

Active TinyMLOps

Hossein (Maziar) Aqasizade

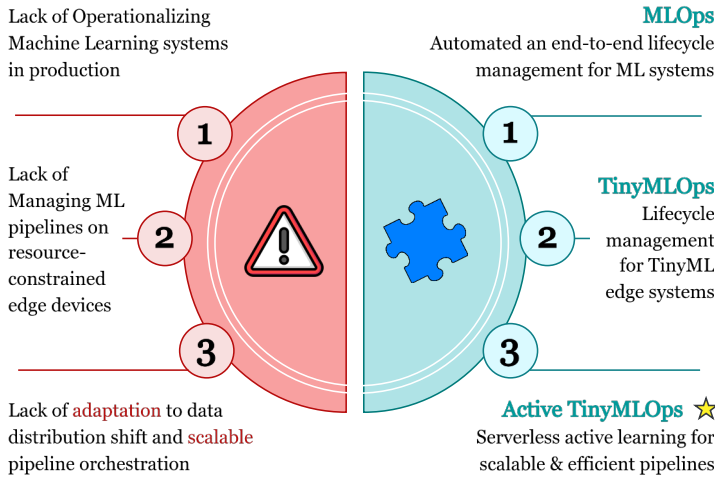
Supervised by Prof. Kasim Sinan Yildirim, Prof. Massimo Vecchio, Dr. Mattia Antonini

1 Problem & Challenges

Applying AI in real-world, decentralized environments is challenging due to **resource constraints**, **dynamic data shifts**, and the need for **continuous adaptation**. Moreover, labeled data is often limited or costly to obtain, making it difficult to sustain model performance over time. Addressing these challenges requires data-efficient learning and scalable orchestration mechanisms for **adaptive pipelines**.

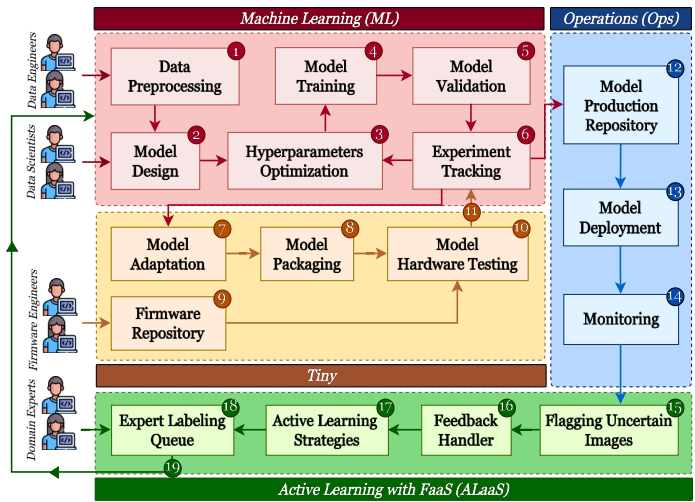
2 Motivation for Active TinyMLOps

Active TinyMLOps evolves from MLOps and TinyMLOps to bridge the gaps in operationalization, pipeline management, and adaptive scalability for resource-constrained devices.



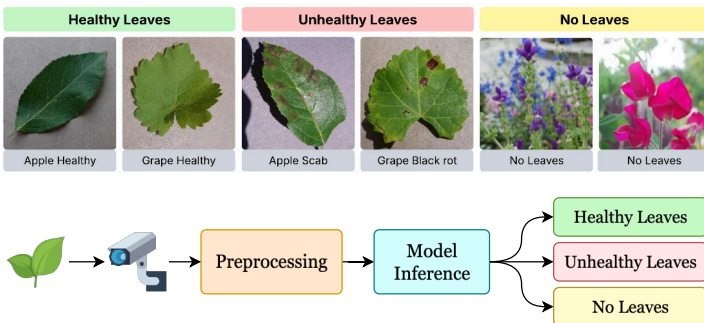
3 Active TinyMLOps Pipeline

We present **Active TinyMLOps**, an end-to-end framework that integrates **TinyMLOps** with **active learning** to enable data-efficient, continuously adaptive ML pipelines.



4 Use case: Leaf Disease Detection [1]

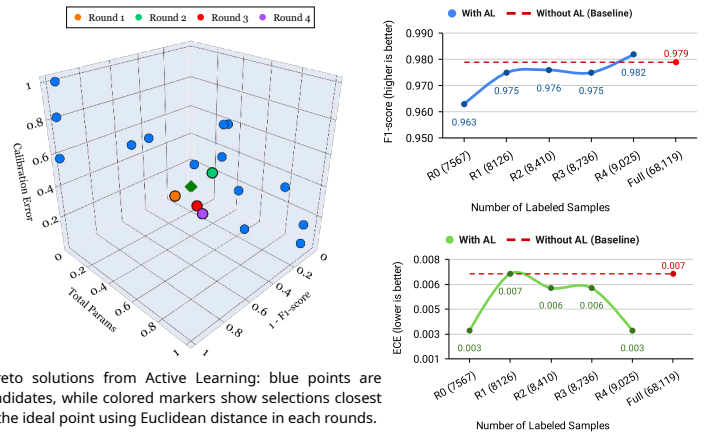
The proposed framework was evaluated on a plant health classification task using the augmented PlantVillage dataset in a simulated incremental setting.



[1] Aqasizade H., et al. "A TinyMLOps-Based Edge AI Approach for Early Detection of Emerging Plant Diseases" 2025 IEEE Int. Workshop on Metrology for Industry 4.0 & IoT (MetroInd4.0 & IoT).

5 Preliminary Results

The preliminary results indicate that the proposed workflow in a simulated setting achieved comparable performance to full-dataset (baseline) training using **87% fewer labeled samples**.



6 Expected Outcomes & Future Work

An end-to-end **Active TinyMLOps** loop is illustrated, spanning model optimization, deployment, continuous monitoring, and adaptation. The system efficiently selects informative samples to handle data-distribution shift and to support safe over-the-air model updates. **Serverless (FaaS)** event-driven orchestration will be explored as future work to enhance scalability and enable event-triggered execution.