PhD Thesis Proposal

Title: Low-Cost Multimodal AI for Crop Growth Monitoring and Decision Support in Smallholder Farms

Summary

This PhD research aims to design and evaluate a low-cost multimodal artificial intelligence framework [1] for smallholder agriculture. The proposed system will combine computer vision, farmer-provided text input, and LLM-generated crop descriptions to monitor crop growth, detect early signs of agricultural problems, and support farmer decision-making [2].

Recent advances in multimodal learning and vision-language models have demonstrated strong capabilities in combining visual and textual information. However, their application in real-world agricultural contexts remains limited due to high computational requirements, lack of domain adaptation, and the constraints of smallholder farming environments. This project aims to bridge this gap by developing a lightweight and scalable multimodal AI solution adapted to low-resource settings[3].

Images captured through smartphones or low-cost cameras will be used to analyze crop conditions such as plant growth stage, leaf color, disease symptoms, pest damage, weed presence, water stress, and nutrient deficiency[4]. These visual observations will be combined with heterogeneous textual data, such as farmer observations, local field knowledge, crop history, weather conditions, and treatment records. A key scientific challenge lies in effectively fusing these multimodal sources to produce robust and context-aware predictions under real-world constraints[5].

In addition, large language models or vision-language models will generate automatic descriptions of visual crop conditions, helping transform image-based observations into understandable explanations [6]. This component aims to improve transparency, usability, and trust in AI-based systems, particularly for non-expert users. The research will investigate efficient adaptation strategies for these models in low-data agricultural contexts.

In addition to crop monitoring, the project will develop a multimodal decision-support pipeline integrating detection, reasoning, and recommendation [7]. The system will identify crop anomalies and suggest appropriate interventions (e.g., irrigation, pest control, nutrient management), while addressing challenges related to reliability, uncertainty, and user trust.

The expected contribution is an accessible and scalable AI-based decision-support framework that enables smallholder farmers to track crop development over time, detect problems at an early stage, and receive simple guidance adapted to resource-limited agricultural contexts. Collaboration with some partners could support field data collection, agronomic expertise, validation with farmers, and the evaluation of the proposed AI framework in real agricultural conditions[8].

Objectives:

The main objective of this PhD is to develop a low-cost multimodal AI framework for crop monitoring and decision support in smallholder farming. This goal is structured into the following objectives:

1. **Multimodal Data Fusion**: Develop methods to effectively integrate visual crop data and heterogeneous textual information for robust crop analysis.
2. **Low-Cost and Robust Model Adaptation：** Develop computer vision models for low-quality data and investigate efficient adaptation of foundation models to low-resource agricultural settings.
3. **Multimodal Reasoning with LLMs:** Investigate the use of LLMs and vision-language models to generate semantic descriptions and explanations of crop conditions.
4. **Decision Support System**:Build a pipeline capable of detecting crop issues and recommending appropriate agricultural interventions.
5. **Real-World Validation**: Evaluate the proposed framework through field experiments, focusing on robustness, usability, and practical impact.

Supervisors:

Main **supervisor**: Aladine Chetouani <aladine.chetouani@univ-paris13.fr>
Co-supervisor: Zuheng Ming <zuheng.ming@univ-paris13.fr>

References:

[1] BALTRUŠAITIS, Tadas, AHUJA, Chaitanya, et MORENCY, Louis-Philippe. Multimodal machine learning: A survey and taxonomy. IEEE transactions on pattern analysis and machine intelligence, 2018, vol. 41, no 2, p. 423-443.
[2] KAMILARIS, Andreas et PRENAFETA-BOLDÚ, Francesc X. Deep learning in agriculture: A survey. Computers and electronics in agriculture, 2018, vol. 147, p. 70-90.
[3] SAPKOTA, Ranjan, QURESHI, Rizwan, HADI, Muhammad Usman, et al. Multi-modal LLMs in agriculture: A comprehensive review. IEEE Transactions on Automation Science and Engineering, 2025.
[4] WANG, Yiqun, WANG, Fahai, CHEN, Wenbai, et al. A large language model for multimodal identification of crop diseases and pests. Scientific Reports, 2025, vol. 15, no 1, p. 21959.
[5] BOUDIAF, Abderrahmene, HUSSAIN, Irfan, et JAVED, Sajid. AgriChat: A Multimodal Large Language Model for Agriculture Image Understanding. arXiv preprint arXiv:2603.16934, 2026.
[6] YANG, Bo, CHEN, Yunkui, FENG, Lanfei, et al. Agrigpt-vl: Agricultural vision-language understanding suite. arXiv preprint arXiv:2510.04002, 2025.
[7] Zhang, M., Xu, Z., Wang, P., Li, R., Wang, L., Liu, Q., Xu, J., Zhang, X. and Wu, S., 2026, May. AgriDoctor: A multimodal intelligent assistant for agriculture. In ICASSP 2026-2026 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 2741-2745).
[8] Wu, H., Du, Z., Zhong, D., Wang, Y. and Tao, C., 2025. FSVLM: A vision-language model for remote sensing farmland segmentation. IEEE Transactions on Geoscience and Remote Sensing, 63, pp.1-13.