Artificial intelligence methods for biomedical imaging and omics data

Workshop - AI per la Medicina e la Salute

Carlo Alberto Barbano[1,2],
Marco Beccuti[1],
Francesca Cordero[1],
Desislav Nikolaev Ivanov[1],
Nicola Licheri[1],
Simone Pernice[1],
Alberto Presta[1],
Riccardo Renzulli[1]
Marco Grangetto[1]

riccardo.renzulli@unito.it
simone.pernice@unito.it

[1] *Computer Science dept., University of Turin, Italy*
[2] *LTCI, Télécom Paris, IP Paris, France*

Ital-IA
ITALIA INTELLIGENZA ARTIFICIALE
Terzo Convegno Nazionale
Pisa 29-31 Maggio 2023
cini National Lab AIIS

EIDOS Lab

- Digital Pathology
- Digital Radiology
- Neuroimaging

qBio Lab

- Omics data for biomarkers discovery
- Computational epidemiology

EIDOS Lab

We are the **computer vision** and **image processing** group in the Computer Science department of the University of Turin.

Our Lab is part of the Italian Association for research in Computer Vision, Pattern Recognition and machine learning (CVPL).

https://eidos.di.unito.it/

EIDOS Lab

Digital Pathology

Digital Radiology

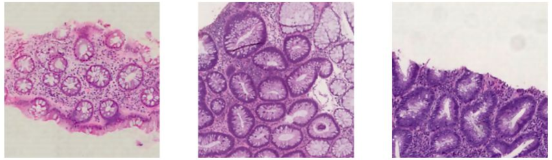
Neuroimaging

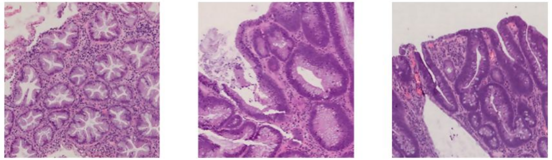
qBio Lab

Omics data for biomarkers discovery

Computational epidemiology

UniToPatho


(a) NORM (b) TA.LG (c) TA.HG


(d) HP (e) TVA.LG (f) TVA.HG

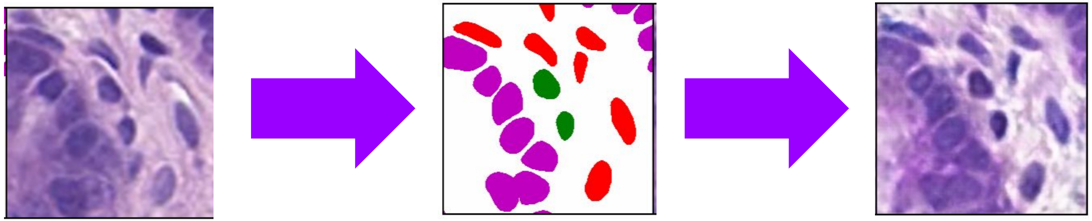
UniToPatho [1] is an annotated dataset of **9536** hematoxylin and eosin stained **patches** extracted from **292 whole-slide** images, meant for training deep neural networks for **colorectal polyps classification** and **adenomas grading**. Each slide belongs to a different patient and is annotated by expert pathologists, according to **six classes** as follows:

NORM - Normal tissue;
TA.LG - Tubular Adenoma, Low-Grade dysplasia;
TA.HG - Tubular Adenoma, High-Grade dysplasia;
HP - Hyperplastic Polyp;
TVA.LG - Tubulo-Villous Adenoma, Low-Grade dysplasia.
TVA.HG - Tubulo-Villous Adenoma, High-Grade dysplasia;

https://ieee-dataport.org/open-access/unitopatho

Generative models for synthetic data augmentation

Typically, medical data is not big data. In order to suffice this lackness, with a Generative Adversarial Network (GAN) we **generate new data** which contain the information we desire the deep model to learn. The new synthetic data can be used for data **augmentation**.



EIDOS Lab

- Digital Pathology
- **Digital Radiology**
- Neuroimaging

qBio Lab

- Omics data for biomarkers discovery
- Computational epidemiology

COVID-19 and the C.o.R.S.A. project

2019

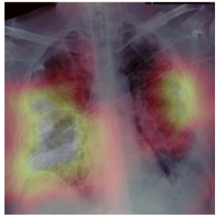
Initial outbreak in Wuhan quickly spread worldwide
COVID-19 pandemic

2020

Early efforts to build **datasets** [2]

CORDA: Covid **R**adiographic images **D**ata-set for **AI**
Collaboration with *Città della Salute e della Scienza* and *San Luigi Gonzaga*

- 898 CXRs on patients with fever or respiratory symptoms
- Virus testing to determine COVID infection
- Collected in March and April 2020 (peak of the epidemic in Italy)



2021

Direct COVID-19 diagnosis from CXR is hard with **limited data**
Differences in negative vs positive population have a major impact

2022

Kick-off Piemonte regional project **Co.R.S.A.** (Covid Radiographic imaging System based on AI)

2023

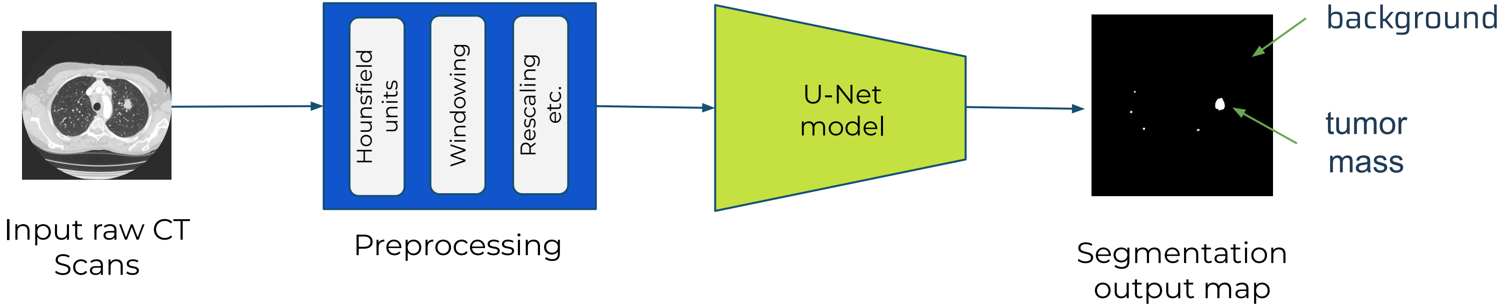
CORDA dataset released*, 3000+ images (both CXRs and CTs)
More hospitals (mauriziano, molinette, san luigi, monzino)

*https://zenodo.org/record/7501816#.ZAeAv9LMJhE

UniToChest

https://zenodo.org/record/5797912#.Y_yoztLMJhE

UniToChest [3] is a collection of anonymized **306440** chest CT scan slices coupled with the proper lung nodule segmentation map, for a total of **10071** nodules from **623** different patients

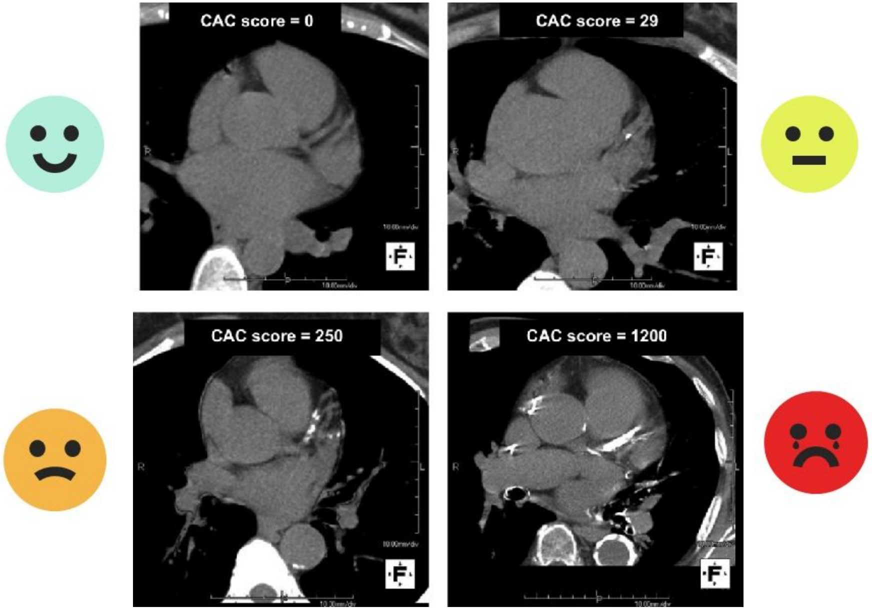


Input raw CT Scans

Preprocessing

Segmentation output map

Calcium score prediction

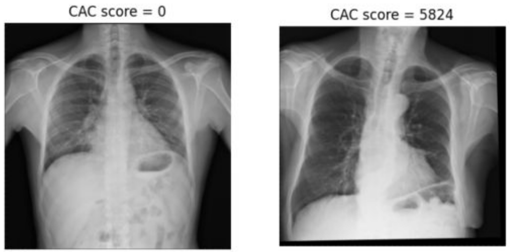


Coronary artery calcium (CAC) -> amount of calcified plaque in the coronary arteries.

Coronary artery disease (CAD) (heart attacks or other cardiovascular events).

High CAC → High CAD

New dataset (publicly available soon) collected with Citta della Salute e della Scienza di Torino with total of 507 radiographs coming from 507 different patients.



EIDOS Lab

- Digital Pathology
- Digital Radiology
- **Neuroimaging**

qBio Lab

- Omics data for biomarkers discovery
- Computational epidemiology

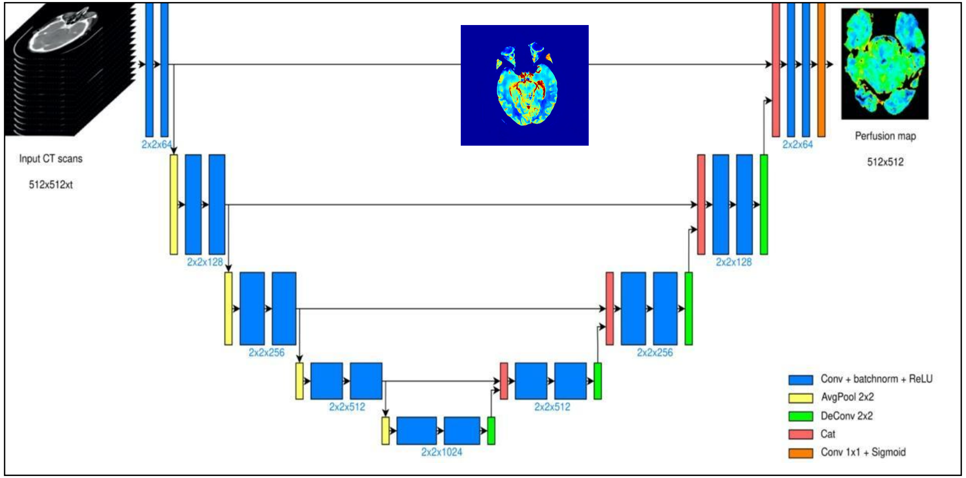
UnitoBrain

UniToBrain is a dataset of Computed Tomography (CT) **perfusion images** (CTP). It includes 258 patients [4].

https://ieee-dataport.org/open-access/unitobrain

Deep learning **model** [5] should be **robust** to:

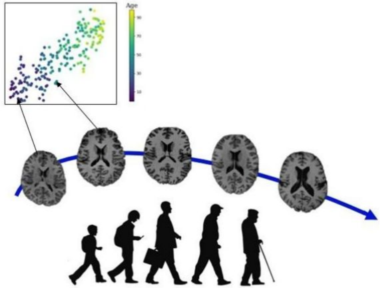
- reduced number of acquisitions -> lower dose of radiations
- acquisition errors/noise

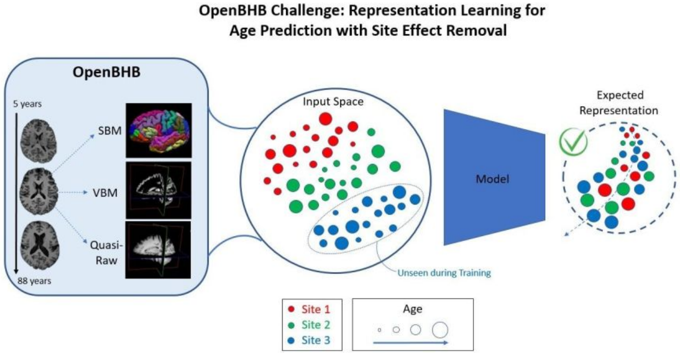


Brain age prediction

Brain aging involves complex biological processes, such as cortical thinning, that are highly heterogeneous across individuals, suggesting that **people do not age in the same manner**.

Accurate modeling of brain ageing help **better understand neurodegenerative disorders** and in finding **new biomarkers**.





Novel contrastive learning regression loss for robust brain age prediction using MRI scans achieves **state-of-the-art performance** on the OpenBHB challenge, yielding the best generalization capability and robustness to site-related noise [6].

Takeaway

Lessons learned so far

- **Datasets** should be **unbiased** and **balanced**
- Importance of **preprocessing** (especially on small dataset)
- **Public** dataset & GDPR and **ethical** issues
- Inspect (try to **understand**, to constrain) **extracted features**
- Exploit **learning complexity** to unveil unexpected phenomena **challenging prior assumptions**

References

[1] C. A. Barbano, D. Perlo, E. Tartaglione, A. Fiandrotti, L. Bertero, P. Cassoni, M. Grangetto, **Unitopatho, a labeled histopathological dataset for colorectal polyps classification and adenoma dysplasia grading**, in: IEEE ICIP, 2021.

[2] E. Tartaglione, C. A. Barbano, C. Berzovini, M. Calandri, M. Grangetto, **Unveiling covid-19 from chest x-ray with deep learning: a hurdles race with small data**, International Journal of Environmental Research and Public Health 17 (2020) 6933.

[3] H. A. H. Chaudhry, R. Renzulli, D. Perlo, F. Santinelli, S. Tibaldi, C. Cristiano, M. Grosso, G. Limerutti, A. Fiandrotti, M. Grangetto, et al., **Unitochest: A lung image dataset for segmentation of cancerous nodules on ct scans**, in: ICIAP, 2022.

[4] U. A. Gava, F. D'Agata, E. Tartaglione, R. Renzulli, M. Grangetto, F. Bertolino, A. Santonocito, E. Bennink, G. Vaudano, A. Boghi, M. Bergui, **Neural network-derived perfusion maps: a model-free approach to computed tomography perfusion in patients with acute ischemic stroke**, Frontiers in Neuroinformatics (2023).

[5] D. Perlo, E. Tartaglione, U. Gava, F. D'Agata, E. Benninck, M. Bergui, **Unitobrain dataset: A brain perfusion dataset**, in: ICIAP, 2022.

[6] C. A. Barbano, B. Dufumier, E. Duchesnay, M. Grangetto, P. Gori, **Contrastive learning for regression in multi-site brain age prediction**, in: IEEE ISBI 2023.

qBio Lab

We are the ***Quantitative Biology*** (q-Bio) group of the Computer Science Department in the University of Turin.

Our interests regard mainly the **development of computational models able to integrate several sources of data and knowledge** to give new insights into the comprehension of the biological mechanisms at the bases of the disease considered and in the definition of patient tailored therapies.

https://qbio.di.unito.it/

EIDOS Lab

- Digital Pathology
- Digital Radiology
- Neuroimaging

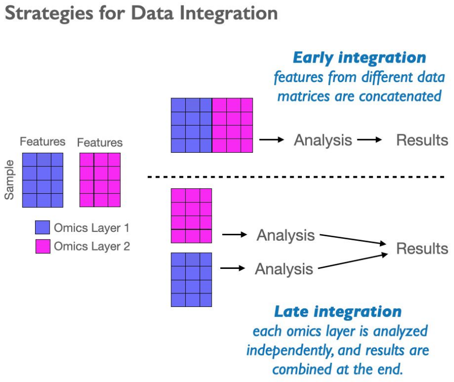
qBio Lab

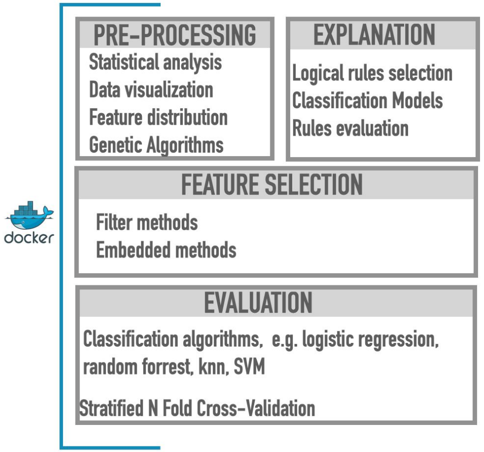
- **Omics data for biomarkers discovery**
- Computational epidemiology

Multi-omics data integration

FeatSEE
Feature Selection, Extraction, Evaluation

Modular framework to ***define workflows of analysis for multi-omics integration***, for automated machine learning usable by researchers without computational skills, to help them in the functional interpretation of results from a clinical/biological point of view.





PRE-PROCESSING
- Statistical analysis
- Data visualization
- Feature distribution
- Genetic Algorithms

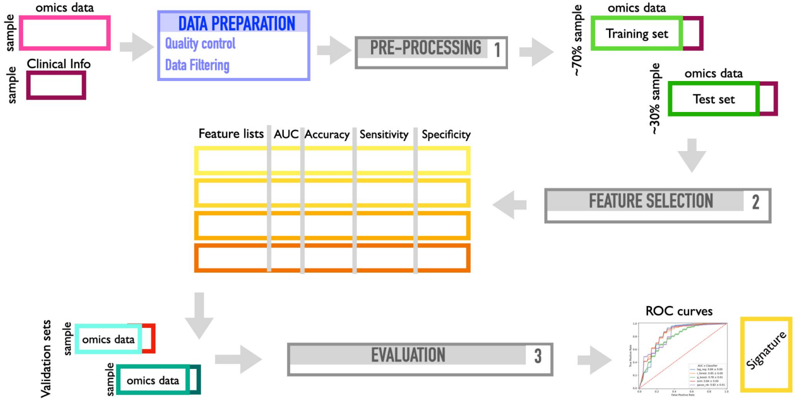
EXPLANATION
- Logical rules selection
- Classification Models
- Rules evaluation

FEATURE SELECTION
- Filter methods
- Embedded methods

EVALUATION

Classification algorithms, e.g. logistic regression, random forrest, knn, SVM

Stratified N Fold Cross-Validation

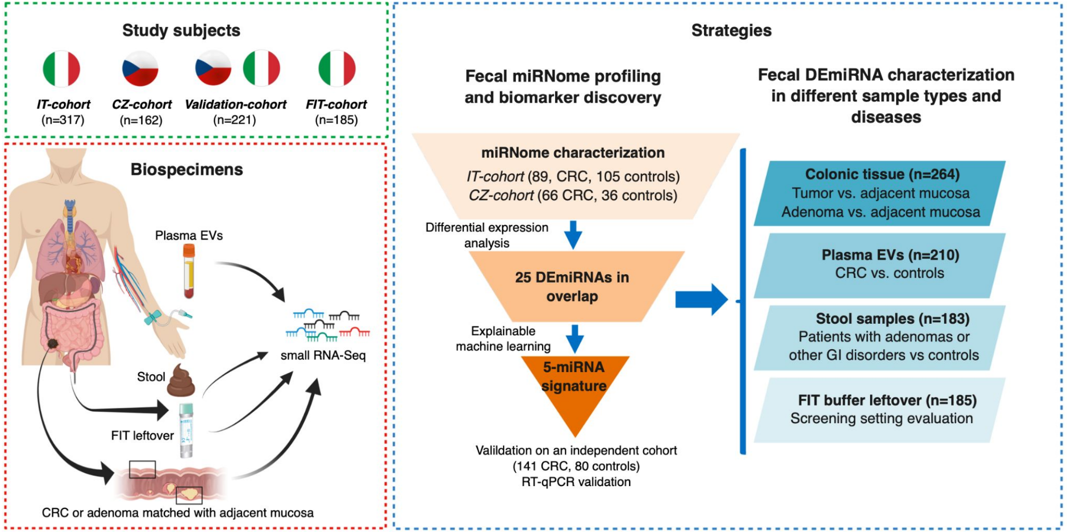


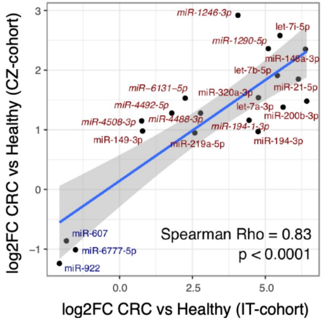
Licheri N., et al. In preparation

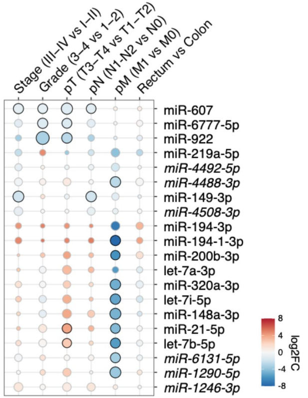
Multi-omics data integration

Colorectal Cancer (CRC): Small RNA seq

***Aim**: identification of biomarkers for CRC with low stages vs high stages.*







Identification of 20 miRNA differentially expressed in CRC patients versus controls in both Italian and Prague cohorts by DeSeq2 library, model corrected by sex and age.

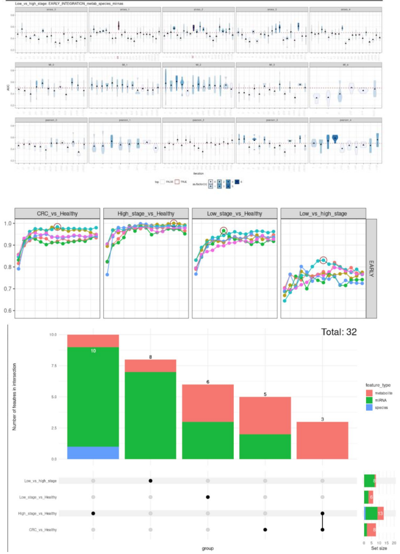
Pardini., et al. Gastroenterology. Under review.

Multi-omics data integration

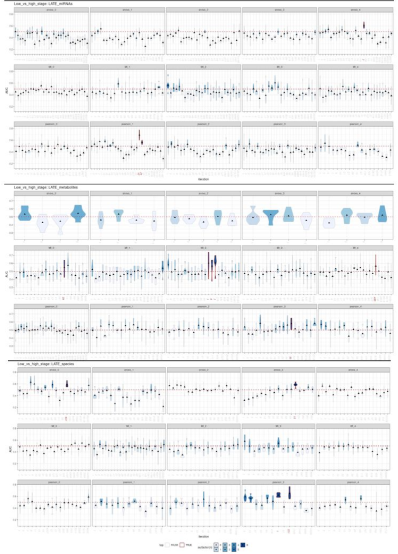
Colorectal Cancer (CRC): Small RNA seq · Metagenomic Sequencing · Metabolomics

***Aim**: identification of biomarkers for CRC with low stages vs high stages.*

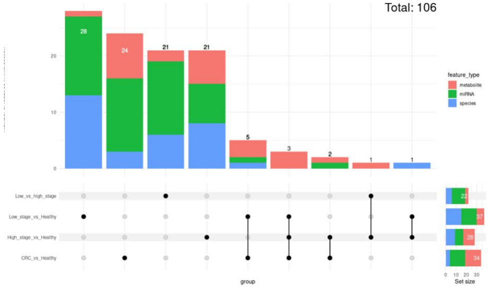
Early approach

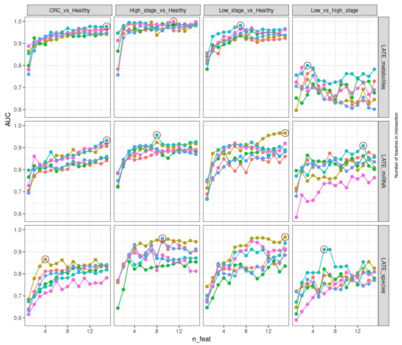


Late approach



Overview on meta results



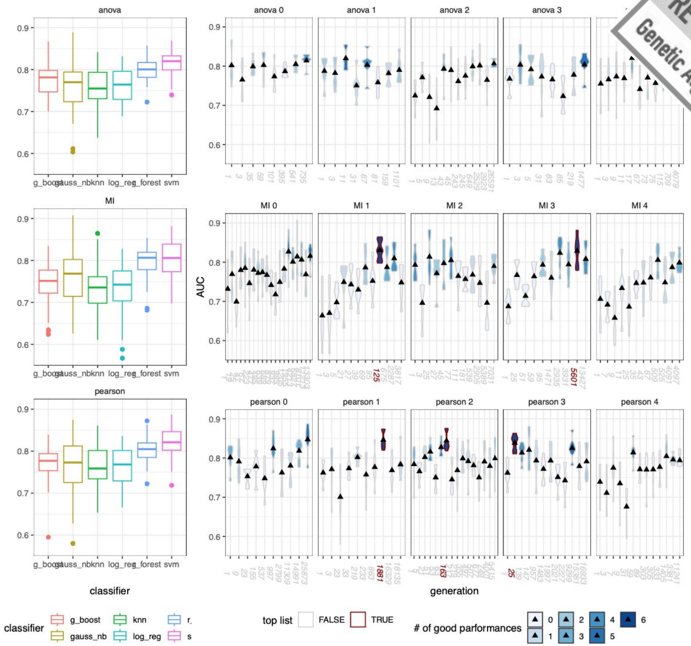


WORKING IN PROGRESS...

Multi-omics data integration

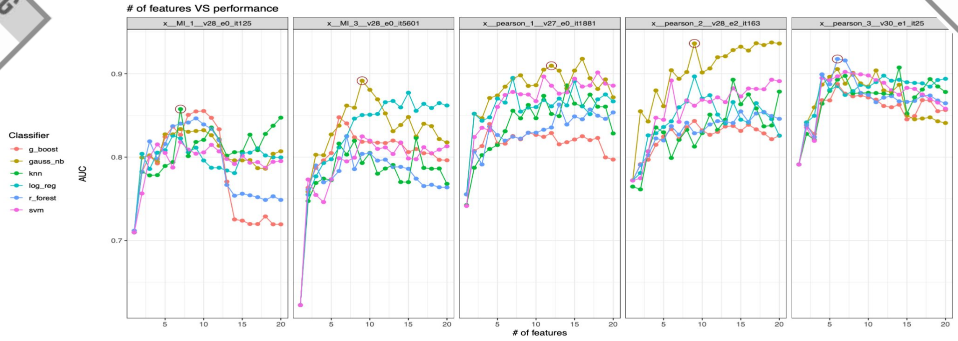
Metastatic Colorectal Cancer (mCRC): Copy number variation, gene expression, mutational profiles

Aim: *identification of prognostic biomarkers.*



PRE-PROCESSING
Genetic Algorithms

FEATURE SELECTION
Filter methods



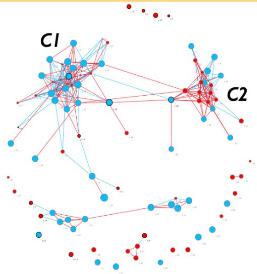
WORKING IN PROGRESS...

Multi-omics data integration

Embryo: *Basal, Clinical and Embryological Features*

Aim: *morphokinetic features that identify cleavage-stage embryos with the best chance to reach the expanded blastocyst stage*

ICSI - Selected Rules



BL group Rules

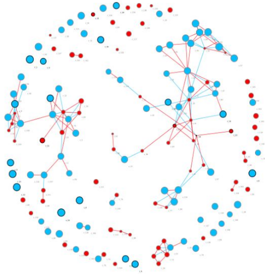
fertilization_rate <= 89.444 and imcs > 6.000 and t4t2 <= 12.040
fertilization_rate <= 94.444 and t4 <= 38.250 and t4t2 <= 12.248
fertilization_rate <= 94.444 and imcs > 6.250 and tpnf <= 23.960
imcs > 6.000 and t4 <= 39.125 and t8 <= 56.085
t3 <= 38.165 and t4t3 <= 1.375
t4 <= 38.292 and t4t2 <= 12.875 and t4t2 > 6.005
t4 > 31.623 and t8 <= 52.627 and afc <= 28.500
t4 > 31.793 and t8 <= 56.085 and t4t2 <= 12.830
tpnf <= 21.918 and t4t3 <= 1.375
tpnf <= 22.460 and t4 > 31.458 and t4t2 <= 12.332
tpnf <= 22.835 and t4t2 <= 11.918

nBL group Rules

bmi > 19.350 and t4 > 34.415 and t8 > 52.667
maturation_rate > 45.455 and t8 > 52.667 and
and t2tpnf <= 10.465
t3 > 27.585 and t4 > 35.125 and t8 > 51.500

	nBL Rules	BL Rules
Embryological features	7	26
Basal features	1	1
Clinical features	1	3

IVF - Selected Rules



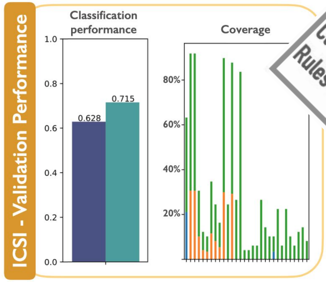
BL group Rules

fertilization_rate > 57.273 and imcs > 6.300 and t4-t2 <= 13.370
fertilization_rate > 57.273 and t4-t2 <= 12.915 and afc > 15.500
fertilization_rate > 65.461 and imcs > 6.400 and t3-t2 <= 12.325
imcs > 6.150 and t3-t2 <= 12.325 and t4-t3 <= 1.880
imcs > 6.350 and t4-t2 <= 12.745
imcs > 6.350 and t4-t2 <= 14.795 and total_fsh <= 1706.000
imcs > 6.400 and t4 <= 39.875 and t4-t2 <= 12.960
t2 <= 27.540 and t3-t2 <= 12.710
t2 <= 32.625 and t4-t3 <= 1.120
t2 > 22.290 and t8 <= 52.955 and t3-t2 > 9.915
t2-tpnf <= 3.415 and t4-t3 <= 0.790 and afc > 15.500
t3-t2 <= 12.835 and t3-t2 > 9.875 and t8-t4 <= 17.875
t4 <= 45.125 and t3-t2 <= 12.325 and t3-t2 > 8.000
t4 <= 45.125 and t3-t2 > 9.835
t8 <= 74.085 and t2-tpnf <= 3.415 and t4-t2 <= 12.745
t8 <= 75.170 and t3-t2 > 9.875 and t4-t2 <= 12.745

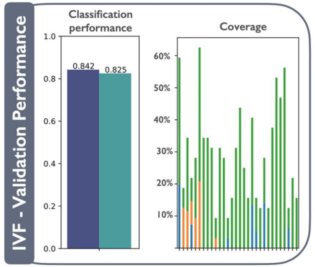
nBL group Rules

day_3_fsh <= 21.500 and t8 > 52.585 and t4-t3 > 1.120
fertilized_oocytes <= 15.500 and t4 > 36.125 and t8 > 60.750
t2 > 24.210 and t8-t4 > 17.875 and day_3_amh <= 11.350
t4 > 39.665 and t8-t4 > 19.460
t4-t3 > 1.120 and t8-t4 > 17.875

	nBL Rules	BL Rules
Embryological features	10	38
Basal features	2	1
Clinical features	1	5



EXPLANATION
Logical rules selection
Classification Models
Rules evaluation



These studies demonstrated an improved predictive ability for evaluation of embryo viability when compared with embryologists' traditional morphokinetic grading methods

Canosa and Licheri, et al, under review

EIDOS Lab

Digital Pathology

Digital Radiology

Neuroimaging

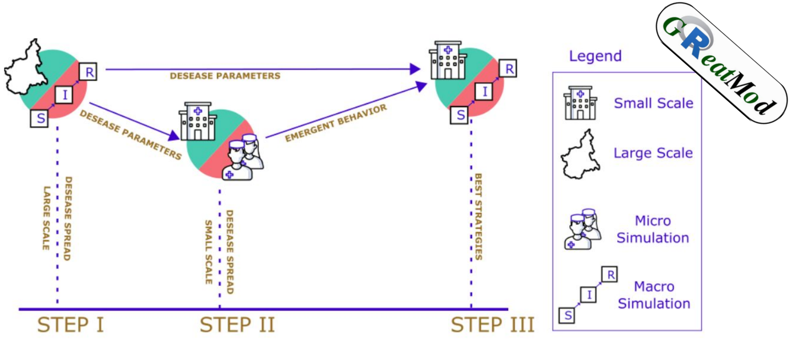
qBio Lab

Omics data for biomarkers discovery

Computational epidemiology

Computational Epidemiology

TrustAlert: development of AI solutions to integrate into a **smart platform** for providing early ***warnings***, ***monitoring*** and ***forecasting*** tools to ***public health*** response agencies and local healthcare services for anticipating medical needs.



Computational epidemiology exploits **AI** and **Simulation** to successfully support epidemiologists, healthcare professionals, and decision-makers to understand and control the spatio-temporal spread of infectious diseases.

- 2023 project founded by "La Compagnia di San Paolo" and "Fondazione CDP" for the second edition of the grant dedicated to AI
-GreatMod: https://qbioturin.github.io/epimod/

THE END

EIDOS Lab	**qBio Lab**