

**Title:** Representation of multi-modal big health data for precision cardiovascular risk assessment

**Author:** Andrea Mario Vergani

**Supervisors:** Marco Masseroli, Emanuele Di Angelantonio

**Tutor:** Francesca Ieva

**PhD program coordinator:** Piercesare Secchi

### **Abstract**

The growing availability of multi-modal biobank-scale data offers invaluable opportunities for studying the impacts of omics and medical imaging on biological mechanisms and disease. In this context, the thesis focuses on the relevance of complex health data modalities for personalised cardiovascular risk assessment.

Specifically, in the first part of the thesis, we leveraged the interactions between cardiac magnetic resonance imaging, electrocardiogram, and genetic predisposition data to project such heterogeneous sources in a cross-modal representation and used the latter to predict the future occurrence of cardiovascular events in healthy individuals, thus shedding light on the relevance of multi-modal integration of medical test data for cardiovascular risk definition. In the second part, instead, we explored the role of unconventional phenotypes to predict incident cardiovascular events, particularly focusing on deep representation learning-derived latent factors from cardiac magnetic resonance imaging and single-nucleotide polymorphism data, and also investigating the relevance of domain knowledge to define useful features from exome sequences. Our studies clarified the relationships between the analysed health data modalities in the context of cardiovascular risk definition, and highlighted that novel features may improve disease prediction over more established measures derived from cardiac imaging and genetic data.

Throughout the dissertation, we also focused on the explainability of the proposed methodological frameworks and the interpretation of results, which are key aspects to bringing our research closer to clinical translation.