COMPUTER AND CONTROL ENGINEERING

ENI Young Talent Award - Enhancing Agricultural Resilience: Deep Learning-Based Plant Disease Detection and Prediction

Funded By	ENI S.P.A. [P.iva/CF:00905811006]

Supervisor	CERQUITELLI TANIA - tania.cerquitelli@polito.it

Contact	CERQUITELLI TANIA - tania.cerquitelli@polito.it GIOVANNI BOSCHIN MORRA LIA - lia.morra@polito.it

Context of the research activity	Efficient and timely plant disease detection is vital for guaranteeing food security and climate change adaptation in developing nations. Despite progress in AI, challenges persist due to limited, biased datasets and disparities between lab and open-field conditions. Practical constraints demand efficient solutions capable of running on low-resource computational platforms. This PhD aims to craft a computational-efficient solution for disease severity assessment, adapting advanced Neural Architectural Search to smaller datasets, and creating adaptable models for agricultural sustainability in developing countries. The position is reserved to candidates who have participated in the selection of the competition "Debut in Research: Young Talents from Africa" of the year 2023.

	State of the art and research gaps Early detection of plant diseases is crucial in averting farm devastation and mitigating risks to food security. While naked-eye detection methods pose considerable complexity, primarily due to inherent drawbacks such as challenges in distinguishing similarities among diseases, identifying their causes, and discerning various disease stages in plants, experts including plant pathologists and farmers face significant hurdles in this detection process. To address this issue, researchers have delved into the realm of computer vision integrated with machine learning, yielding promising outcomes. Notably, current disease phenotyping methods based on RGB images can be classified into two different problem formulations: (i) Disease detection: defined as an absence or presence of disease and (ii) Disease quantification or severity prediction, defined as the extent to which individual leaf has been affected. While disease detection has been extensively tackled in the literature, considerably less effort has been developed to disease quantification and severity prediction.

Objectives	Despite achieving promising results in laboratory studies, there are several hurdles that prevent the widespread application of computer vision techniques for plant disease management. These challenges primarily revolve around dataset issues, complexities in image backgrounds, practical deployment concerns, and model optimization limitations. Firstly, dataset insufficiency poses a significant hurdle. While publicly available datasets exist, the lack of severity annotated datasets crucial for plant disease severity research remains evident. Annotating images for severity is a tedious process, demanding both efficiency and accuracy. Addressing this, automating annotation with advanced software and employing semi-supervised training methods emerge as potential solutions, yet the accuracy challenge persists, both in manual and software-driven annotation. Another substantial challenge is dataset imbalance and bias, impacting model generalizability and potentially leading to overestimated performance estimates. Complexities arising from diverse image backgrounds, especially in natural environments, pose additional obstacles. These complexities include misclassification due to resemblances between ground stains and disease symptoms, reflections in natural lighting, and the simultaneous occurrence of multiple diseases on a single leaf. Publicly available datasets are affected by dataset bias, in terms of background, acquisition modalities and geographical distribution, besides being often acquired in laboratory conditions. Practical deployment of such models demands considerations of computing resources and model size optimization. Cloud infrastructures are needed to run large models, but are expensive to operate, require high-bandwidth connectivity and may pose security concerns. Solutions operating on low-resource, mobile devices would facilitate adoption at large, but require the design of ad-hoc networks that can balance requirements in terms of accuracy, fine-grained categorization and computational requirements. Research objectives The present PhD proposal aims at designing a mature, robust, and frugal deep learning-based solution for the automated assessment of plant disease severity, specifically designed to meet the needs and requirements of developing countries. The research objectives of this proposal are three-fold: - RQ1: Enhance the accuracy and efficiency of plant disease severity prediction by leveraging advanced methodologies like Neural Architectural Search (NAS) The focus here is to not only improve prediction accuracy but also minimize the computational costs associated by leverating NAS and Hyper-Parameter Optimization (HPO). This objective aims to streamline the prediction process while maintaining or improving model performance. - RQ2: Design innovative methodologies for NAS and HPO on real-life, small datasets: This objective aims to introduce novel methodologies to improve HPO and NAS on small scale datasets, e.g., by taking into account the uncertainty induced by the smaller scale of the dataset, or by leveraging multiple data sources – both labeled and unlabeled - in the optimization process. - RQ3: Contribute to the creation of robust and adaptable models designed explicitly for agricultural applications. This involves, on one hand, to identify the sources of potential bias in dataset acquisition, and to exploit this domain knowledge to design more robust datasets and/or trained models, enhancing

	the model's resilience to handle diverse datasets and adapt effectively to varying conditions within agricultural environments. The objective is to develop models that demonstrate adaptability and reliability in different agricultural scenarios. Research plan The activities will be carried out in three phases, roughly corresponding to the three years. In phase 1, the PhD candidate will start with a survey of the relevant literature, identify existing challenges and limitations of models that predict plant disease severity, and clearly define the specific problems that the study aims to address within the context of advanced NAS. The candidate will also investigate available datasets, delve into potential sources of dataset bias, and acquire relevant datasets while minimizing sources of biases. A baseline implementation is expected within the first year. In Phase 2, the goal is to integrate Hyperparameter Algorithms into Neural Architecture Search (NAS) for optimizing plant disease severity prediction models, experimenting with diverse strategies for improved efficiency, and ensuring compatibility with low-resource computational infrastructures. Phase 3 focuses on rigorously evaluating the developed model using tailored datasets, comparing its performance with existing methods, documenting and analyzing results, and emphasizing advancements through NAS integration. Results will be disseminated at conferences and international journals targeting both foundational and applied data science and computer vision. Conferences include International Conference on Computer Vision (ICCV), European Conference on Computer Vision (ECCV), European Conference on Machine Learning and Knowledge Discovery (ECML-PKDD) and IEEE International Conference on Data Science and Advanced Analytics (DSAA). Targeted journals include IEEE Transactions on Image Processing, IEEE Transactions on Neural Networks and Learning Systems, IEEE Internet of Things Journal, Pattern Recognition, Expert Systems with Applications, Machine Vision and Applications.

Skills and competencies for the development of the activity	Good knowledge of machine learning and deep learning principles Proficiency in DL frameworks (e.g., TensorFlow, PyTorch) Familiarity with image processing techniques Excellent data analysis and interpretation skills