

**Ph.D. in Information Technology
Thesis Defense**

October 16th, 2026

at 2:00 pm

Nuova Sala Conferenze – building 20A

Sofia Mongardi – XXXVIII Cycle

Improving Interpretable Machine Learning in Genomics and Precision Medicine for Gene Characterization, Spatial Analysis, and Disease Risk Prediction

Supervisor: Prof. Marco Masseroli

Abstract:

Recent advances in high-throughput biological and medical technologies have led to the generation of large and complex biological and biomedical datasets. These developments have accelerated the adoption of artificial intelligence and machine learning approaches in biology and medicine, enabling applications such as disease detection, classification, and risk prediction with the ultimate goal of supporting precision medicine. However, despite the increasing availability of such data, their analysis remains challenging due to high dimensionality, heterogeneity, noise, limited sample sizes, missing data, and interpretability. Consequently, there is a pressing need for the development of robust computational approaches that not only provide accurate predictions but also ensure biological and clinical interpretability. To address these challenges, this thesis focuses on the development and application of computational approaches that combine predictive accuracy with interpretability, leveraging prior biological knowledge, multimodal data, and additional relevant information. By integrating knowledge from curated biological databases, morphological and spatial information, and clinical or lifestyle features, these methods aim to enhance the reliability, robustness, and interpretability of models applied to high-dimensional biomedical data.

A key focus of this work is on enhancing feature selection and predictive modeling in transcriptomics datasets through the development of computational approaches that integrate prior biological knowledge into their modeling frameworks. Using information on gene functions, pathways, and molecular interactions, such approaches can enhance the interpretability of predictive models, guiding the identification of features that are not only statistically informative but also biologically meaningful.

Spatially resolved transcriptomics adds another layer of complexity, as the functional organization of tissues depends not only on gene expression but also on the spatial and morphological context surrounding each cell. Graph-based models, including graph neural networks, have emerged as powerful tools for learning representations from such data, enabling tasks such as spatial domain detection, denoising, and prediction of cellular states. Yet, challenges remain in graph construction, as naive spatial adjacency does not always reflect true biological relationships, and incorporating morphological or tissue-level information remains an open problem.

Beyond molecular data, integrating complementary clinical and physiological measurements can further enhance predictive modeling for biomedical applications. For example, cardiovascular risk prediction may benefit from the inclusion of behavioral and lifestyle metrics, such as sleep patterns, which can capture additional dimensions of health not reflected in classical biomarkers and improve the assessment and stratification of disease risk. Evaluating the effectiveness of such data for risk prediction may provide more comprehensive, personalized, and clinically actionable insights.

In conclusion, this thesis advances computational methods for the extraction of biologically and clinically meaningful insights from complex biological and biomedical datasets. It introduces methods that enhance the interpretability of feature selection and predictive models through the integration of prior biological knowledge, develops novel graph-based approaches for spatial transcriptomics that jointly leverage molecular and morphological features, and integrates complementary clinical and physiological information to improve disease risk prediction. Collectively, these contributions aim to advance biomedical data analysis and support the translation of computational findings into actionable insights for computational biology and precision medicine.

PhD Committee

Prof. Rosario Michael Piro, Politecnico di Milano

Prof. Joana Sofia de Pinho Gonçalves, Technische Universiteit Delft

Prof. Paolo Soda, Università Campus Bio-Medico di Roma