

Computational Intelligence for Computational Biology

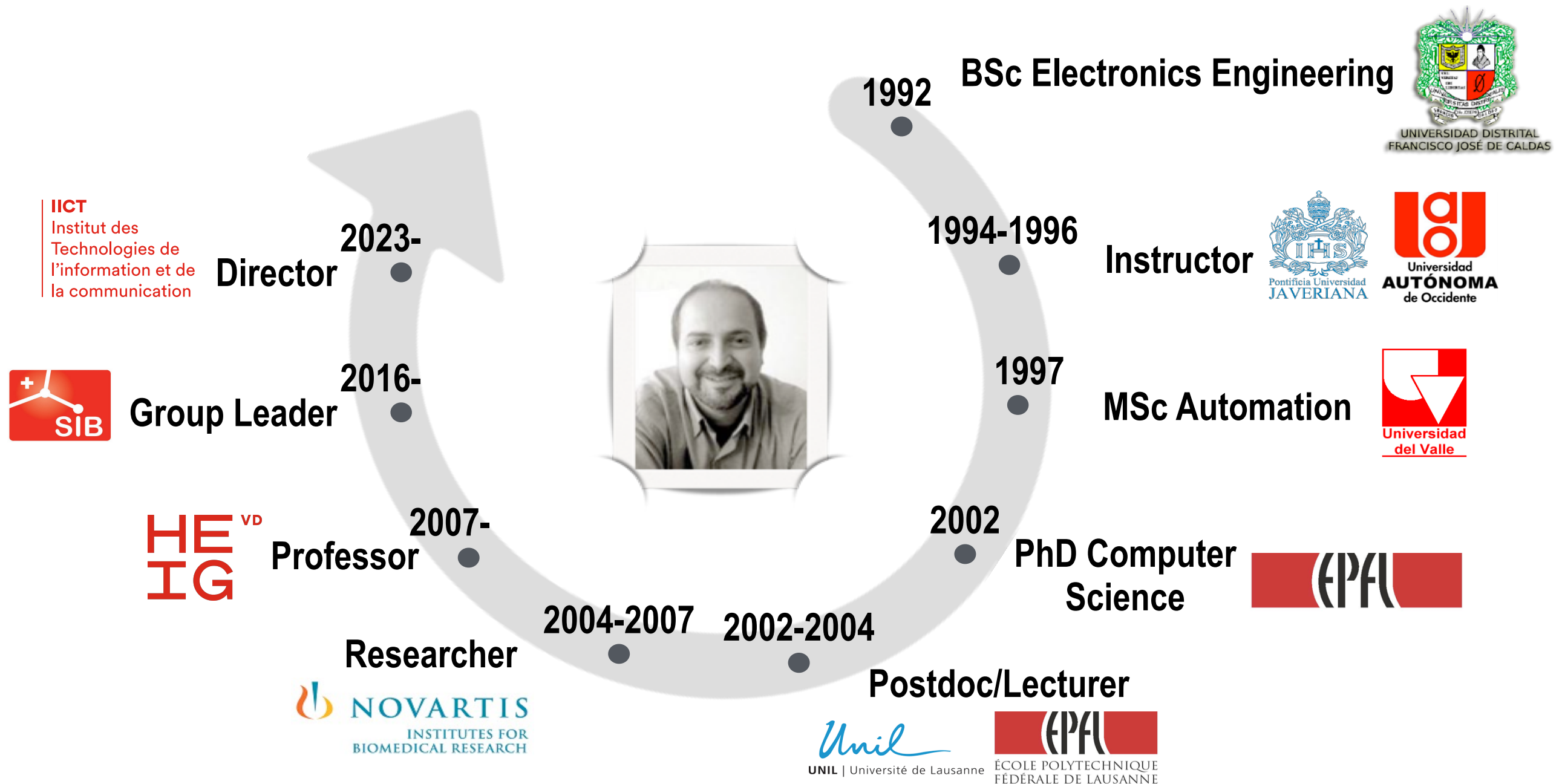


Swiss Institute of
Bioinformatics

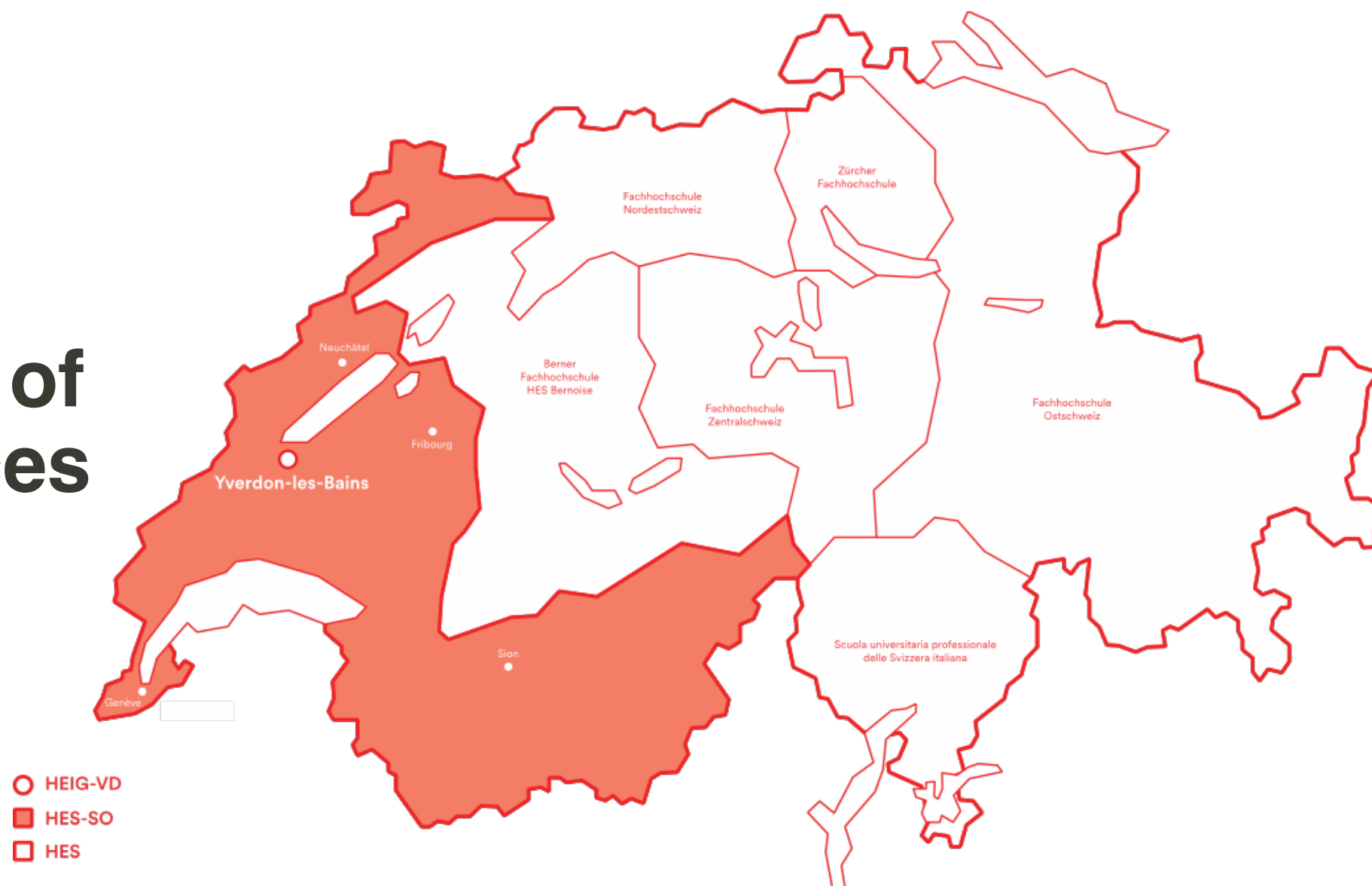
Carlos-Andrés Peña-Reyes

HEIG-VD / HES-SO

Prof. Carlos Andrés Peña



Our University of Applied Sciences HES-SO



4 missions

- Bachelor and Master Education
- Continued education
- **Applied Research and Development**
- National and International Collaborations

Our research unit: the IICT



VD

IICT

Institut des
Technologies de
l'information et de
la communication

Our mission

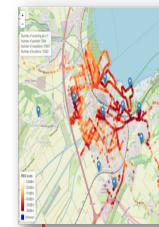
Applied research and development, in
collaboration with partners from industry,
academia, and public institutions



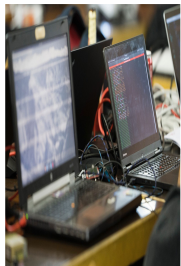
Cybersecurity



Data science



Internet of Things



Software
engineering



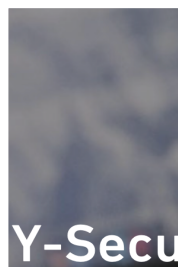
Biomedical
applications



Advanced
communication
systems



HEE - Health,
Engineering and
Economics



Y-Security



Optimisation et
durabilité



Our research group

3 main research interests

1. Biomarker discovery and characterization

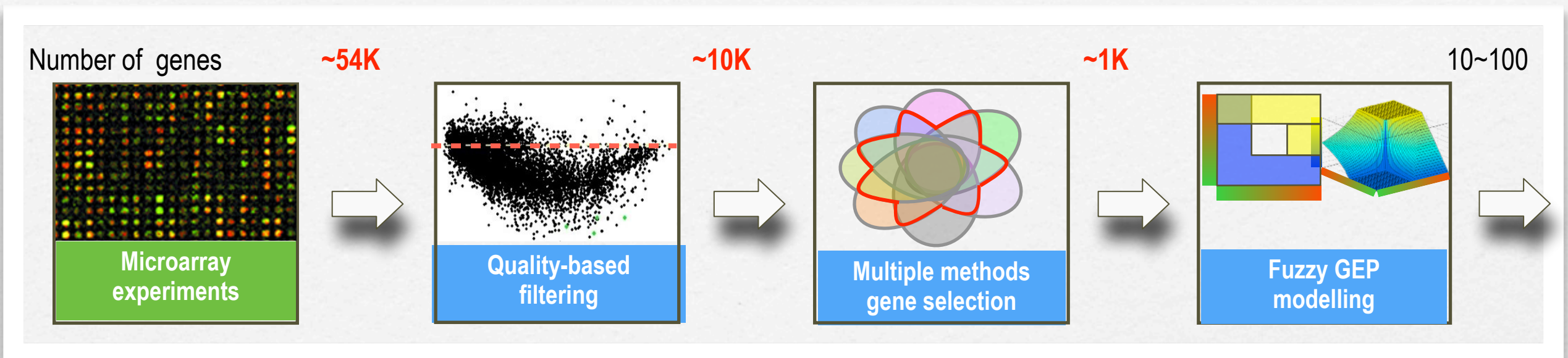
Robust feature selection and (explainable) predictive modelling

Machine learning-based,
data-driven biomarker discovery and characterisation,
particularly from high-dimensional datasets
such as RNA-seq transcriptomics.

We have developed pipelines combining
feature selection and fuzzy modelling,
which are currently being applied to model
biotechnological production processes in industrial environments.

Gene-expression profiling (GEP)

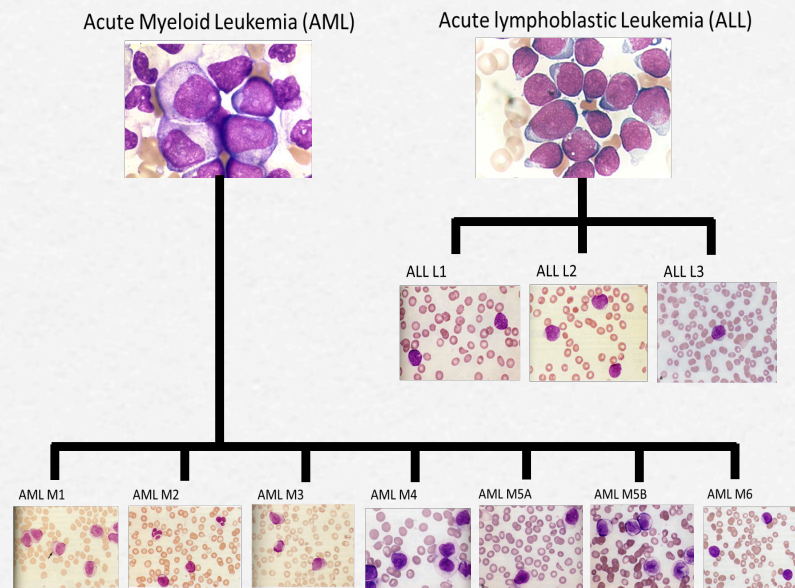
Innovative gene-selection techniques and GEP modeling



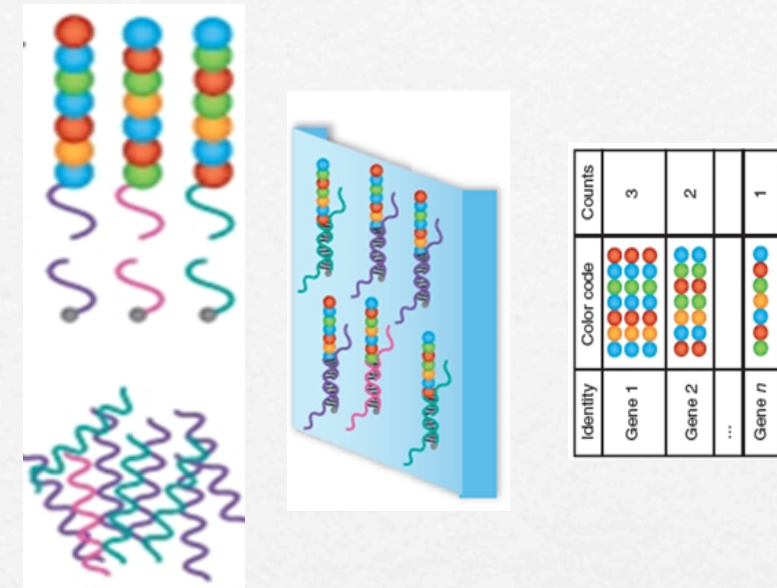
- ❑ Role of smoking on **lung cancer** development
- ❑ Novel test for diagnosing and classifying 18 **leukemia** subtypes
- ❑ Modelling biomarker activity for **colorectal cancer** diagnosis
- ❑ Feature selection and fuzzy modelling for **biotech processes**



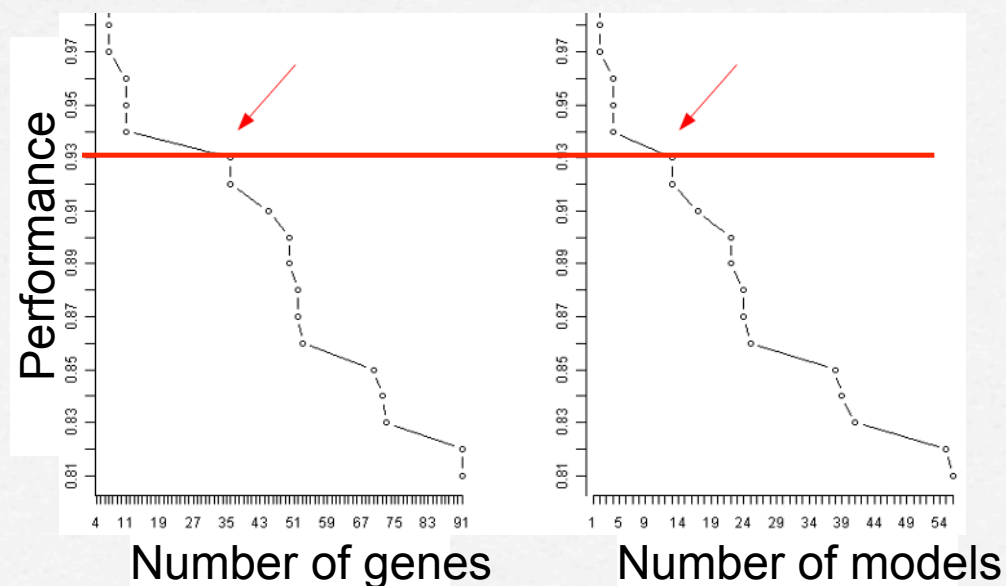
Gene and model selection: nanoFUGE



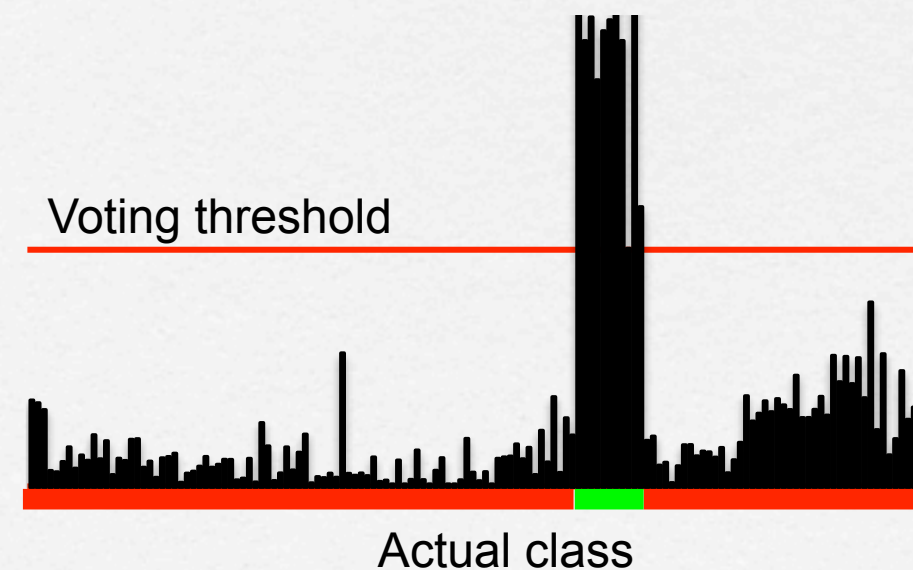
Leukemia: a complex disease



Nanostring: Reliable data source



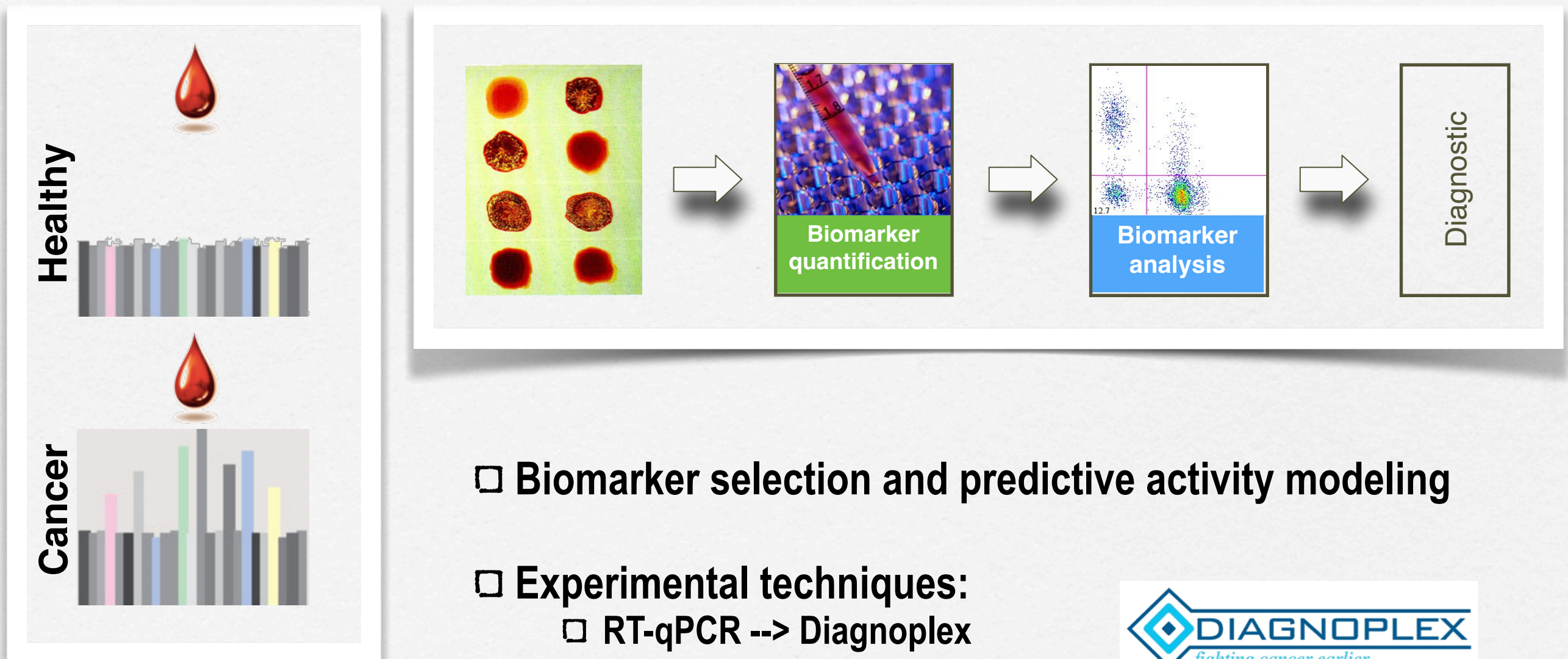
Selection of predictive gene pool



Final multi-model voting strategy

Biomarker discovery

Fuzzy modeling biomarker activity for cancer diagnosis



- ❑ Biomarker selection and predictive activity modeling
- ❑ Experimental techniques:
 - ❑ RT-qPCR --> Diagnoplex
 - ❑ Elisa-test --> Ayanda



Fuzzy modeling biomarker activity for colorectal cancer diagnosis

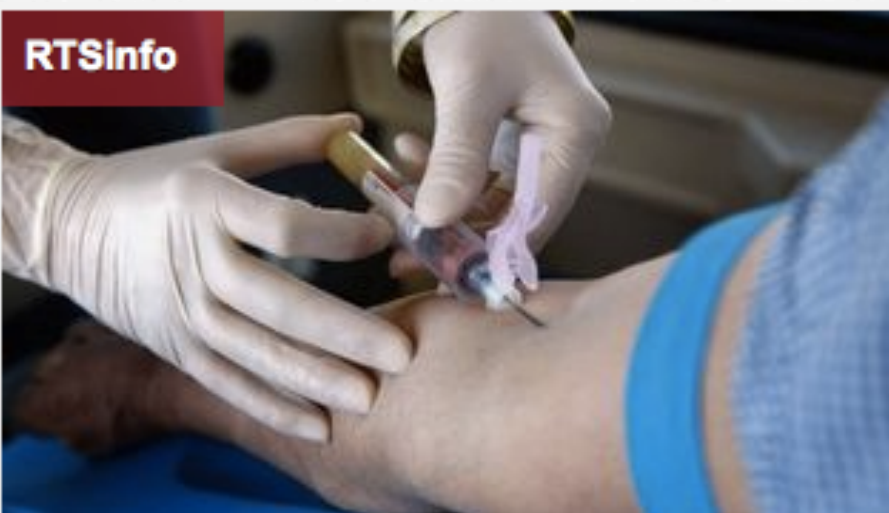
2010-2014



AP&T Alimentary Pharmacology and Therapeutics

A novel gene expression signature in peripheral blood mononuclear cells for early detection of colorectal cancer

C. Nichita^{*,1}, L. Ciarloni^{†,‡,1}, S. Monnier-Benoit[†], S. Hosseinian[†], G. Dorta^{*} & C. Rüegg^{‡,§}



RTSInfo

Un test de dépistage précoce du cancer du colon développé à Fribourg

Des chercheurs fribourgeois ont développé un test de détection précoce du cancer colorectal à partir d'une prise de sang. Le procédé doit être commercialisé cette année.

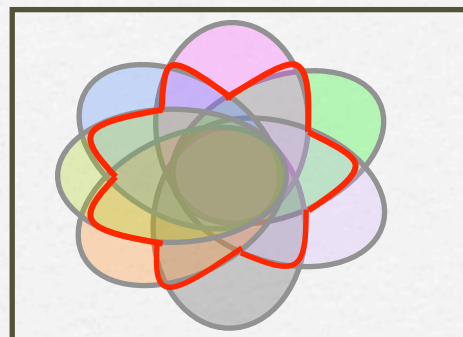
24 janvier 2014



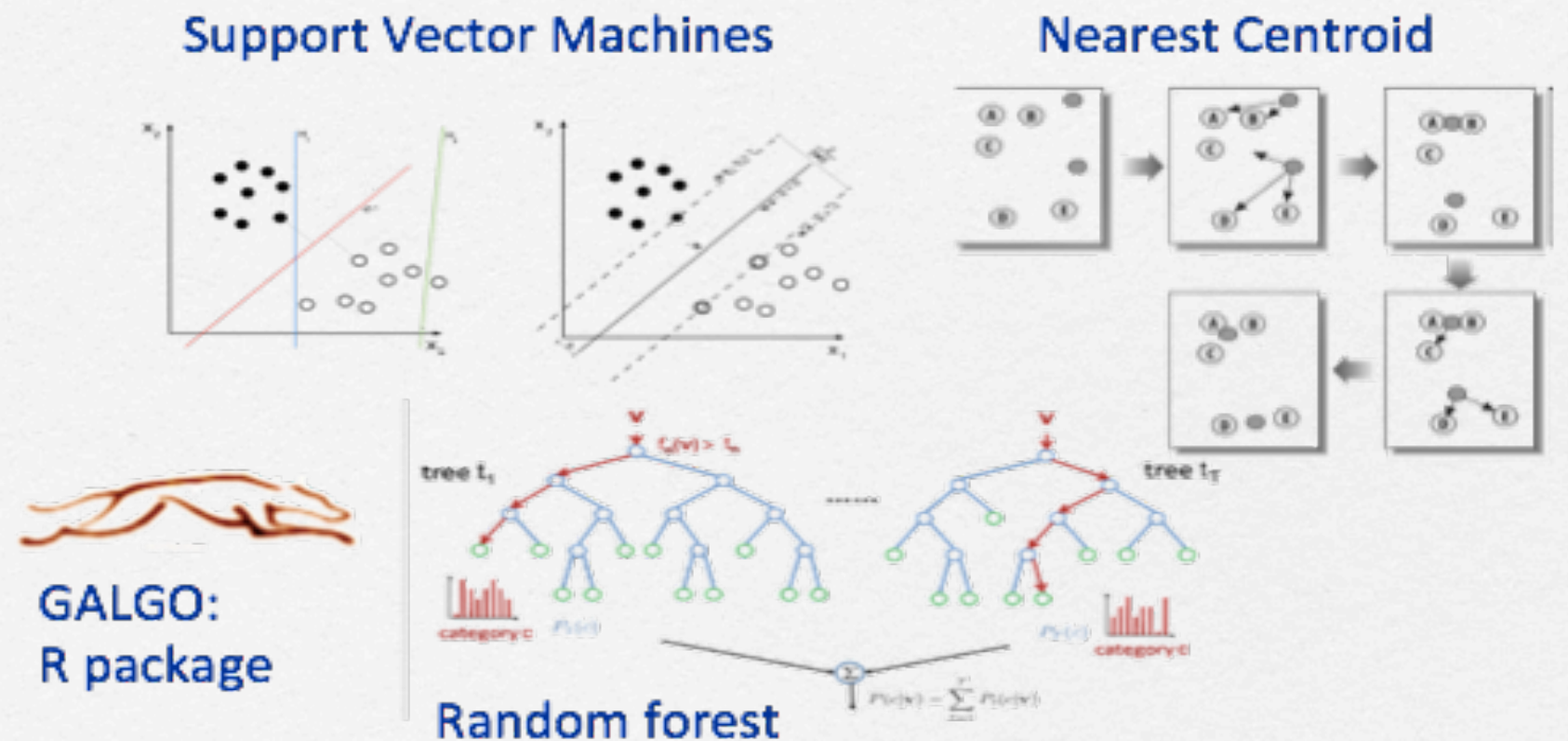
Dépister le cancer du colon avec une prise de sang

Robust gene selection

Feature selection based on a mix of several methodologies



Mix of methods
for gene selection

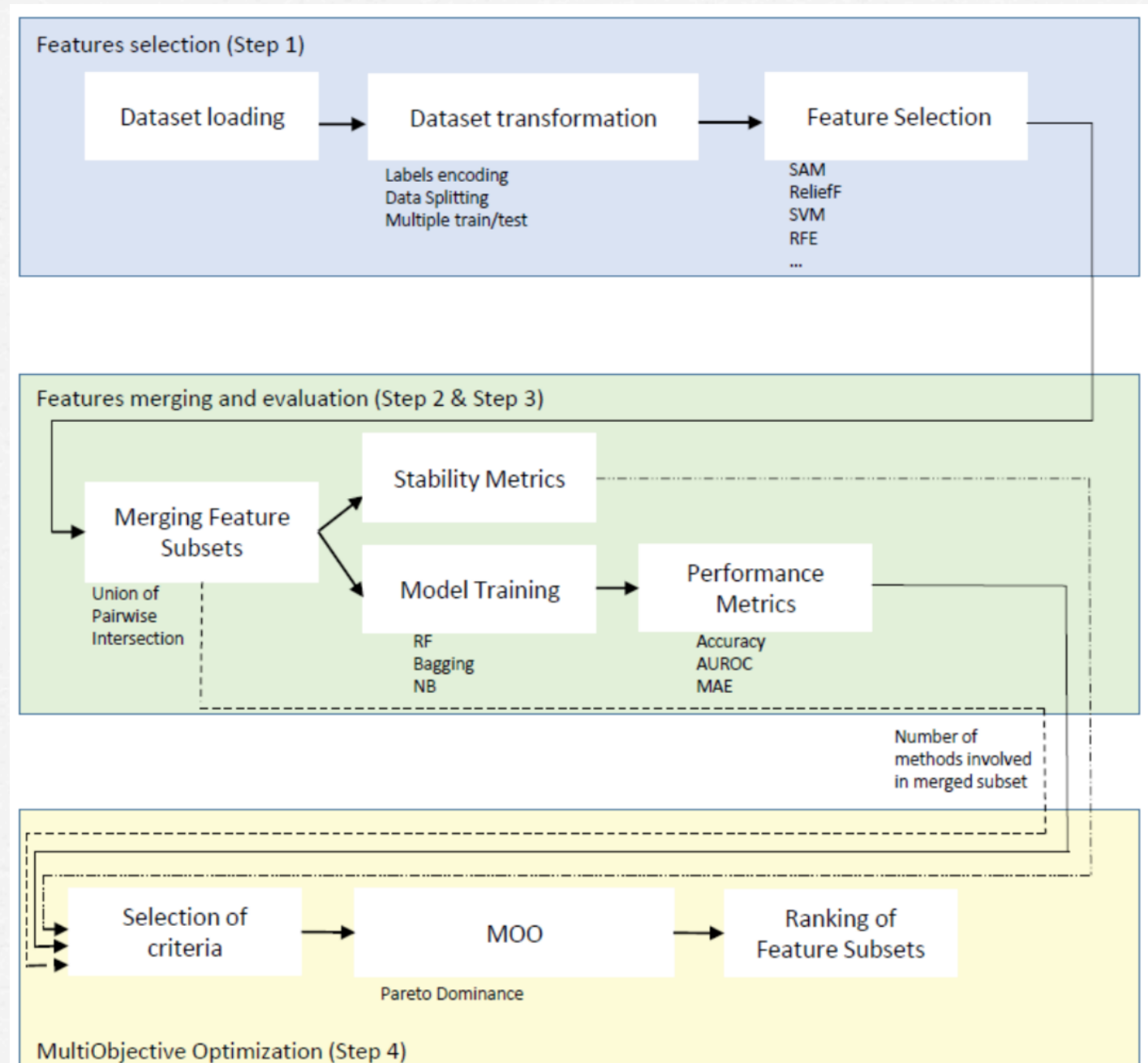


- Supervised, multivariate, multiclassifier approach
- Reduce method-specific biases

MOOSE-FS

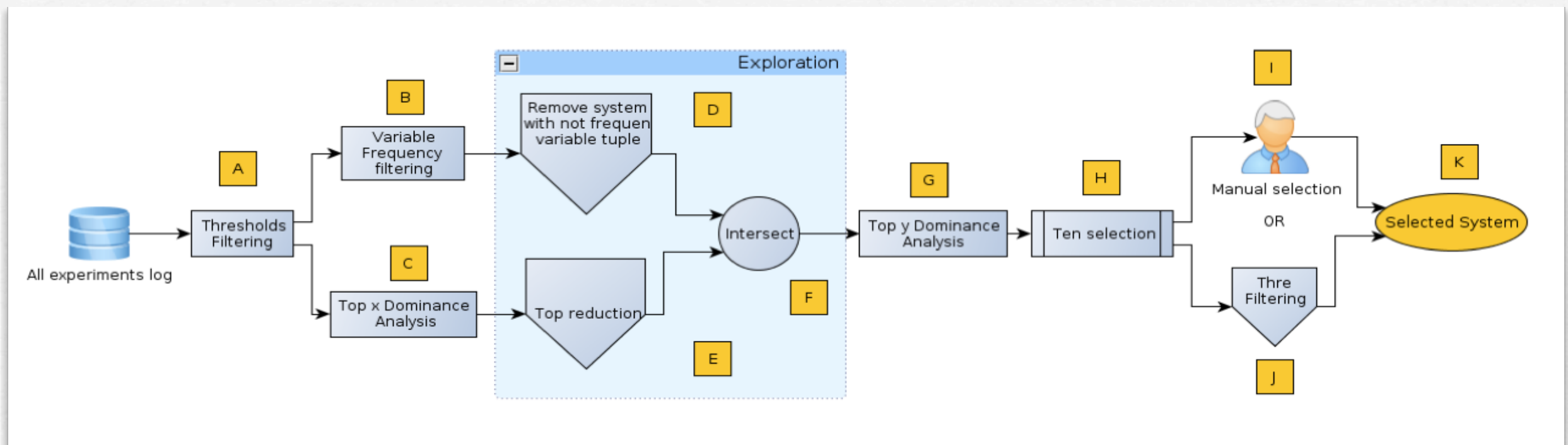
Multiple-Objective Optimization for Ensemble Feature Selection

- **Ensemble feature selection** pipeline using **multi-objective optimization** based on Dr. Zahra Mungloo-Dilmohamud's thesis.
- The benefit of this pipeline is
 - **Diversity and bias reduction** thanks to ensemble feature selection
 - **Robustness and performance** reinforced by MOO
- We aim to offer users greater control and **customization** by emphasizing flexibility and providing diverse options within the pipeline.



Model selection workflow

Robust construction of a final predictive model

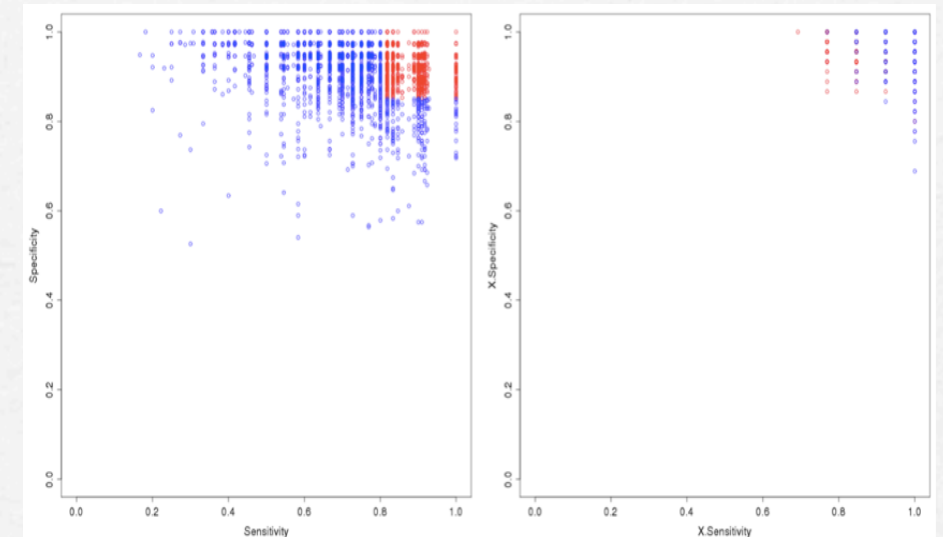
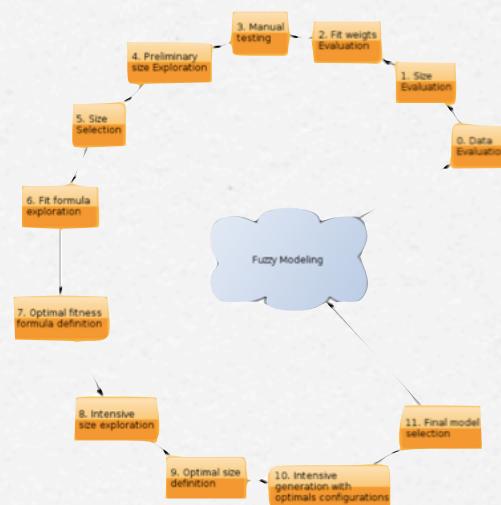
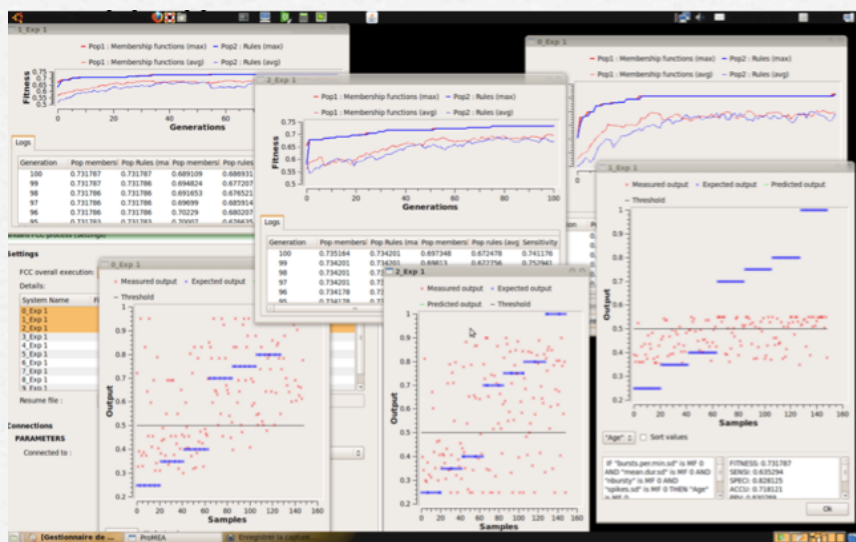


- Modeller-driven interactive process
- Multiple criteria for assessing model quality
 - Multivariate consistency
 - Performance-driven multicriteria dominance analysis

Building Computational Tools

SimplicityBio: A spin-off adventure (2013-->)

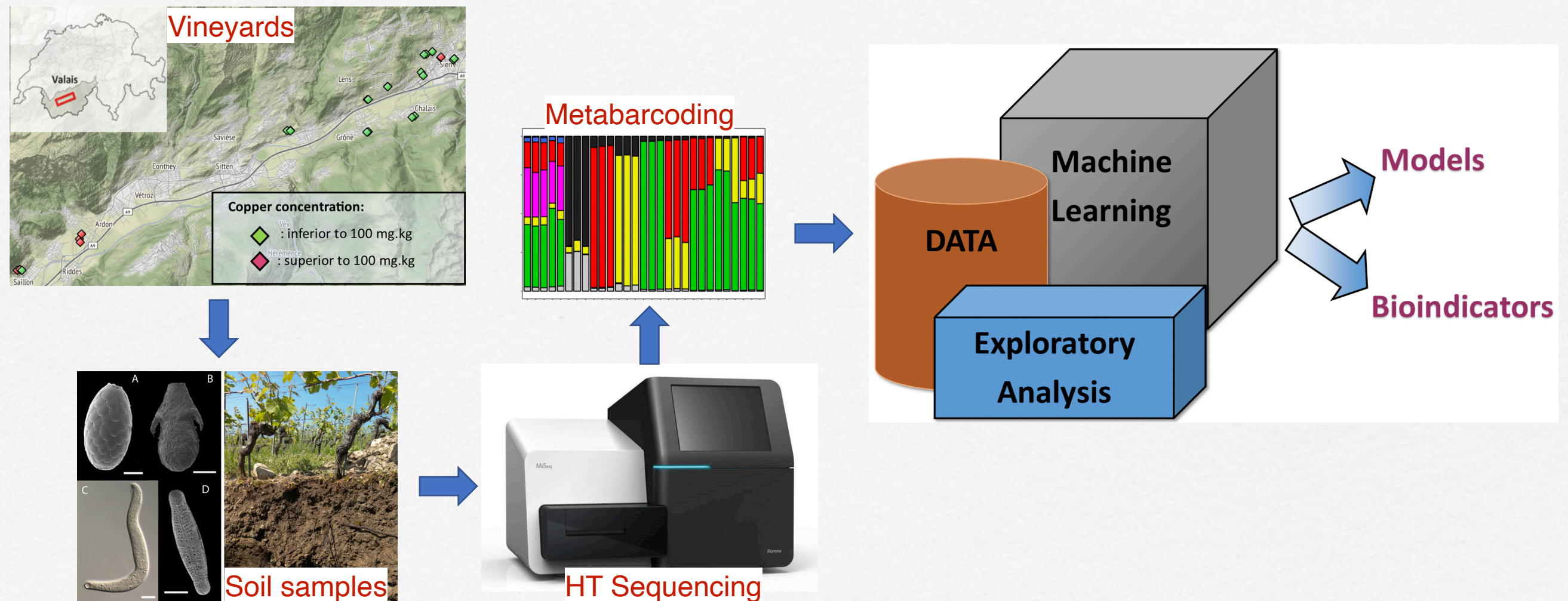
☐ Licensed software: FUGE-LC and FUGE-MM



☐ Business model based on experience from previous projects

- ☐ Biomarker discovery services
- ☐ Pursue on Leukemia test results
- ☐ CTI project with HEIG-VD to start in 2016

MaLDIveS : Bioindication by sequencing



2. Machine Learning on Biological Sequences

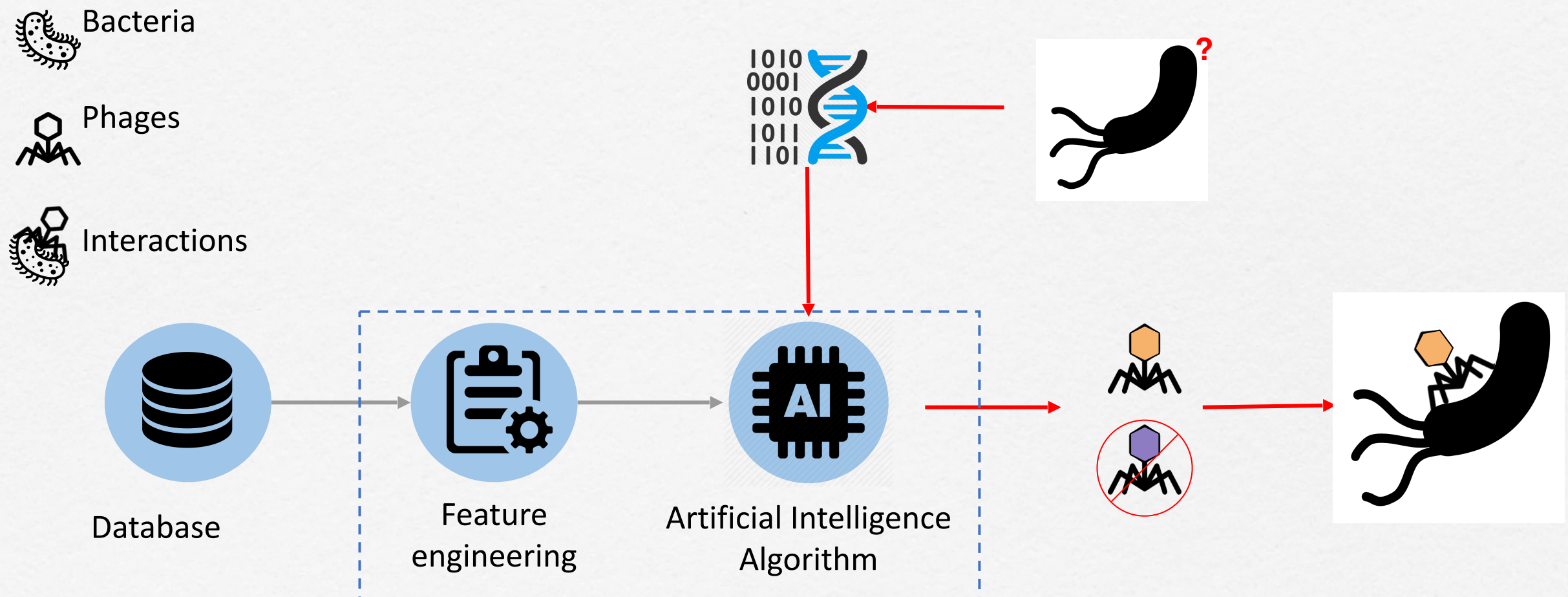
Feature-based ML and feature-free DL methodologies

Predictive modelling using biological sequence data, including genomics and proteomics, with a focus on sequence-based patterns for inference and classification tasks.

Particular interest in genome-based phage-bacteria interaction prediction.

INPHINITY: Interaction prediction from genomics data

Paradigm shift in phage therapy against multi-resistant bacteria

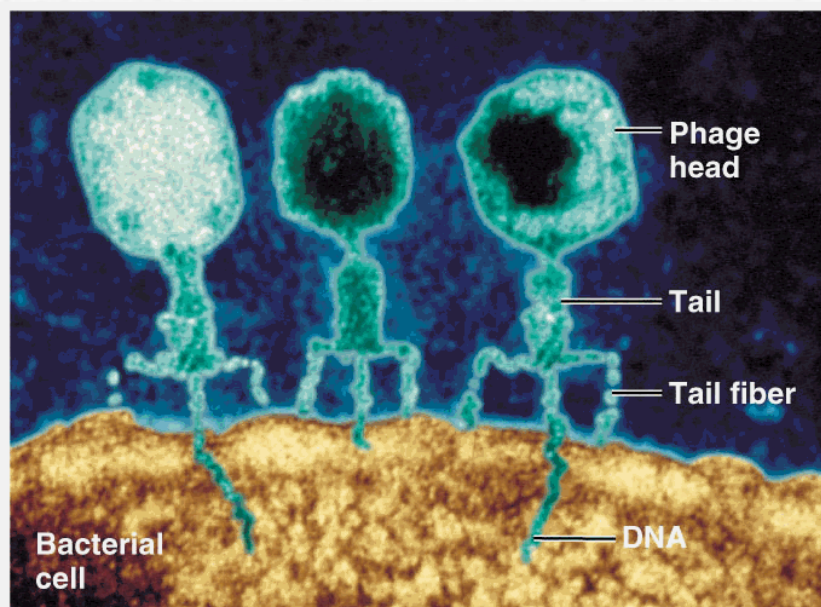


INPHINITY: Feature Engineering

PPI: Protein-protein interactions

“As in all organisms, **protein-protein interactions (PPIs)** play a fundamental role in phage biology. PPIs are important in all virus-related processes...”

Häuser et al. “Bacteriophage protein–protein interactions,” in Advances in Virus Research. Elsevier, 2012, pp. 219–298.



3575 Proteins in average



86 Proteins in average



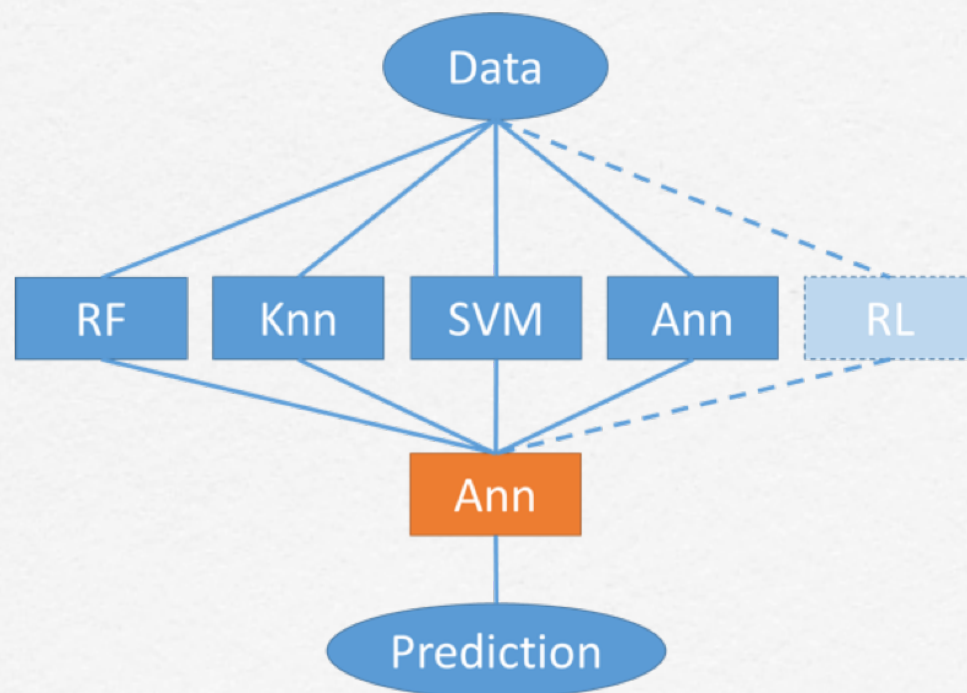
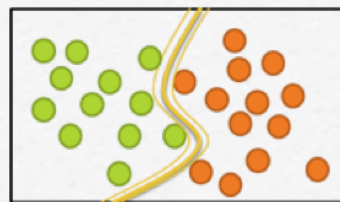
>300'000 PPIs per pair

PPI-based features:

- Domain-interaction scores: 1 score per PPI
- Chemical composition data: 27 features per protein

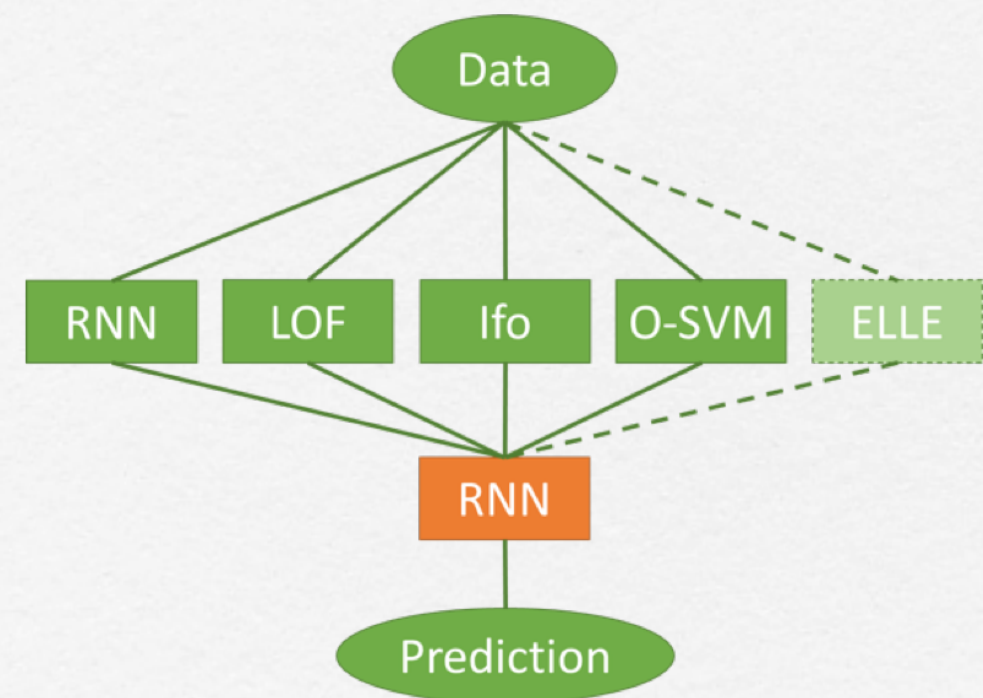
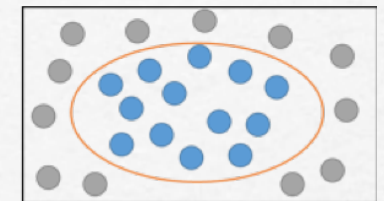
INPHINITY: Machine learning approaches

Ensemble learning techniques



Supervised

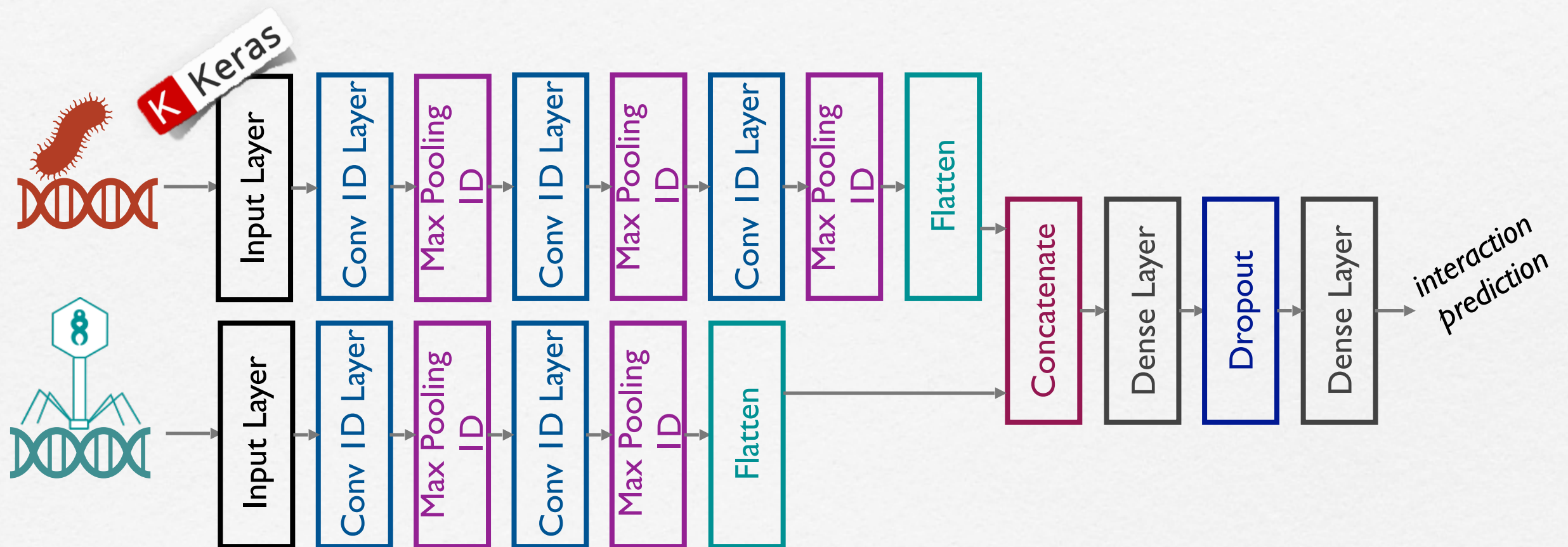
One-Class learning techniques



Unsupervised

INPHINITY+: Deep learning on genomics data

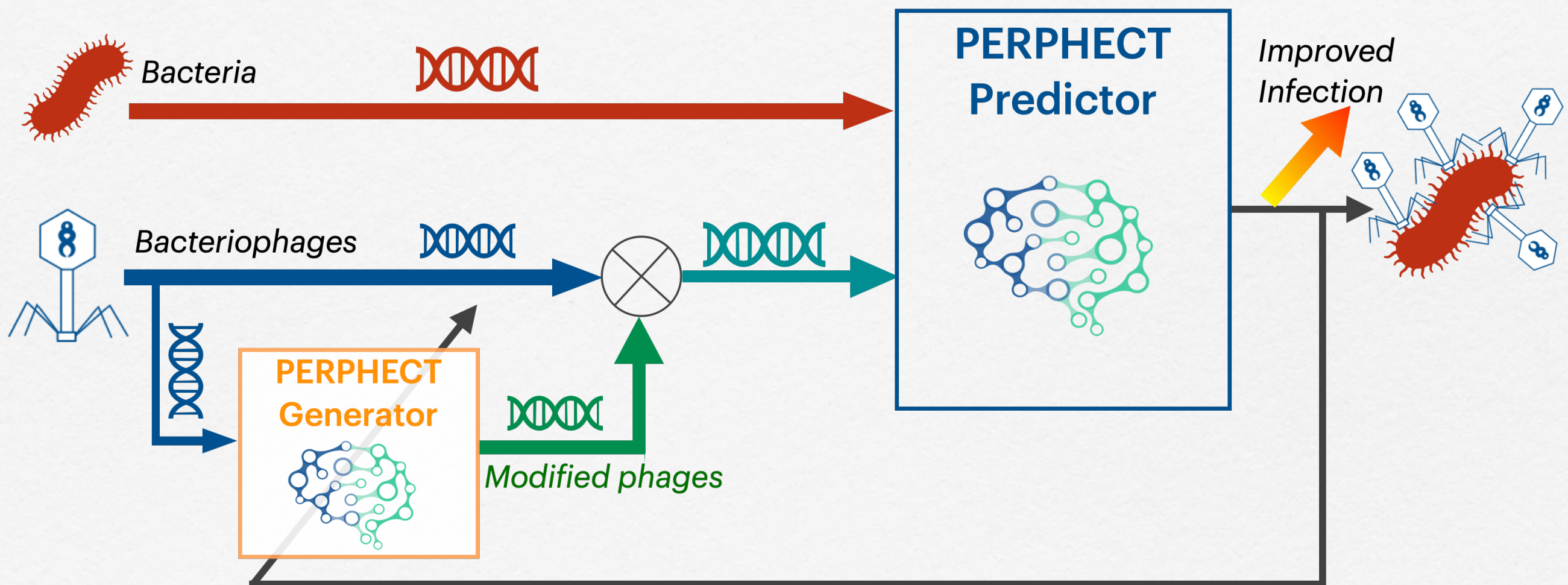
Feature-free phage-bacteria interaction therapy against multi-resistant bacteria



Multi-input **1D Convolutional Neural Network** to process genome sequences

PERPHECT: Engineering viruses to fight bacteria

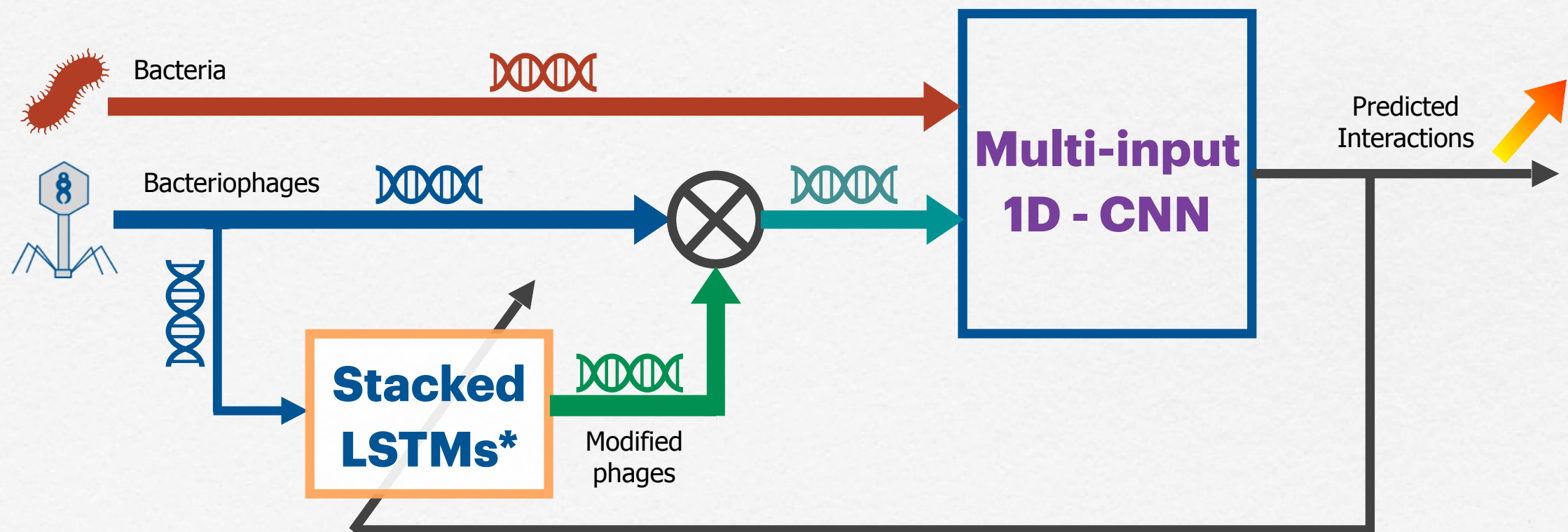
Deep p Generative Networks for Bacteriphage Genetic Editing



PERPHECT: Engineering viruses to fight bacteria

Deep Generative Networks for Bacteriophage Genetic Edition

- ❖ Deep Generative Models for Phage Genetic Edition
- ❖ Composed of a predictor and a sequence generator



* Explored alternatives: GANs, Autoencoders, Evolutionary algorithms, Transformers

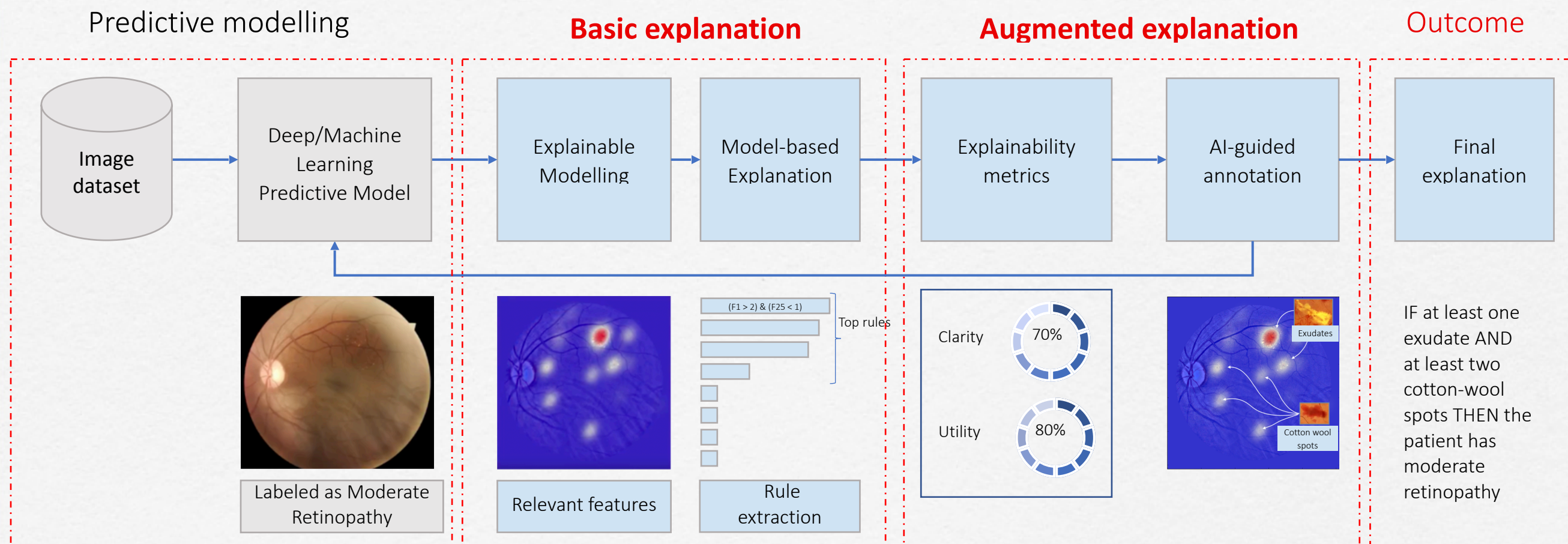
3. Explaining AI-based diagnostic decisions

Building explanation models and promoting model interpretability

Methods and tools for Explainable AI (XAI), spanning from inherently interpretable ML algorithms (such as Fuzzy CoCo) to the extraction of rules and domain knowledge from black-box predictive models.

These XAI approaches are primarily applied to life science problems, including medical diagnostic decision support and the elucidation of mechanisms of action from multi-omics data.

Explaining AI-based diagnostic decisions

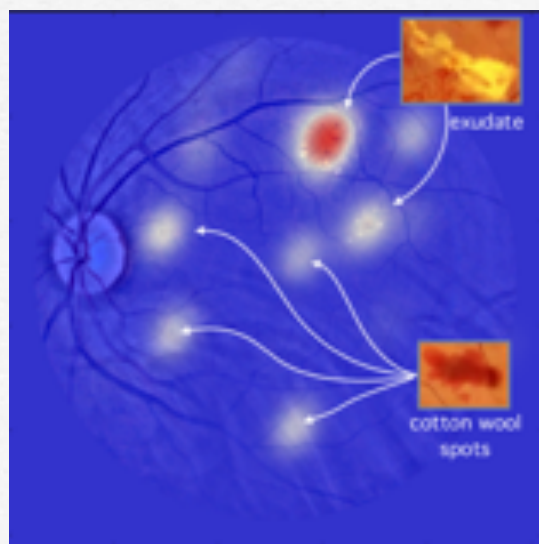


D-Rex: Explaining AI-based decisions

Tuberculosis

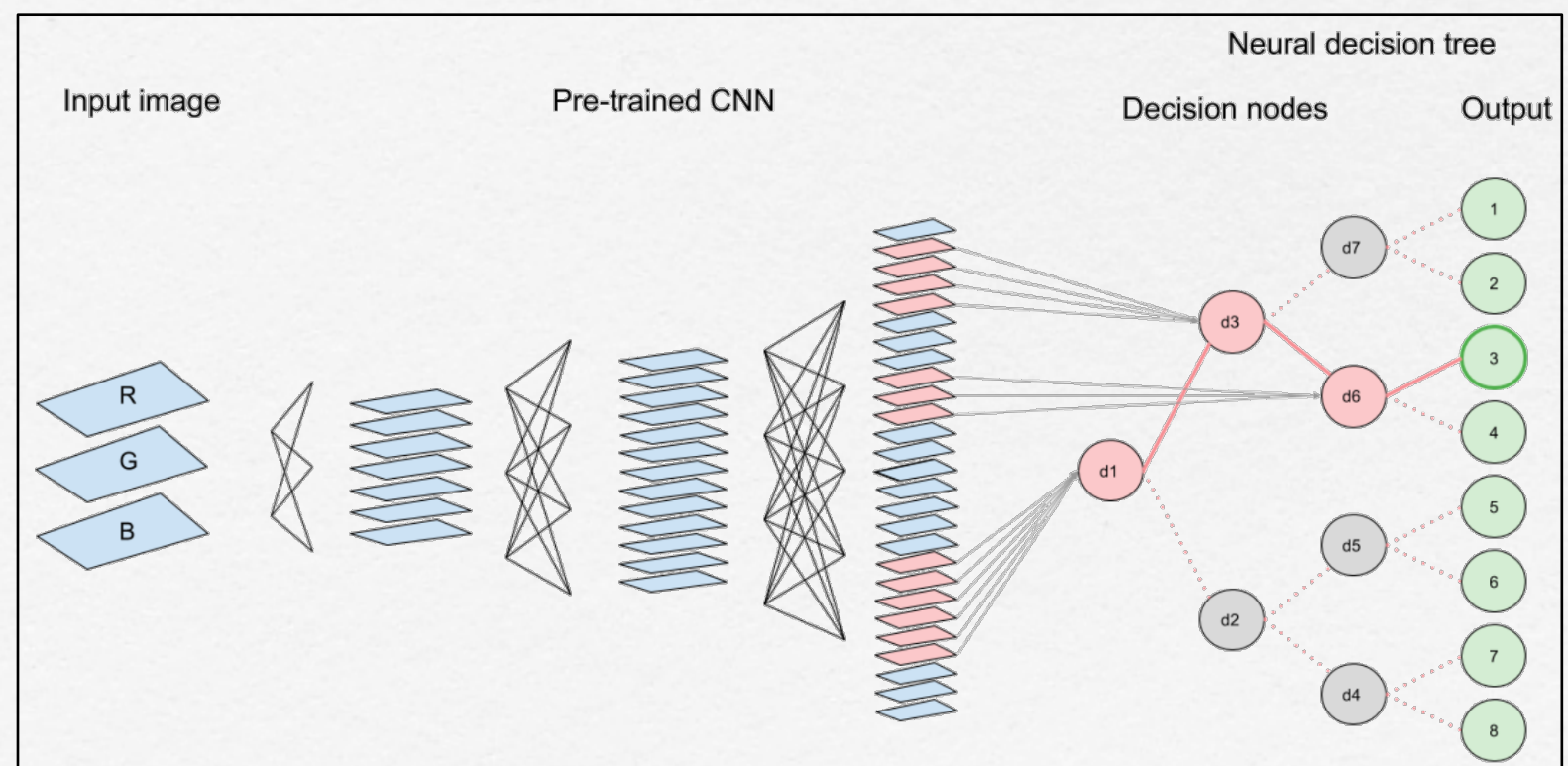


Diabetic retinopathy



Input highlighting (ROI)

Rule extraction from deep
neural networks



HES-XPLAIN

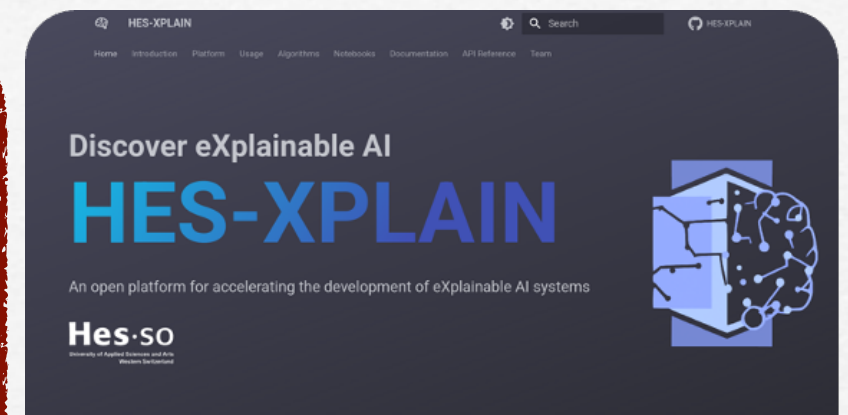
An open platform for accelerating the development of eXplainable AI systems



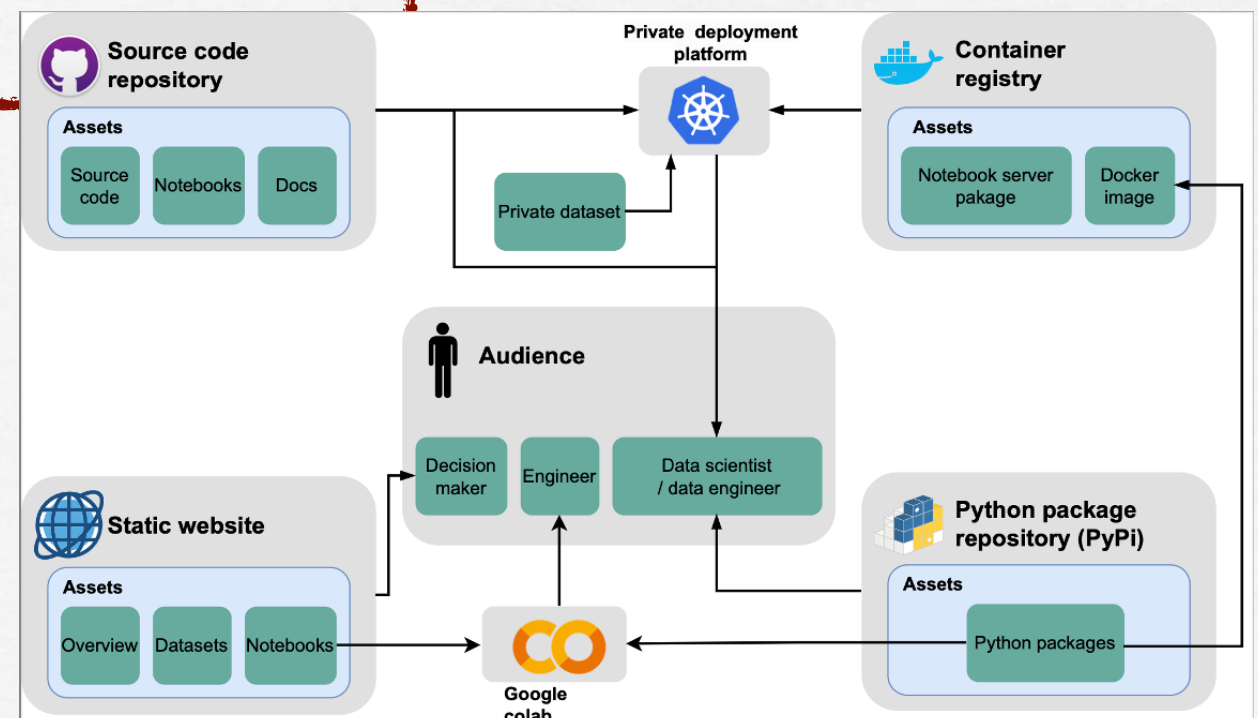
class = « axe throwing »



class = « tug of war »



<https://hes-xplain.github.io/>





IMPTOX

-

EU PROJECT



Understanding the Complex Role of

Micro- and Nanoplastics

combined with

**Environmental
Contaminants**

on Human Health

Focus:

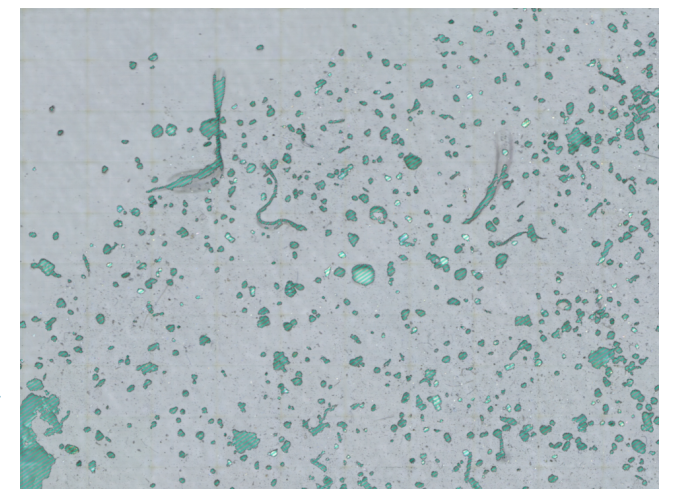
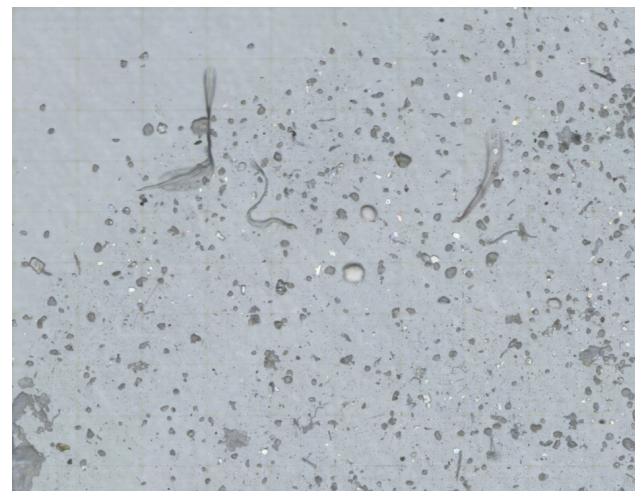
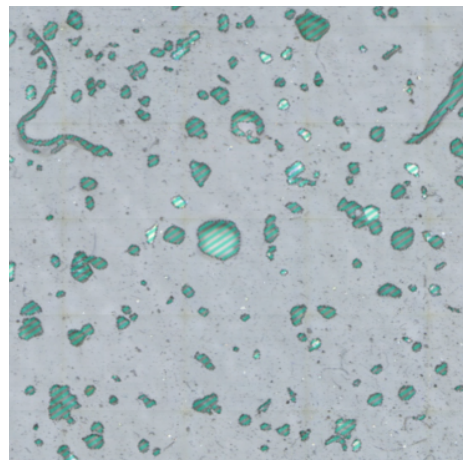
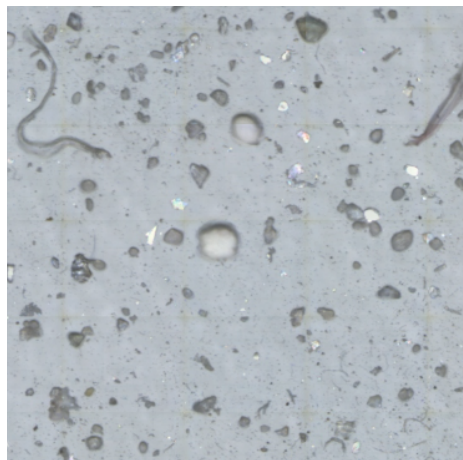
Allergy and Asthma



- Our task: Automatizing MNP recognition through
 - Machine and deep-learning to detect MNPs in μ FTIR images (combined microscopic and spectral data)

**HE^{VD}
IG**

Hes·SO
Haute Ecole Spécialisée
de Suisse occidentale



Research on micro- and nanoplastics and their effects on human health

Context

Plastic particles are ubiquitous (water, atmosphere, food...) and abundant. An estimated 5.25 trillion plastic particles are circulating in the oceans

Aim

Studying the impact of micro- and nanoplastics (MNPs) on food safety and human health. The project focuses on the interaction between the immune system and MNPs.

Relevance

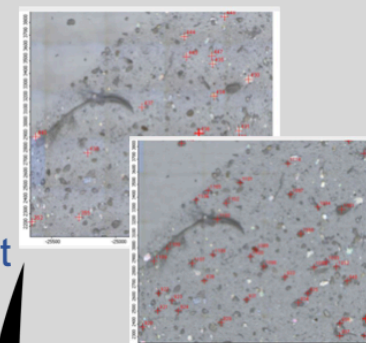
- AI-powered tools improve micro-FTIR analysis,
 - reduce bias
 - preserve manual control
 - Accelerate microparticle detection.

Contact: Prof. Dr Carlos Andrés Peña – carlos.pena@heig-vd.ch



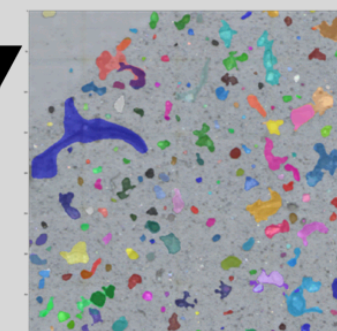
Particle
Selector

Biased and inefficient
target identification



Particle
Detector

Transfer-learning
approach for an
improved detection
and data acquisition



Project funded by
H2020 (EU)