

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

**October 26th, 2026
at 2:15 pm
Room “Sala Schiavoni” – building 20A**

Michele CRAIGHERO – XXXVIII Cycle

Learning in the Wild: Domain Adaptation and Personalization for Time Series in Distributed Environments

Supervisor: Prof. Giacomo Boracchi

Abstract:

Time series data are pervasive in real-world applications such as healthcare, cybersecurity, finance, and industrial manufacturing, where modeling temporal dynamics is essential for effective monitoring, prediction, and decision making. Despite their importance, time series data pose significant challenges for machine learning solutions due to their limited interpretability, sensitivity to sensor characteristics and changes over time. These challenges are further amplified in user-oriented and biomedical applications, where inter-subject variability, different sensing devices, and evolving physiological conditions lead to complex and heterogeneous data distributions.

These variations in data distributions are commonly known as domain shift, which represents a major obstacle to the deployment of machine learning models in real-world settings. In fact, domain shift leads to substantial performance degradation when models trained on a source domain are tested on a target one. While numerous domain adaptation and personalization techniques have been proposed to mitigate this problem, most existing approaches assume centralized access to data and do not consider the constraints of realistic distributed environments, where labeled data are scarce and data sharing is limited due to privacy and communication requirements.

This thesis investigates the impact of domain shift in distributed learning environments and proposes novel domain adaptation and personalization solutions tailored to time series and biomedical applications. The research follows a two-stage progression, starting from algorithmic innovation and then focusing on practical deployment.

In the first stage, we present a suite of algorithmic solutions to address the dual challenges of domain shift and data scarcity. Firstly, we focus on a labels-at-server Semi-Supervised Federated Learning (SSFL) setting, in which clients continuously collect only unlabeled data while labeled data are available exclusively at a central server. In this context, we propose two unsupervised federated domain adaptation methods: SemiFDA, which mitigates server-client shift through federated covariance alignment, and ClusterSSFDA, a federated clustering strategy designed to address client-client heterogeneity. Furthermore, we introduce K-TCL, a novel kernel-guided pair-mining strategy for Contrastive Learning to contrast the deteriorating impact of class imbalance

and false negatives in time series tasks. We also investigate the security aspect of these distributed systems by proposing BadAvg, a novel time-resilient backdoor attack against Federated Contrastive Learning. This highlights the need to integrate advanced defense mechanisms in such systems.

In the second stage, we bridge the gap between theory and practice through two critical case studies. First, we examine a real-world biomedical application, namely systolic complex detection in Seismocardiogram (SCG) signals. We demonstrate how models trained in controlled laboratory settings fail in unconstrained, daily-life conditions, thus requiring adaptation and personalization strategies for robust deployment. Finally, we establish the hardware feasibility of these concepts by presenting the first implementation of model personalization for Human Activity Recognition (HAR) directly on a resource-constrained STM32 microcontroller. This feasibility study confirms that local fine-tuning is not only computationally possible but essential for robust, decentralized machine learning in the wild.

PhD Committee

Prof. Matteo Matteucci, Politecnico di Milano

Prof. Claudio Gallicchio, Università di Pisa

Prof. Simone Scardapane, Sapienza Università di Roma