DIPARTIMENTO DI INGEGNERIA
CORSO DI DOTTORATO IN INGEGNERIA INDUSTRIALE E DELL'INFORMAZIONE -
PHD COURSE IN INDUSTRIAL AND INFORMATION ENGINEERING -
37TH CYCLE

Title of the research activity:	**Edge machine learning and signal processing for wireless networked systems**
State of the Art:	We are currently experiencing a new technological revolution, where the massive growth of data traffic, transmitted and generated in wireless and wired telecommunication networks and user equipment's (mobile phones, wearable devices, autonomous vehicles, etc.) is accompanied by a pervasive use of artificial intelligence (AI) and machine learning (ML) algorithms aimed to extract specific, task-oriented, meanings from the data, as well as to enable communication functionalities, or optimizing the communication network performance. In communication systems, the widespread success of (deep) ML approaches, has been firstly envisaged to fulfill classical telecommunications goals, such as channel estimation, detection, coding, traffic shaping, network discovery, and so forth [1]. The motivating idea to resort to data-driven ML algorithms, is that they can outperform classical model-based solutions and designs when the models are not matched to the practical scenarios, thus representing a viable alternative for communication and network reconfiguration in heterogenous scenarios. However, a much more intriguing and challenging framework looks at a deeper interplay of ML (computation) tasks with wireless communication networks that, rather than being only perceived as technological enablers for ML tasks performed by devices connected to the Internet, are conversely integrated with, and optimally managed for, the specific statistical learning task they have to support, with a paradigm known as Edge Machine Learning [2]-[8]. The massive data to be processed nowadays by AI applications, are typically distributed across a cloudified network, or generated in real time by thousands of IoT devices. As well known, this makes impossible to analyze the data in a central fashion both for capacity issues of the network and the prohibitive computational complexity, which in most cases does not even scale linearly with the data-size. To overcome these criticalities, in the last decade a huge research effort has been dedicated to developing both statistical learning algorithms and platforms with distributed and collaborative architectures. However, such algorithmic and architectural solutions are not well suited to situations where the learning result is requested back in a noticeably short time, possibly by the same equipment that generated the data, such as users or devices connected by a wireless network, whose specific nature and constraints are not always taken into account. For example, a challenging situation is when the inference or learning tasks must be performed by mobile or simple devices, which may not have at their disposal either enough energy, or computational capabilities, or both.

	This could be the case of small drones that have to identify or classify objects during their mission or exploit computer vision deep learning algorithms for collision avoidance and pose estimation: if they do not have enough computational or energy capabilities on board, they may ask a server of a (wireless connected) network to perform the task and send back the outcome. The problem would become even more challenging in the presence of a swarm of drones, or a network of IoT devices, that need to cooperate for a common mission, or learning task. The drones must rely on a secure, fast, and low-latency communication network, to act as a seamless networked system and possibly employ what they learn within their control algorithms to stabilize the flight and to perform the mission. As another example, a plethora of IoT devices may have extremely limited batteries and computational capabilities, although possibly a less stringent end-to-end delay for their (possibly co-operative) learning task. In this framework, the recently introduced 5G networks offer amazing wireless communications capabilities, by enabling different new services using the same communication platform, which are characterized by a significant communication performance enhancement, by means of a) Ultra-reliable and low-latency communications (URLLC) b) Enhanced mobile broadband (EMBB) c) Massive Machine Type Communications (eMTC). d) Pervasive use of cloud capabilities and the Wireless Network EDGE [4]. This last characteristic of 5G networks, differently from classical cloud-based architectures, moves the computational services to the network edge, i.e., closer to the end-user and the wireless connected devices. This is crucial, together with the other three pillars of 5G networks, to enable a high-data-rate learning framework where the computational burden of a (mobile) equipment (drone, sensor, IoT device, mobile phone, etc.) could be demanded (off-loaded) to the network, still preserving acceptable end-to-end delay. This Edge Computing vision, formerly known as Fog Computing [9], has been also recently standardized by ETSI under the name of Multi-Access Edge Computing (MEC).
Short description and objectives of the research activity:	In the described framework the research activity will be focused on a holistic view of the learning goals and algorithms, the network resources (bandwidth, rate, power), and the edge computational capabilities. More specifically the research activity, that will be positioned in the hot topic of Edge Machine Learning, will focus on: a) exploring the trade-offs of learning accuracy and resource allocation for dynamic training and inference of (single agent) machine learning tasks, which can be possibly split between the end-user and the edge of the wireless network, rather than being simply off-loaded to the edge. The goal is to explore trade-offs between energy consumption, delay and learning accuracy, by proper allocation of the available resources (radio, data-rate, scheduling, edge computations, etc.) as proposed in some seminal works for classical computation off-loading [10].

	b) data compression techniques to reduce the wireless rate transmission while preserving the accuracy of the learning tasks. For instance, the research activity will investigate approaches inspired by the information theoretic bottleneck method [11] that has been largely shown to play a fundamental role in the design and performance evaluation of deep machine learning algorithms [12][13]. c) Collaborative machine learning at the edges of 5G wireless networks and beyond, where the existing literature on wireless distributed estimation [14] [15] and Federated Learning [16] [17] will be casted and extended to the proposed holistic view for the joint network & learning design and optimization, which will be the core of the research activity. The said research objectives, will exploit an interdisciplinary approach that finds its theoretical roots on statistical learning and signal processing, wireless communications, information theory, network science, distributed and stochastic optimization, coupled with deep machine learning architectures, such as (graph) convolutional neural networks.
Bibliography:	[1] D. Gündüz, P. de Kerret, N. D. Sidiropoulos, D. Gesbert, C. R. Murthy and M. van der Schaar, "Machine Learning in the Air," in IEEE Journal on Selected Areas in Communications, vol. 37, no. 10, pp. 2184-2199, Oct. 2019. [2] S. Wang et al., "When Edge Meets Learning: Adaptive Control for Resource-Constrained Distributed Machine Learning," IEEE INFOCOM 2018 - IEEE Conference on Computer Communications, 2018, pp. 63-71, [3] T. K. Rodrigues, K. Suto, H. Nishiyama, J. Liu and N. Kato, "Machine Learning Meets Computation and Communication Control in Evolving Edge and Cloud: Challenges and Future Perspective," in IEEE Communications Surveys & Tutorials, vol. 22, no. 1, pp. 38-67, 2020. [4] T. Taleb, K. Samdanis, B. Mada, H. Flinck, S. Dutta and D. Sabella, "On Multi-Access Edge Computing: A Survey of the Emerging 5G Network Edge Cloud Architecture and Orchestration," in IEEE Communications Surveys & Tutorials, vol. 19, no. 3, pp. 1657-1681, 2017. [5] G. Zhu, D. Liu, Y. Du, C. You, J. Zhang and K. Huang, "Toward an Intelligent Edge: Wireless Communication Meets Machine Learning," in IEEE Communications Magazine, vol. 58, no. 1, pp. 19-25, January 2020. [6] J. Chen and X. Ran, ‘‘Deep learning with edge computing: A review,’’ Proc.IEEE, vol. 107, no. 8, pp. 1655–1674, Aug. 2019. [7] X. Wang, Y. Han, V. C. M. Leung, D. Niyato, X. Yan, and X. Chen, ‘‘Convergence of edge computing and deep learning: A comprehensive survey,’’ IEEE Commun. Surveys Tuts., vol. 22, no. 2, pp. 869–904, 2020. [8] M. Mohammadi Amiri and D. Gündüz, "Machine Learning at the Wireless Edge: Distributed Stochastic Gradient Descent Over-the-Air," IEEE Transactions on Signal Processing, vol. 68, pp. 2155-2169, 2020. [9] M. Mukherjee, L. Shu, and D. Wang, "Survey of Fog Computing: Fundamental, Network Applications, and

	Research Challenges," in IEEE Communications Surveys & Tutorials, vol. 20, no. 3, pp. 1826-1857, 2018. [10] M. Merluzzi, P. D. Lorenzo and S. Barbarossa, "Wireless Edge Machine Learning: Resource Allocation and Trade-Offs," in IEEE Access, vol. 9, pp. 45377-45398, 2021 [11] N. Tishby, F. C. Pereira, and W. Bialek, ”The information bottleneck method,” arXiv preprint physics/0004057, 2000. [12] O. Shamir, S. Sabato, N. Tishby, “Learning and generalization with the information bottleneck”, Theoretical Computer Science, Vol. 411, n. 29–30, 2010. [13] M. Saxe, et. al, “On the information bottleneck theory of deep learning,” Journal of Statistical Mechanics: Theory and Experiment, IOP Publishing, vol. 2019, n.12, 2019. [14] M. M. Amiria, T. M. Dumanb, D. Gündüzc, S. R. Kulkarni and H. Vincent Poor, "Collaborative Machine Learning at the Wireless Edge with Blind Transmitters," in IEEE Transactions on Wireless Communications, March 2021. [15] C. Battiloro, P. Di Lorenzo, P. Banelli and S. Barbarossa, "Dynamic Resource Optimization for Decentralized Estimation in Energy Harvesting IoT Networks," in IEEE Internet of Things Journal, vol. 8, no. 10, pp. 8530-8542, 15 May15, 2021, [16] T. Li, A. K. Sahu, A. Talwalkar, and V. Smith, ‘‘Federated learning: Challenges, methods, and future directions,’’ IEEE Signal Process. Mag., vol. 37, no. 3, pp. 50–60, May 2020. [17] J. Mills, J. Hu and G. Min, "Communication-Efficient Federated Learning for Wireless Edge Intelligence in IoT," in IEEE Internet of Things Journal, vol. 7, no. 7, pp. 5986-5994, July 2020.
Scientific coordinator (s)	Prof. Paolo Banelli
Contact (s)	paolo.banelli@unipg.it +39-075-585 3626