



Aristotle University of Thessaloniki



# Combining machine learning and network analysis pipelines: the case of microbiome and metabolomics data in colorectal cancer



Ntzioni Eleni

[eleni.ntzioni@gmail.com](mailto:eleni.ntzioni@gmail.com)

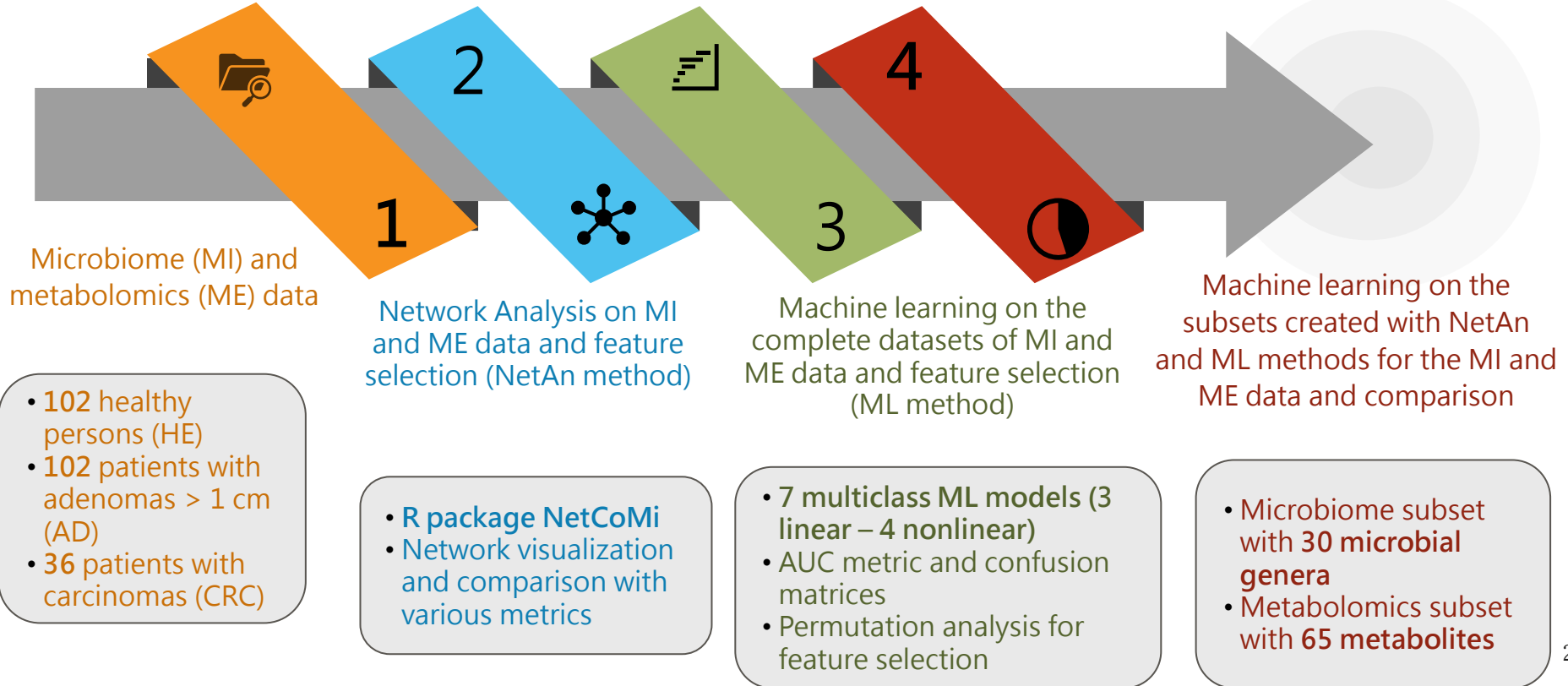
Chouvarda Ioanna

[ioannach@auth.gr](mailto:ioannach@auth.gr)

This work has received funding from H2020 programme under Grant Agreement 952179.



# Data and Methodology



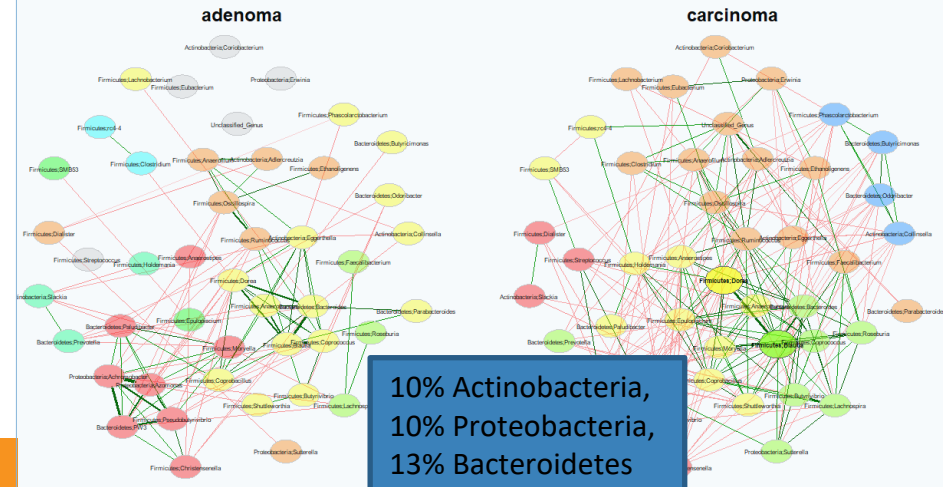


## Results of Network analysis

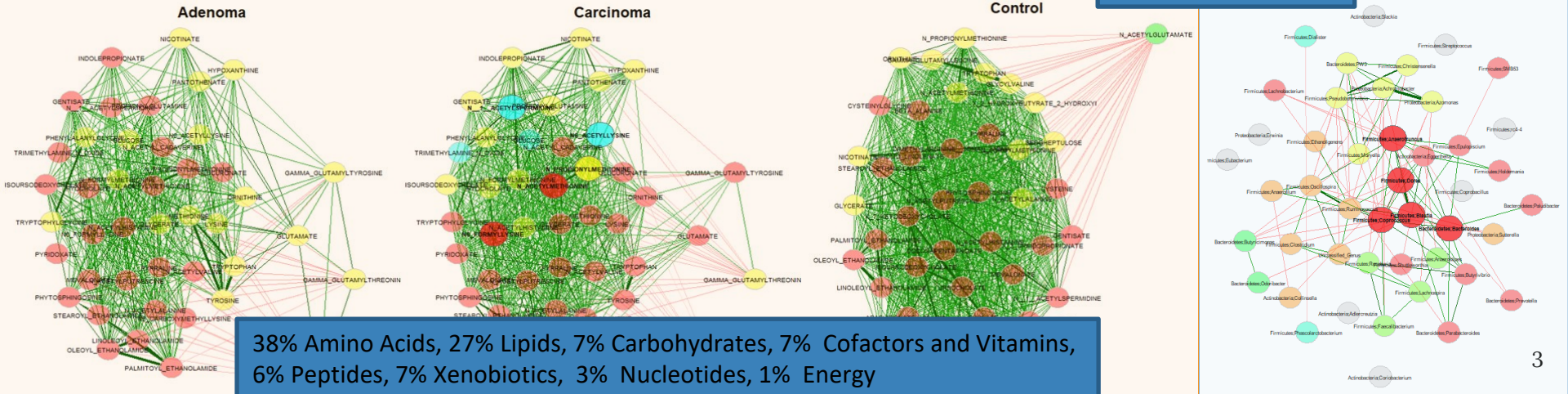
Metrics used for network comparison and feature selection

Degree	Jaccard index
Betweenness centrality	Adjusted rand index
Closeness centrality	Average path length
Eigenvector centrality	Clustering coefficient

## Metabolomics



10% Actinobacteria,  
10% Proteobacteria,  
13% Bacteroidetes  
και 67% Firmicutes

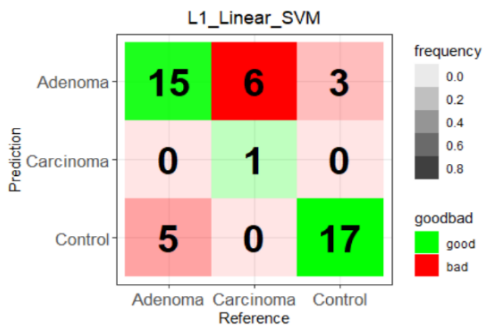
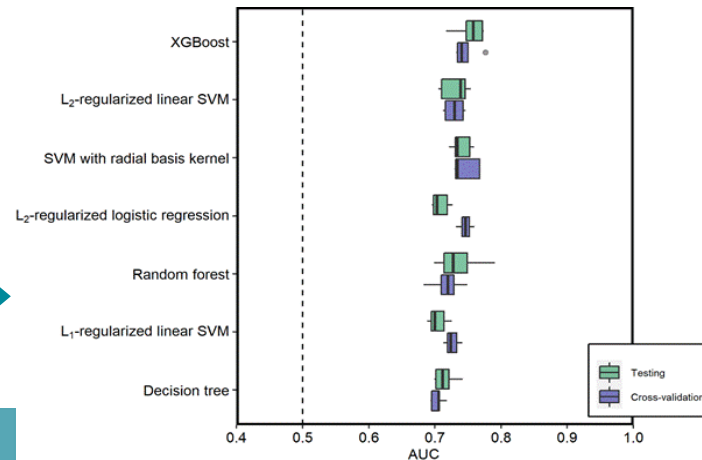
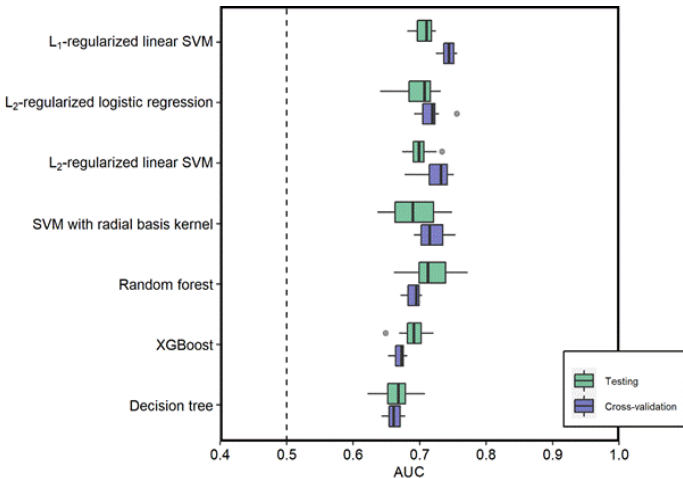
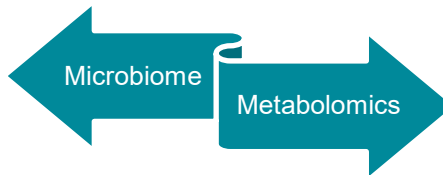


38% Amino Acids, 27% Lipids, 7% Carbohydrates, 7% Cofactors and Vitamins,  
6% Peptides, 7% Xenobiotics, 3% Nucleotides, 1% Energy

# AUC values and ConfMat - NetAn subsets



## Results of ML models

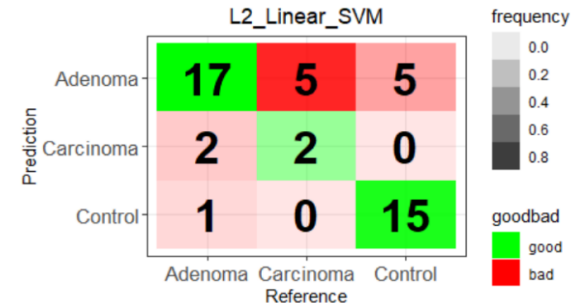


**Average success rates for all 7 models and all methods**

HE ~ 77%	HE ~ 80%
AD ~ 78%	AD ~ 81%

**Best success rates for all 7 models and all methods**

CRC ~ 29 %	CRC ~ 43%
------------	-----------





# Conclusions

- Overlap of the features selected with NetAn method and ML method (MI: 14/30, ME: 8/65).
- Linear models have greater AUC values.
- ML on subsets from NetAn method produced slightly better results, especially for HEALTHY class.
- Slightly better results on the analysis of ME data (AUC: 0.68-0.71) than on MI data (AUC: 0.64-0.70) *ME had a much bigger dataset. Can we combine them?*
- Very low prediction rate for CRC class for both datasets, all approaches and all 7 models.

Results show that the fewer samples of the CRC class had a strong negative effect on the performance of the models. For a set of samples where the number of samples will be equal to the other classes, we consider that the features selected with NetAn may lead to better predictions than a classic feature selection method.

---

## References

1. Kim M, Vogtmann E, Ahlquist D, Devens M, Kisiel J, Taylor W et al. Fecal Metabolomic Signatures in Colorectal Adenoma Patients Are Associated with Gut Microbiota and Early Events of Colorectal Cancer Pathogenesis. *mBio*. 2020;11(1).
2. Peschel S, Müller C, von Mutius E, Boulesteix A, Depner M. NetCoMi: network construction and comparison for microbiome data in R. *Briefings in Bioinformatics*. 2020;22(4).
3. Topçuoğlu B, Lesniak N, Ruffin M, Wiens J, Schloss P. A Framework for Effective Application of Machine Learning to Microbiome-Based Classification Problems. *mBio*. 2020;11(3).