Multimodal AI for Early Detection of Abiotic Stress in Crops Using UAV Imagery and Environmental Time-Series Data

Masoumeh Javanbakhat

1. Motivation

Abiotic stress is a major cause of yeild loss worldwide. These stresses disrupt physiological and metabolic processes, leading to reduced growth, development, and productivity by $50\% - 80\%$, depending on the crop and location. Water stress is one of the major abiotic stresses, limiting photosynthesis, nutrient transport, causing alterations at the genome level, and negatively affecting overall crop health. Early detection of those factors, especially water stress detection, is really crucial. Modern technologies like UAV (drone) remote sensing enable rapid, large-scale monitoring of crop water status, helping detect stress symptoms before they become visible to the human eye [1, 2].

2. Problem Statement and Bioeconomic Impact

Abiotic stresses like drought, heat, and nutrient deficiency reduce crop yields and threaten food security. Traditional detection methods are reactive and often miss the early onset of stress, which is critical for effective intervention. This challenge is central to sustainable agriculture (SDG 2), climate resilience (SDG 13), and efficient resource use (SDG 12). Advances in sensors (e.g., multispectral, thermal) and AI now allow early detection of stress indicators before visible symptoms appear [3]. However, there remains a gap in integrating diverse data sources—such as UAV imagery and environmental sensor streams—into unified, interpretable AI systems. We propose a multimodal AI system to detect early stress symptoms by integrating UAV imagery and environmental time-series data.
Research Question: Can a fusion model of UAV imagery and environmental sensor data detect early abiotic stress in crops more accurately and earlier than unimodal approaches? **Hypotheses 1**: Multimodal models outperform unimodal ones in early stress detection. **Hypotheses 2**: Attention-based explainability reveals agronomically relevant patterns.

3. Proposed Method

We propose a smart AI system that can spot early signs of stress in crops by combining images taken from drones (UAVs) with environmental data collected over time. This approach uses machine learning models to analyze the information and help farmers manage their crops more precisely.

3.1 Data Landscape and Preprocessing

This project will use two primary data modalities: *Multispectral UAV imagery* (RGB, near-infrared, thermal, hyperspectral)[3]: captured via drone flights at regular intervals over test fields. *Environmental time-series sensor data*: including soil moisture, temperature, and humidity, recorded continuously from in-situ field sensors. **Data Collection Strategy**: 1- Deploy or simulate UAV flyovers over crop test plots (e.g., maize or wheat) subjected to varying irrigation or nutrient regimes, to capture multispectral and thermal imagery. 2- Install in-field IoT sensors (e.g., soil moisture, temperature, humidity) or use pre-existing datasets from ATB or EU Horizon projects to gather high-resolution environmental time-series data. **Challenges** & **Mitigation**: *Heterogeneity* & *missing data*: Sensor and tabular data may include gaps due to dropouts or asynchrony. For time-series, we will use interpolation, forward/backward filling, and RNN-based imputation models to preserve temporal structure. For tabular data, we will apply regression imputation, classification-based prediction, or Multiple Imputation by Chained Equations (MICE), depending on variable type and data dependencies. *Class imbalance*: use synthetic oversampling (SMOTE) + weighted loss functions. *Data alignment*: Sensor and imagery data will be synchronized using interpolation and spatial registration.

3.2 AI/ML Pipeline Design

Our proposed model combines UAV multispectral imagery and environmental time-series data using a hybrid model:

- **CNN for UAV Imagery**: CNNs effectively extract spatial and spectral features from multispectral images, crucial for identifying subtle crop stress patterns. Their spatial inductive bias suits the detailed texture and shape variations in plant canopies [4, 5, 6].
- **Transformer for Time-Series Data**: Transformers model long-range temporal dependencies in environmental sensor data better than traditional RNNs, thanks to self-attention mechanisms that adaptively focus on relevant time points for stress signals [7, 8, 9].
- **Mid-Level Attention-Based Fusion**: We adopt a mid-level fusion strategy to allow each modality—UAV imagery and environmental time-series data—to first learn domain-specific features through separate encoders (CNN and Transformer, respectively). This enables richer and more relevant feature extraction before integration. Compared to early fusion, this avoids forcing heterogeneous raw data into a shared space prematurely, and compared to late fusion, it preserves cross-modal interactions essential for detecting subtle, multimodal stress signals. Attention layers in the fusion stage further enhance interpretability by dynamically weighting modality contributions [10, 11].
- **Training** & **Validation**: Supervised training with data augmentation (spatial for images, temporal for sensor data) and cross-validation ensures robustness and generalization across diverse conditions.
- **Deployment**: The AI system will support edge inference on resource-limited devices like drones and field IoT sensors, enabling real-time decisions without constant internet. We will use lightweight architectures (e.g., MobileNet, TinyTransformer) combined with pruning and quantization [12] to reduce model size and computation, ensuring low-power, cost-effective deployment. A hybrid approach—with simple tasks on edge and complex processing on a central server—will provide flexibility across different infrastructure settings.

Challenges in Developing Multimodal Models: *Data Heterogeneity*: Different modalities may have different formats and characteristics, making it challenging to align them. *Modality Gap*: Different modalities may have different interpretations, and it can be difficult to bridge the gap between them. *Cross-Modal Interactions*: Establishing effective interactions between different modalities can be challenging. *Computational Complexity*: Processing high-dimensional multimodal data (e.g., hyperspectral + time-series) demands efficient architecture and resource management [13].

4. Explainability and Scientific Insight

We will use Grad-CAM for image saliency, attention visualization for time-series data, and SHAP values for environmental features to provide biologically meaningful explanations. To validate these, we will apply faithfulness tests (e.g., feature removal impact), stability checks, and involve domain experts to ensure explanations align with agronomic knowledge. This iterative feedback will refine the model and improve trustworthiness. The methods are tailored to stakeholders: detailed visualizations for scientists, intuitive heatmaps for farmers, and summary reports for policymakers—ensuring explanations are accessible and actionable for all users.

5. Collaboration and Scalability

The proposed multimodal AI architecture is inherently modular and can be extended to other domains in the bioeconomy that involve heterogeneous data streams—such as combining animal sensor data with video for livestock stress detection, or integrating satellite imagery and soil chemistry for land restoration analysis. This flexibility allows for cross-disciplinary scalability. Collaboration with plant scientists, agronomists, and remote sensing experts will be integral during model development, validation, and field testing. To ensure reproducibility and transferability, we will adhere to open science practices—using version-controlled codebases, standardized data schemas, and model documentation following FAIR (Findable, Accessible, Interoperable, Reusable) principles. Model retraining protocols and modular APIs will enable adaptation to new crops, climates, or data sources with minimal overhead.

6. Real-World Implications and Risks

This AI system enables earlier and more accurate detection of crop stress than traditional scouting, improving resource efficiency and crop yields—supporting SDGs 2, 12 and 13. It benefits farmers through actionable insights, aids researchers in understanding stress patterns, and informs policymakers for climate-resilient strategies. Potential challenges include: Data scarcity—addressed via augmentation and transfer learning; Model bias—mitigated by diverse training data and domain adaptation; Limited farm infrastructure—overcome with lightweight models for edge, enabling offline use. These strategies ensure practical, scalable deployment to advance sustainability in bioeconomy.

References

[1] Jian-Kang Zhu. “Abiotic stress signaling and responses in plants”. In: *Current Opinion in Plant Biology* 5.5 (2002), pp. 375–379.

[2] Artal Smart Agriculture. *Abiotic Stress in Plants*. Accessed: 2025-06-06. Mar. 2025. URL: https://www.artal.net/en/2025/03/abiotic-stress-in-plants/.

[3] AGRI-FOOD.AI. *AI and Sensors for Plant Stress Detection*. Accessed: 2025-06-02. 2025. URL: https://agri-food.ai/news/ai-and-sensors-for-plant-stress-detection/.

[4] Andreas Kamilaris and Francesc X. Prenafeta-Boldú. “Deep learning in agriculture: A survey”. In: *Computers and Electronics in Agriculture* 147 (2018), pp. 70–90. DOI: 10.1016/j.compag.2018.02.016.

[5] Sharada P Mohanty, David P Hughes, and Marcel Salathé. “Using deep learning for image-based plant disease detection”. In: *Frontiers in Plant Science* 7 (2016), p. 1419. DOI: 10.3389/fpls.2016.01419.

[6] Inkyu Sa et al. “WeedNet: Dense semantic weed classification using multispectral images and MAV for smart farming”. In: *IEEE Robotics and Automation Letters* 3.1 (2018), pp. 588–595. DOI: 10.1109/LRA.2017.2774979.

[7] Georgios Zerveas et al. “A Transformer-based Framework for Multivariate Time Series Representation Learning”. In: *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*. ACM. 2021, pp. 2114–2124. URL: https://arxiv.org/abs/2010.02803.

[8] Neo Wu et al. “Deep Transformer Models for Time Series Forecasting: The Influenza Prevalence Case”. In: *arXiv preprint arXiv:2001.08317* (2020). URL: https://arxiv.org/abs/2001.08317.

[9] Shiyang Li et al. “Enhancing the Locality and Breaking the Memory Bottleneck of Transformer on Time Series Forecasting”. In: *Advances in Neural Information Processing Systems*. Vol. 32. 2019. URL: https://arxiv.org/abs/1907.00235.

[10] Tadas Baltrušaitis, Chaitanya Ahuja, and Louis-Philippe Morency. “Multimodal machine learning: A survey and taxonomy”. In: *ACM Computing Surveys (CSUR)* 51.6 (2019), pp. 1–47.

[11] Jiaxin Li et al. “Deep learning in multimodal remote sensing data fusion: A comprehensive review”. In: *International Journal of Applied Earth Observation and Geoinformation* 112 (2022), p. 102926. ISSN: 1569-8432. DOI: https://doi.org/10.1016/j.jag.2022.102926.

[12] Benoit Jacob et al. “Quantization and Training of Neural Networks for Efficient Integer-Arithmetic-Only Inference”. In: *IEEE Conference on Computer Vision and Pattern Recognition* (2018), pp. 2704–2713. DOI: 10.1109/CVPR.2018.00286.

[13] Noël Vouitsis et al. “Data-Efficient Multimodal Fusion on a Single GPU”. In: *arXiv preprint arXiv:2312.10144* (2023). DOI: 10.48550/arXiv.2312.10144.