

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

June 22, 2022

at 12:00

Room Conferenze "Emilio Gatti"

Edoardo GIACOMELLO – XXXIV Cycle

Addressing Collaborative Machine Learning Challenges in Medical Imaging

Supervisor: Prof. **Daniele Loiacono**

Abstract:

Machine Learning and Deep Learning tools in Medical Imaging are promising approaches to aid physicians and radiologists in performing diagnoses. Machine Learning models that work with imaging data require massive amounts of data. Although many institutes are collaborating to produce publicly available datasets of medical images, the process of data acquisition is severely limited by different challenges. These challenges are mainly related to privacy regulations and the effort of domain experts to assess imaging data quality and produce high-quality ground truth. In turn, the difficulty of managing large datasets of medical imaging translates in a scarcity of data available for research. This Ph.D. thesis studies collaborative machine learning as a methodological approach to overcome the problem of data availability. Collaborative Machine Learning is a vast area of research that includes a set of techniques, such as Distributed Learning and Esembling Methods, to enable multi-centric studies using multiple private datasets. The main idea behind collaborative machine learning is to share knowledge instead of data to overcome potential privacy issues in exchanging sensitive data. However, this approach poses challenges that include data heterogeneity due to the population included in the datasets, and data incompleteness, due to different data acquisition standards and practices among different institutions. This work provides a general taxonomy for classifying the various approaches proposed in the literature and analyzes well-established techniques such as ensemble learning and transfer learning in the context of collaborative machine learning.

Moreover, we analyze more recent contributions based on Federated and Split Learning, comparing their performances according to data heterogeneity and privacy constraints. Our experiments study multiple approaches that exploit ensemble methods, distributed learning, and transfer learning to overcome different challenges, such as data heterogeneity, model heterogeneity, and label heterogeneity using public and private datasets. Finally, we propose our approach to image segmentation based on adversarial networks and generative adversarial networks to study possible approaches to the problem of incomplete medical imaging datasets.

The results are promising, showing that collaborative learning can successfully overcome the issues above. In particular, ensemble learning methods can build a single model from multiple models with different architectures when trained on different data subsets. Moreover, distributed learning approaches proved to be a good design choice when privacy has to be attained, especially in a

context of data heterogeneity. Transfer learning and embedding techniques can enable the training of custom models on smaller private datasets by exploiting the powerful feature extraction modules of Convolutional Neural Networks. Lastly, our approach based on adversarial networks proved to be promising to enable the use of multi-input segmentation models when some of them are missing, thanks to image translation.

PhD Committee

Prof. **Manuel Roveri**, DEIB-Politecnico di Milano

Prof. **Jaume Bacardit**, Newcastle University

Prof. **Giovanni Dimauro**, Universita' degli Studi di Bari