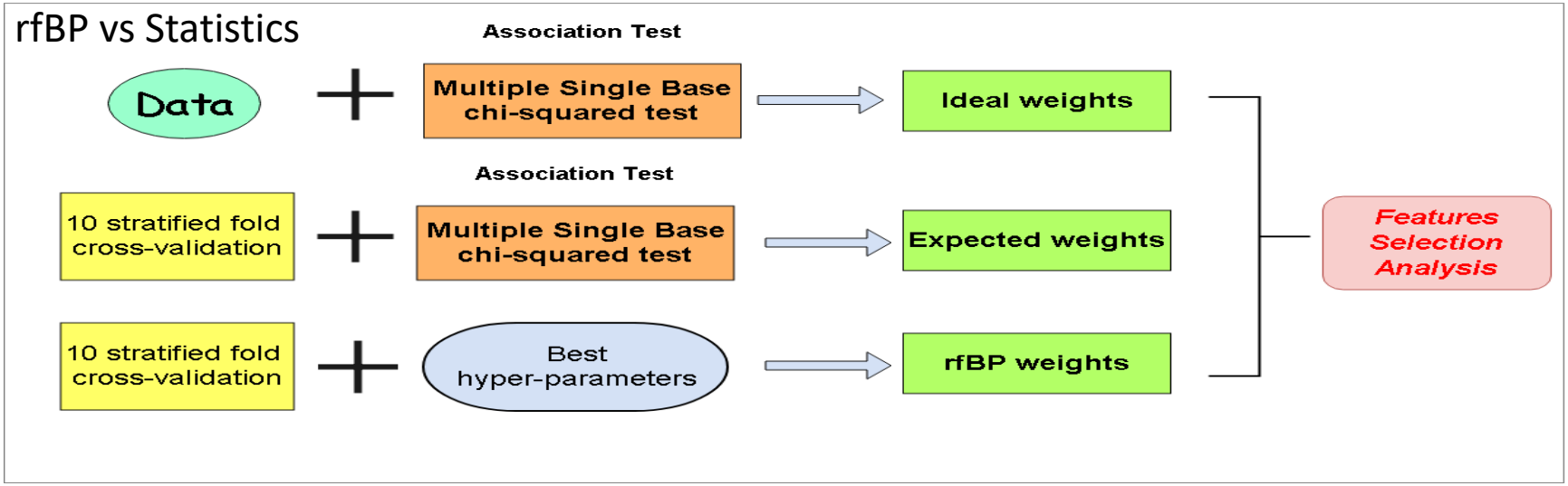
- **Link:** https://github.com/Nico-Curti/rFBP

[1]C. Baldassi et al. ***Unreasonable Effectiveness of Learning Neural Networks: From Accessible States and Robust Ensembles to Basic Algorithmic Schemes, 2016.***

Workflow

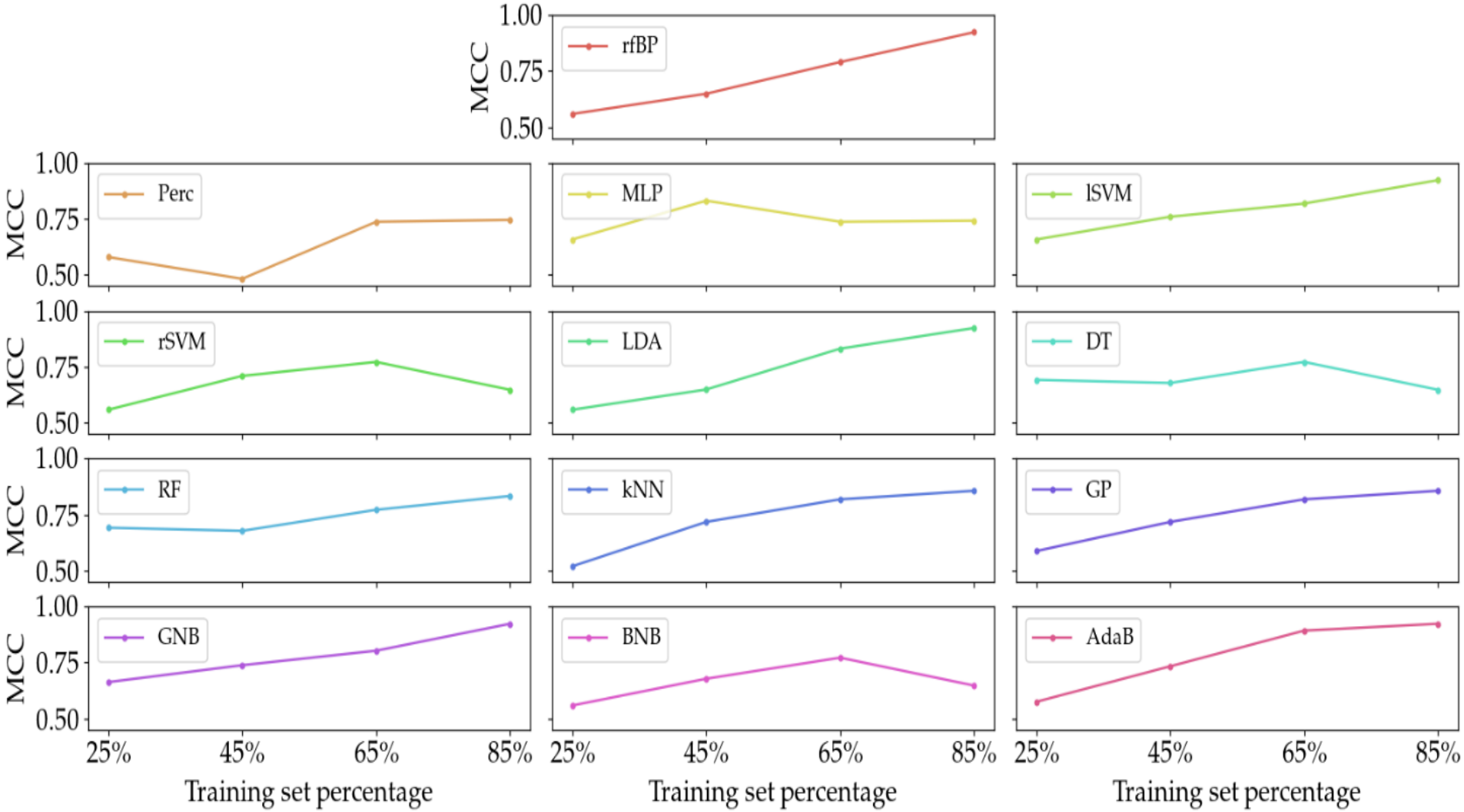
European Commission
Horizon 2020
European Union funding
for Research & Innovation



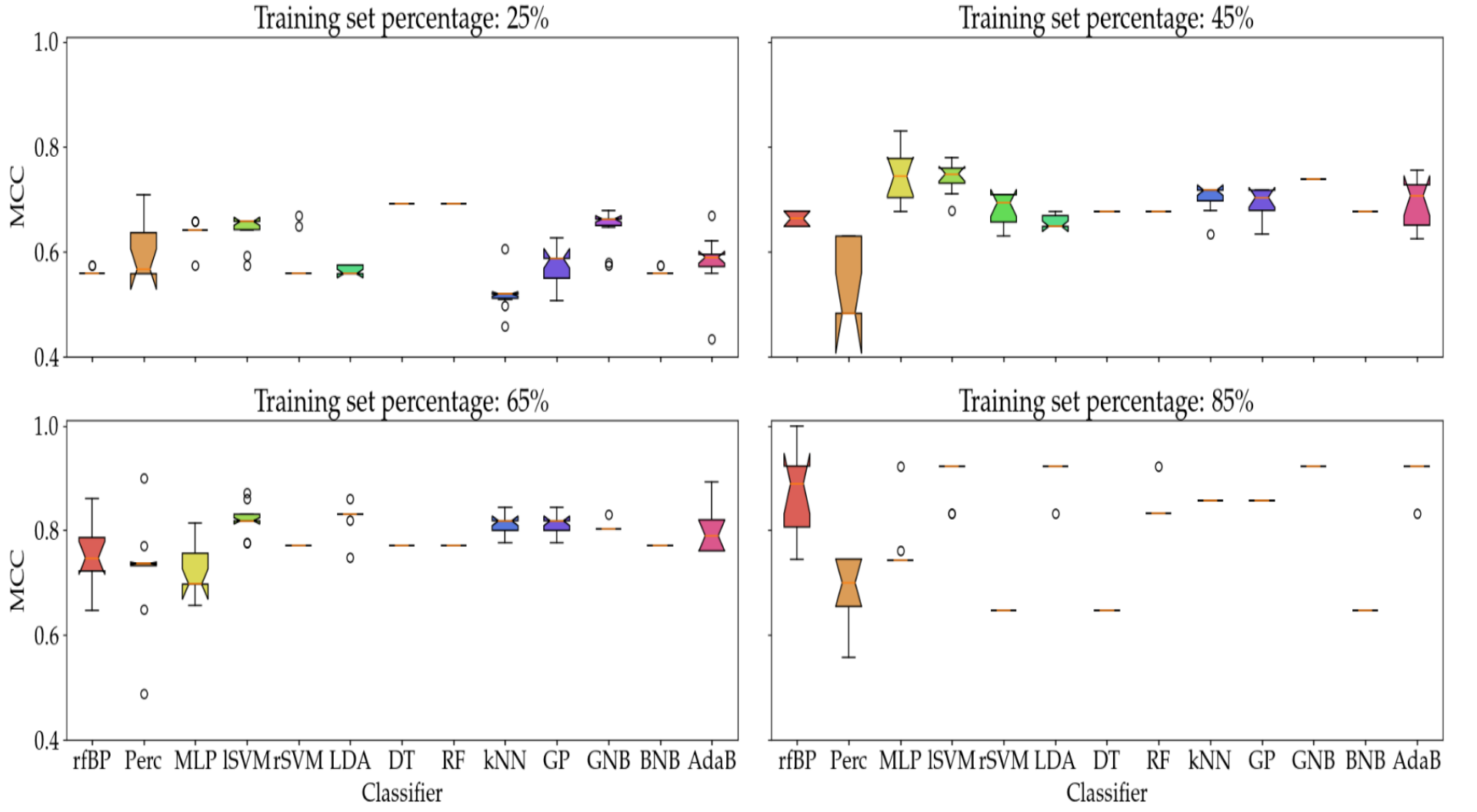


Performances

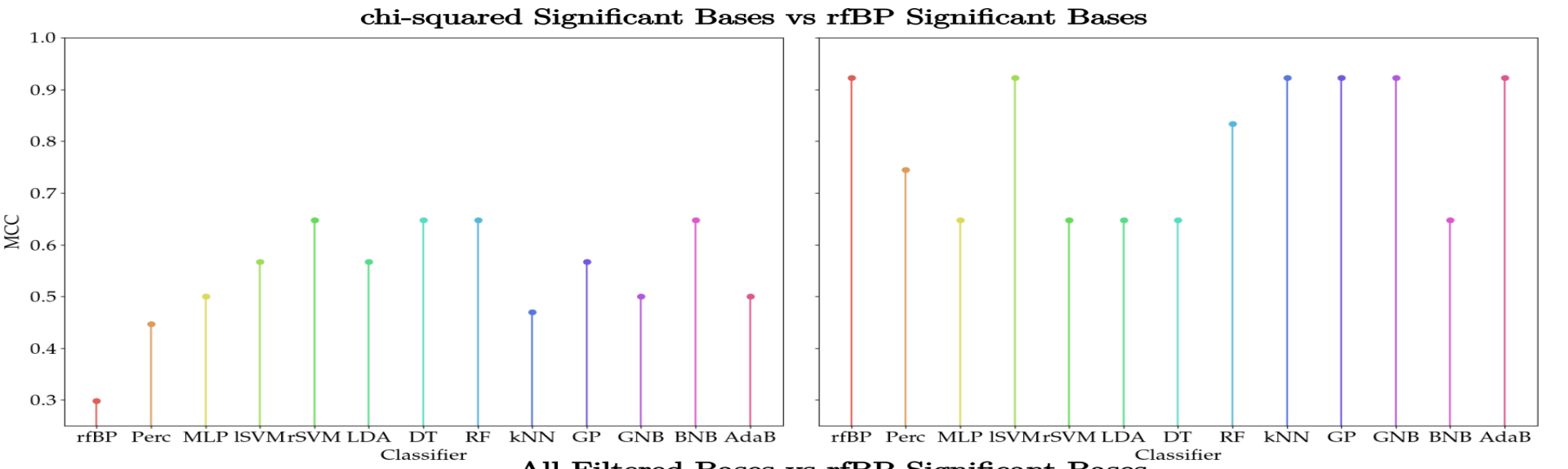
European Commission
Horizon 2020
European Union funding
for Research & Innovation

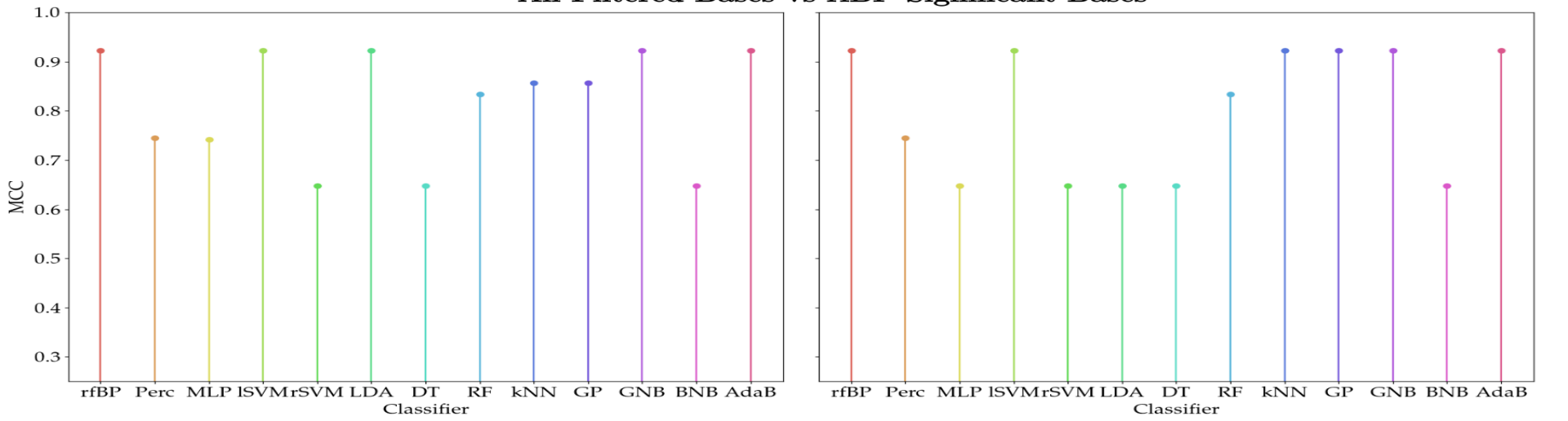


Performances



Features Selection





Conclusion

- Classification of binary SNPs for Source Attribution is optimally performed by Entropy-maximization based algorithm called rfBP
- SNPs binary nature seems to favour rfBP binary properties
- rfBP significant bases are much better to classify than test association significant bases → can be actively used for Features Selection
- New C++ and Python (scikit-learn format) implementations are user-friendly and can be efficiently run on real data

Link: https://github.com/Nico-Curti/rFBP

Acknowledgement