FINAL REVIEW

Evaluating Multimodal Fusion Strategies for Resilient Agricultural Sensing Systems

Project Category: RESEARCH

Guide Name:
Dr.K Sornalakshmi

Student Details:
Ponnuri Aniruddha [RA2112704010015]
Abhay Shaji Valiyaparambil [RA2112704010006]

ABSTRACT

Precision agriculture increasingly relies on integrating heterogeneous sensor streams-such as time series data from soil sensors (measuring moisture, pH, NPK) and high-resolution crop images-to enhance predictive analytics and crop management. However, fusing these diverse data types presents significant challenges due to differences in temporal resolution, modality, and data quality. In this study, we systematically compare three state-of-the-art multimodal fusion networks-MDFCL, GSIFN, and Perceiver IO-on a unique dataset of synchronized soil sensor readings and field images collected under varying conditions.

Each fusion approach encapsulates a distinct strategy: MDFCL constructs modality-specific graphs and aligns their representations through contrastive learning, promoting robustness to sensor corruption. GSIFN interlaces modality-specific masks within a unified Transformer, facilitating strong cross-modal interactions and resilience to partial image loss. Perceiver IO employs an asymmetric attention bottleneck to efficiently compress heterogeneous inputs, enabling scalable, real-time inference with quasi-linear complexity.

Empirical results show that all three models outperform unimodal baselines in predicting agronomic traits, but differ in robustness and computational efficiency. MDFCL is most stable under sensor dropout, GSIFN excels with incomplete imagery, and Perceiver IO balances accuracy with scalability. These findings provide practical guidance for deploying robust sensor and image fusion solutions in precision agriculture, and offer a template for multimodal data integration in other domains

Introduction

The integration of advanced sensing and imaging technologies has revolutionized agriculture by enabling the collection of rich, diverse data streams, such as soil sensor readings and high-resolution crop images. However, extracting actionable insights from these heterogeneous sources requires sophisticated multimodal data fusion, which can capture complex interactions between soil conditions, environmental factors, and plant health. Traditional single-modality or basic fusion methods often miss subtle but critical relationships, such as early soil nutrient changes that precede visible plant symptoms.

Recent advances in deep learning have unlocked new possibilities for robust and efficient multimodal integration, supporting more accurate crop monitoring and decision-making. This project benchmarks three state-of-the-art fusion strategies-MDFCL, GSIFN, and Perceiver IO-on synchronized time-series sensor and image datasets. Each approach offers a unique method for encoding and combining data, aiming to improve the resilience, scalability, and efficiency of agricultural sensing systems. By systematically evaluating these models, the project provides practical insights for deploying robust, adaptive data fusion solutions in precision agriculture

Literature Survey

S. No	Title *(Name of the journal, author and publication details)*	Methodology *(Provide a Summary of key studies and their findings)*	Identification of gaps and limitations. *(Identify the limitations of the Research Paper)*
1	A Systematic Review on Monitoring and Advanced Control Strategies in Smart Agriculture Authors: Syeda Iqra Hassan, Muhammad Mansoor Alam, Usman Illahi, Mohammed A. Al Ghamdi, Sultan H. Almotiri, Mazliham Mohd Su'ud	• The study systematically reviews advanced control strategies in smart agriculture, including IoT, AI, imaging techniques (e.g., multispectral, hyperspectral cameras), drones, and machine learning. • It emphasizes the importance of automation to address challenges like plant diseases, irrigation management, and nutrient optimization. • Highlights AI-based approaches like CNNs and Random Forests for crop monitoring, disease detection, and yield prediction.	• Limited Geographic Application: Most studies focus on specific regions, with little exploration of diverse agricultural contexts or developing countries. • Scalability Concerns: High costs and technical complexities make scalability and adoption by small-scale farmers challenging. • Technology Integration Issues: Difficulty in harmonizing various technologies like IoT, AI, and drones for seamless operations in real-world scenarios.
2	Integrating Artificial Intelligence and Internet of Things (IoT) for Enhanced Crop Monitoring and Management in Precision Agriculture Authors: Kushagra Sharma, Shiv Kumar Shivandu	• The paper explores the integration of AI and IoT technologies to enhance precision agriculture, emphasizing their role in crop monitoring and management.	• Scalability Issues: The high cost of implementation limits adoption among small and medium-sized farmers.

Literature Survey

		• It highlights advancements in technologies such as high-throughput phenotyping, remote sensing, spectral imaging, and agricultural robotics. • Wireless sensor networks (WSNs) and georeferencing technologies (e.g., GPS, DGPS) are identified as critical for precise, real-time data collection in monitoring soil health and crop growth. • The integration of advanced communication technologies like 5G is emphasized for improving connectivity in precision agriculture.	• Technological Integration Challenges: Difficulty in harmonizing IoT devices, AI models, and other precision agriculture tools in real-world scenarios. • Limited Dataset Diversity: A lack of diverse, localized datasets for different geographical conditions restricts the effectiveness of the solutions. • Connectivity Constraints: Dependence on reliable broadband connectivity, which remains inaccessible in many rural areas, hampers widespread adoption.

Literature Survey

4	**Multimodal Data Fusion via Contrastive Learning (MDFCL)** Gong, X., Liu, M., Liu, Q., Guo, Y., Wang, G. (2025)	• Constructs separate graphs for each modality: sensor readings (nodes encode measurements, edges encode temporal/causal proximity) and image features (nodes as spatial regions/superpixels, similarity edges). • Uses an unsupervised contrastive objective to align graph embeddings by pulling together corresponding node pairs and pushing apart non-matching ones. • Augmentations like edge dropout promote robustness to missing/noisy data. Excels in low-label regimes and maintains accuracy even with single-modality input.	• Inadequate handling of temporal misalignment between high-frequency sensor data and low-frequency images, risking loss of lead–lag relationships. • Scarcity of labeled cross-modal datasets; reliance on curated augmentations may not generalize across crops or conditions. • Insufficient evaluation under noisy or incomplete data, which is critical in real-world agriculture.
5	**Graph-Structured & Interlaced-Masked Fusion Network (GSIFN)** Jin, Y. (2024)	• Processes each modality independently, projecting a 9-dim sensor vector and 512-dim ResNet-18 image embedding into a shared 128-dim space. These tokens form a length-2 sequence for a multi-head self-attention block with an interlaced mask alternating attention between sensor and image tokens. • Parallel LSTM side-tasks reconstruct unimodal inputs to prevent overfitting and redundancy. Achieves high	• Overfitting and modality bias: may over-rely on one dominant modality if the other is degraded or missing. • Temporal alignment issues: may not fully address misalignment between sensor and image data • .Limited generalization: typically validated on

Literature Survey

		accuracy with much lower parameter count and FLOPs than naïve transformers.	restricted datasets, so transferability to diverse crops/conditions is uncertain.
6	**Perceiver IO (Latent Bottleneck Transformer)** Jaegle, A., Borgeaud, S., Caron, M., Brock, A., Ramesh, A., Zisserman, A., Vinyals, O., and Carreira, J. (2021)	• Uses a fixed set of learnable latent vectors that cross-attend to both sensor and image tokens, confining computation to a compact latent space. • This decouples complexity from input length, allowing efficient processing of long time-series data under memory constraints. • The pooled latent array is fed to a lightweight classifier, making the model suitable for edge devices.	• Limited adaptability to edge environments: real-time performance on resource-constrained agricultural hardware is untested. • Insufficient evaluation under noisy/incomplete data, as most tests use clean datasets. • Temporal alignment between modalities is not explicitly addressed.

Research Objective

Objective:

Develop an IoT-based sensing infrastructure that integrates environmental sensors (temperature, humidity, soil moisture, light intensity) and timestamped crop images to generate comprehensive, multimodal datasets.

Rationale:

A robust IoT infrastructure enables real-time, synchronized collection of diverse environmental and crop health data, providing a rich foundation for accurate monitoring and advanced analytics in precision agriculture.

Objective:

Systematically benchmark and compare the predictive performance of three leading fusion models-MDFCL, GSIFN, and Perceiver IO-on nutrient-level classification tasks across multiple crop types.

Rationale:

Rigorous comparison of state-of-the-art fusion models on synchronized sensor and image data helps identify the most accurate and reliable architectures for nutrient monitoring, ensuring optimal model selection for agricultural applications.

Research Objective

Objective:

Analyze each model's ability to integrate heterogeneous sensor and image data, with particular attention to their robustness, accuracy, and efficiency under both ideal and degraded data conditions.

Rationale:

Evaluating models under real-world challenges such as sensor noise, missing data, and corrupted images ensures that the chosen fusion strategy is resilient and effective for practical deployment in agricultural environments.

Objective:

Identify the fusion strategy that offers the best trade-off between predictive reliability and practical deployability for real-time crop monitoring in resource-constrained agricultural settings.

Rationale:

Balancing predictive performance with computational efficiency is essential for real-time, edge-based deployments. This ensures advanced monitoring solutions remain accessible and scalable for small and medium-scale farmers, maximizing the impact of precision agriculture.

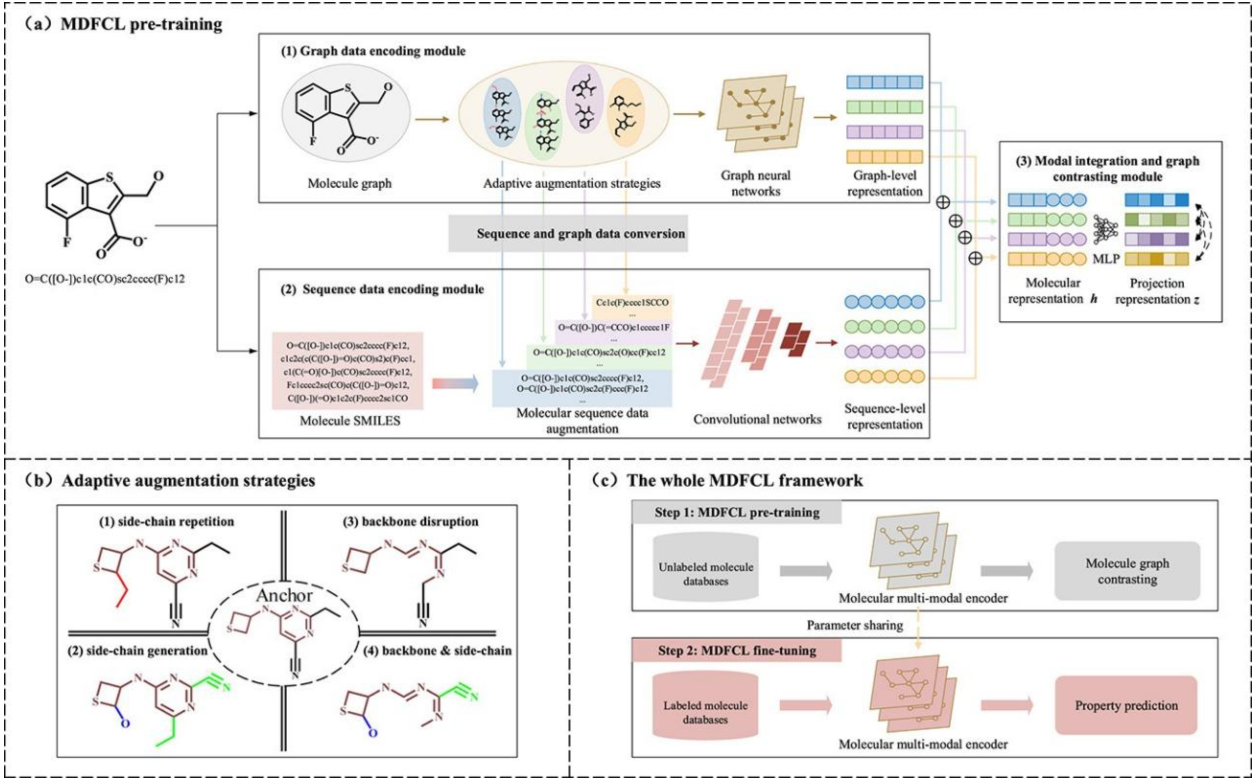
MDFCL – Multimodal Data Fusion via Contrastive Learning

MDFCL (Multimodal Data Fusion via Contrastive Learning) was originally designed to use a special technique (called contrastive learning) to match graph features from different inputs like sensors and images. But in our project, it was simplified into a classic fusion model that just combines features from both sources.

Our Adaptation

- our sensor data (like NPK, pH, etc.) goes through a small MLP (a simple two-layer neural network), and it outputs a 32-dimensional vector.
- our crop image is processed by ResNet-18, giving a 512-dimensional feature vector.
- These two vectors (32 + 512) are stuck together (concatenated).
- This combined vector is passed into another MLP classifier to predict the potassium level (Low, Normal, High).

MDFCL – Multimodal Data Fusion via Contrastive Learning



4. Strengths:

- Simple to understand and implement.
- Works well when both sensor and image data are clean.
- The architecture is transparent and easy to modify.

5. Limitations:

- Doesn't have any special mechanism for aligning or fusing information across modalities.
- Weak against missing modalities—especially image dropout.
- Has the most parameters and slowest performance out of the three models.

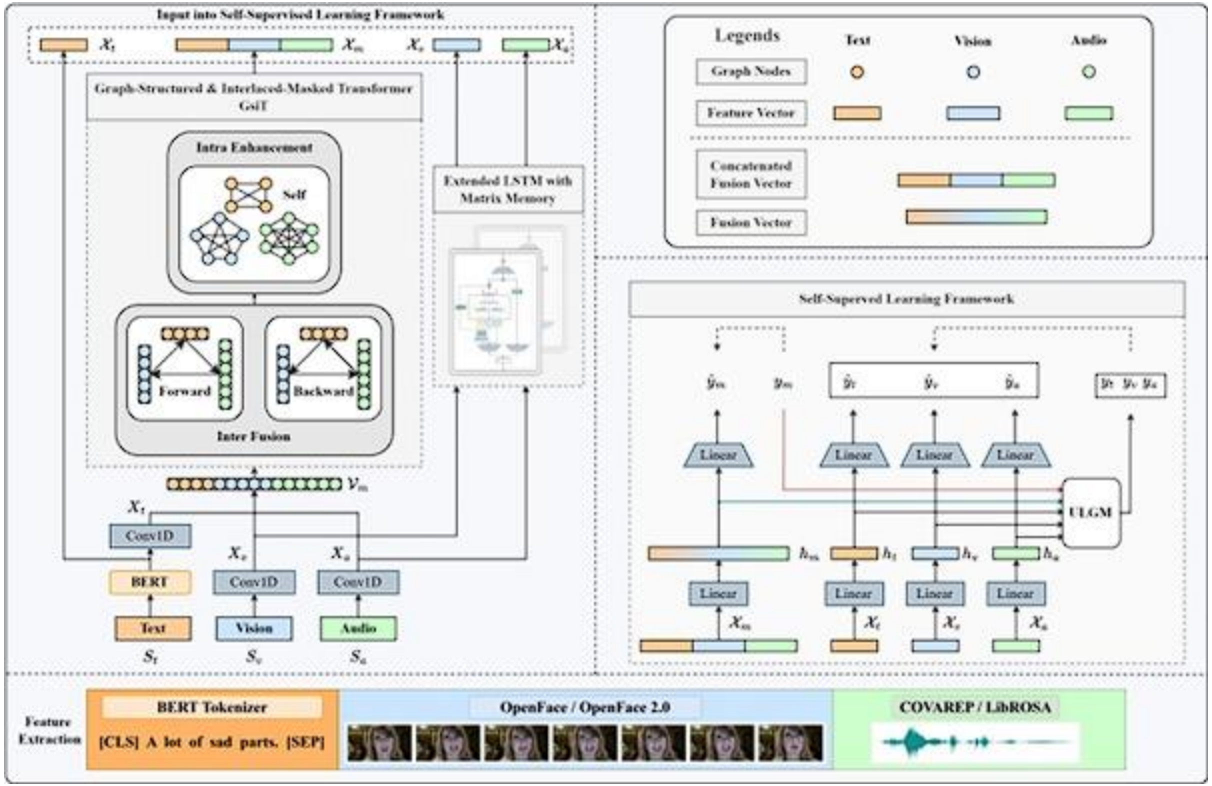
GSIFN – Graph-Structured Interlaced Fusion Network

GSIFN (Graph-Structured Interlaced Fusion Network) is a compact model that uses attention to let the image and sensor data "talk" to each other. It's designed to combine both data types while keeping things fast and lightweight.

Our Adaptation

- our sensor data (like NPK, pH) is passed through a small linear layer to get a 128-length vector.
- The crop image goes through ResNet-18 to extract features, then it's also turned into a 128-length vector.
- These two vectors (sensor + image) are stacked together—like two tokens in a sentence.
- A shared attention layer (MultiheadAttention) lets them "look at" each other and exchange useful information.
- The two tokens are averaged and passed through a small classifier to predict potassium level (Low / Medium / High).

GSIFN – Graph-Structured Interlaced Fusion Network



4. Strengths:

- Super lightweight and fast—great for devices with limited resources.
- Learns from both sensor and image inputs effectively.
- Still performs well even if one input is slightly damaged.

5. Limitations:

- Only uses one attention layer and just two tokens so it may not capture deep relationships.
- If one modality is completely missing, performance drops more than with Perceiver IO.
- Doesn't scale well if you want to add more input types in the future.

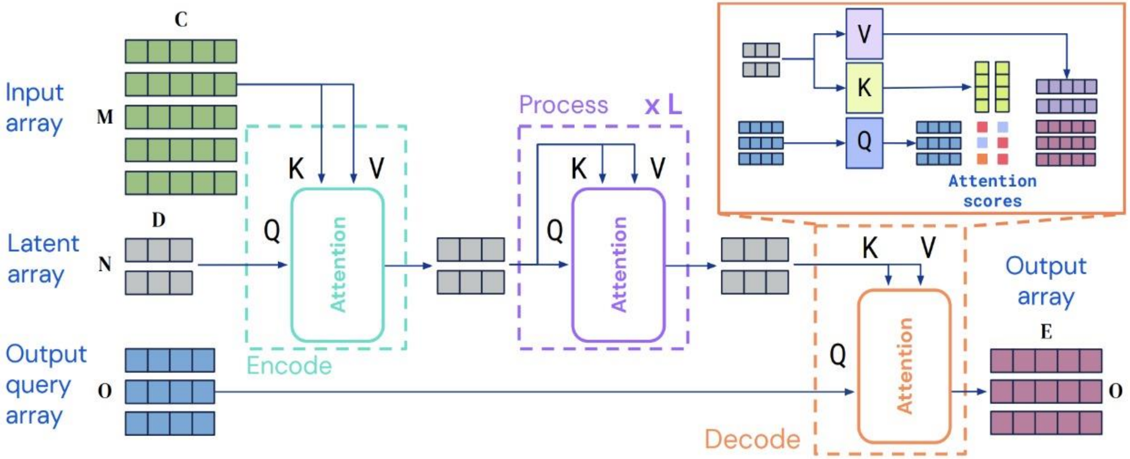
Perceiver IO – Latent Bottleneck Transformer

Perceiver IO is a type of neural network that learns from very different kinds of data at once—like numbers from sensors and pixels from images. It pulls all this info into a smaller “memory” space called latents, so it can focus on the most useful signals.Instead of making the Transformer handle all our raw data directly (which is slow), it only works on the small set of latent vectors. This makes it faster and more memory-efficient.

Our Adaptation

- our soil sensor data (N, P, K, pH, etc.) goes through a small MLP (a kind of mini neural network).
- our crop image goes through a pre-trained ResNet (used for image features).
- Each of these gives a “feature vector” summarizing the input.
- A set of learnable “latent vectors” (like attention sponges) interact with these inputs using a Transformer.
- The updated latents are averaged and passed to a classifier to predict the potassium category (Low/Normal/High).

Perceiver IO – Latent Bottleneck Transformer



4. Strengths:
   - Very good at combining different types of input.
   - Robust when one modality is missing or degraded.
   - Scales well—adding more inputs doesn't explode memory usage.

Limitations:
   - Slightly more complex than MLP-based models.
   - Needs careful tuning of how many latents to use (too few = bad learning, too many = slow).

Data Fusion for Comprehensive Analysis

1. **Diverse Fusion Strategies:**

   MDFCL, GSIFN, and Perceiver IO each implement distinct multimodal fusion techniques-ranging from graph-based contrastive learning to interlaced-masked attention and latent bottleneck transformers-to integrate sensor and image data.

2. **Standardized Benchmarking:**

   All models are evaluated on the same synchronized, preprocessed datasets to ensure a fair and rigorous comparison of their predictive performance in classifying crop nutrient levels.

3. **Robustness and Efficiency:**

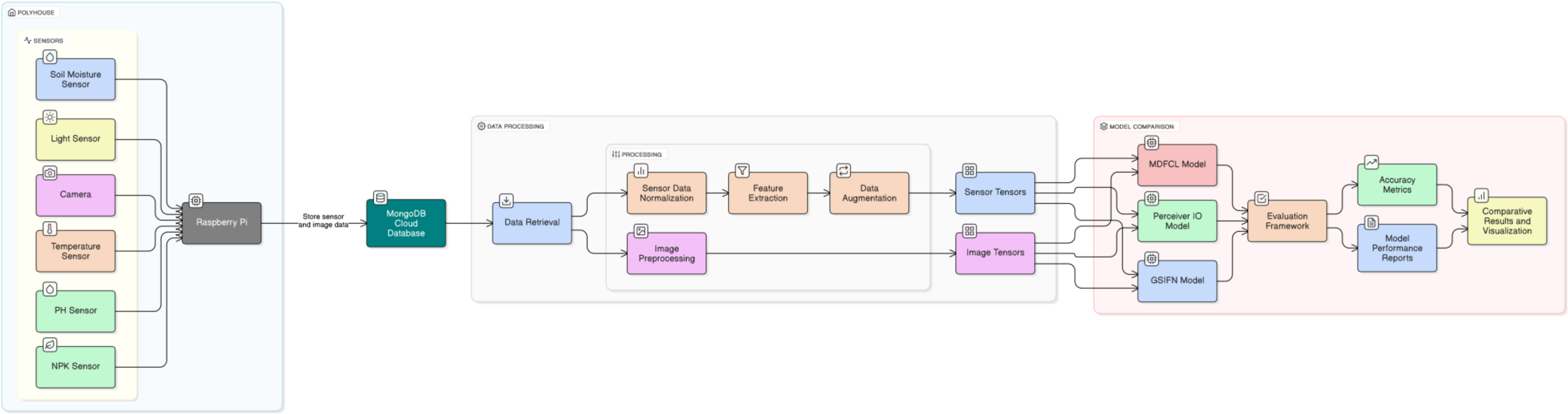
   The models are tested under both clean and degraded data conditions, assessing their resilience to sensor failures, image corruption, and missing data, as well as their computational efficiency on edge devices.

4. **Deployment-Oriented Insights:**

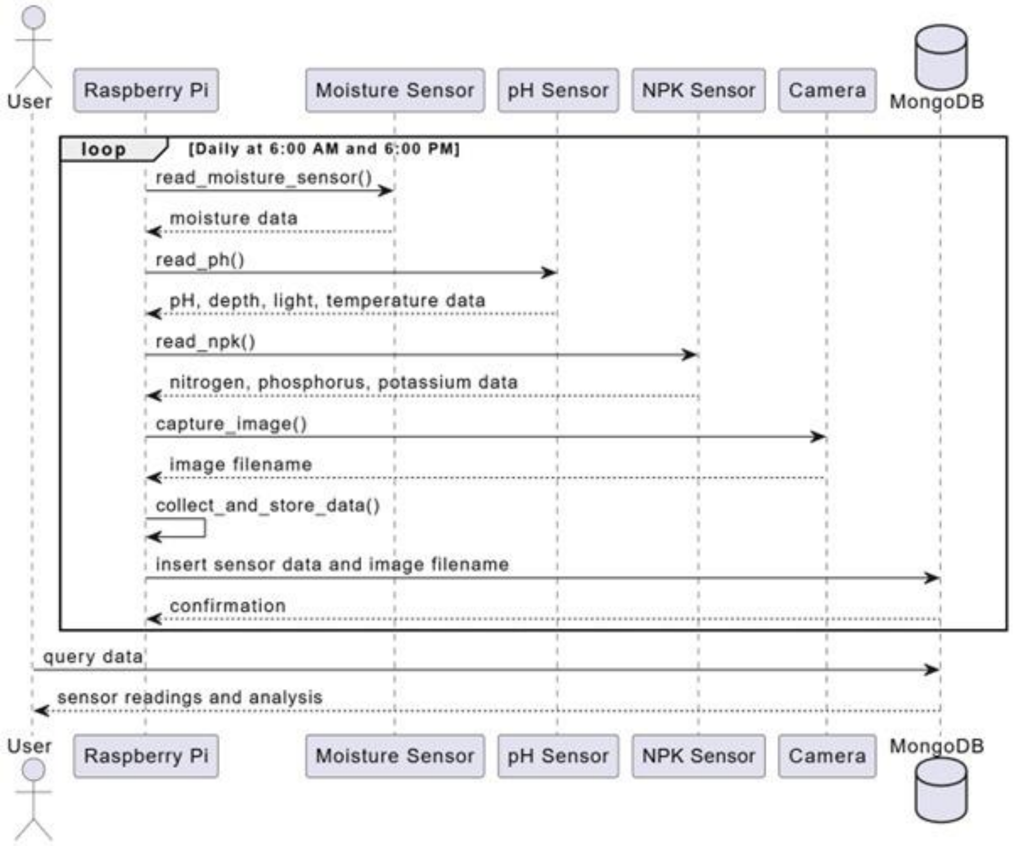
   The comparative analysis highlights which model is best suited for specific agricultural scenarios, balancing accuracy, robustness, and resource requirements to inform practical deployment decisions.

Model	Accuracy	Robustness	Complexity	Edge-ready	Fusion Method
MDFCL	97.7%	Poor (image dropout)	High (slowest)	❌	Feature concatenation
GSIFN	100%	Good	Low (fastest)	✅	Cross-attention (2-token)
Perceiver IO	97.7%	Excellent	Moderate	⚠️	Latent bottleneck attention

Architecture Diagram



Sequence diagram



OUTPUT

```
raspi4@raspi: ~/Pfiles

Microsoft Windows [Version 10.0.26100.3194]
(c) Microsoft Corporation. All rights reserved.

C:\Users\Aniruddha>ssh raspi4@raspi.local
raspi4@raspi.local's password:
Linux raspi 6.6.62+rpt-rpi-v8 #1 SMP PREEMPT Debian 1:6.6.62-1+rpt1 (2024-11-25) aarch64

The programs included with the Debian GNU/Linux system are free software;
the exact distribution terms for each program are described in the
individual files in /usr/share/doc/*/copyright.

Debian GNU/Linux comes with ABSOLUTELY NO WARRANTY, to the extent
permitted by applicable law.
Last login: Sat Mar  1 09:32:24 2025
raspi4@raspi:~ $ cd Pfiles/
raspi4@raspi:~/Pfiles $  source sensor/bin/activate
(sensor) raspi4@raspi:~/Pfiles $ python final.py
Successfully connected to MongoDB.
Moisture Sensor - Raw ADC Value: 0, Mapped Percentage: 0%
NPK Sensor - Nitrogen (N): 1 mg/kg, Phosphorus (P): 1 mg/kg, Potassium (K): 3 mg/kg
Image saved as images/image_20250301-095737.jpg
pH Sensor - pH: 37.76, Depth: 0.0, Light: 181.0, Temperature: 35.0
Sensor data written to MongoDB: {'timestamp': '2025-03-01 09:57:37', 'moisture': {'value': 0, 'percentage': 0}, 'ph_sensor': {'ph_value': 37.76, 'depth': 0.0, 'light': 181.0, 'temperature
': 35.0}, 'npk_sensor': {'nitrogen': 1, 'phosphorus': 1, 'potassium': 3}, 'image_filename': 'images/image_20250301-095737.jpg', '_id': ObjectId('67c28cb98d9bac3977411fc6')}
Moisture Sensor - Raw ADC Value: 0, Mapped Percentage: 0%
pH Sensor - pH: 37.76, Depth: 0.0, Light: 180.0, Temperature: 35.0
NPK Sensor - Nitrogen (N): 1 mg/kg, Phosphorus (P): 1 mg/kg, Potassium (K): 3 mg/kg
Image saved as images/image_20250301-095746.jpg
Sensor data written to MongoDB: {'timestamp': '2025-03-01 09:57:46', 'moisture': {'value': 0, 'percentage': 0}, 'ph_sensor': {'ph_value': 37.76, 'depth': 0.0, 'light': 180.0, 'temperature
': 35.0}, 'npk_sensor': {'nitrogen': 1, 'phosphorus': 1, 'potassium': 3}, 'image_filename': 'images/image_20250301-095746.jpg', '_id': ObjectId('67c28cc28d9bac3977411fc7')}
Moisture Sensor - Raw ADC Value: 0, Mapped Percentage: 0%
pH Sensor - pH: 37.76, Depth: 0.0, Light: 179.0, Temperature: 35.0
NPK Sensor - Nitrogen (N): 1 mg/kg, Phosphorus (P): 1 mg/kg, Potassium (K): 3 mg/kg
Image saved as images/image_20250301-095757.jpg
Sensor data written to MongoDB: {'timestamp': '2025-03-01 09:57:57', 'moisture': {'value': 0, 'percentage': 0}, 'ph_sensor': {'ph_value': 37.76, 'depth': 0.0, 'light': 179.0, 'temperature
': 35.0}, 'npk_sensor': {'nitrogen': 1, 'phosphorus': 1, 'potassium': 3}, 'image_filename': 'images/image_20250301-095757.jpg', '_id': ObjectId('67c28ccd8d9bac3977411fc8')}
Moisture Sensor - Raw ADC Value: 0, Mapped Percentage: 0%
pH Sensor - pH: 37.76, Depth: 0.0, Light: 185.0, Temperature: 35.0
NPK Sensor - Nitrogen (N): 1 mg/kg, Phosphorus (P): 1 mg/kg, Potassium (K): 3 mg/kg
Image saved as images/image_20250301-095806.jpg
Sensor data written to MongoDB: {'timestamp': '2025-03-01 09:58:06', 'moisture': {'value': 0, 'percentage': 0}, 'ph_sensor': {'ph_value': 37.76, 'depth': 0.0, 'light': 185.0, 'temperature
': 35.0}, 'npk_sensor': {'nitrogen': 1, 'phosphorus': 1, 'potassium': 3}, 'image_filename': 'images/image_20250301-095806.jpg', '_id': ObjectId('67c28cd68d9bac3977411fc9')}
^CData collection stopped by user.
/home/raspi4/Pfiles/final.py:202: RuntimeWarning: No channels have been set up yet - nothing to clean up!  Try cleaning up at the end of your program instead!
  GPIO.cleanup()
(sensor) raspi4@raspi:~/Pfiles $
```

SDG Alignment of Project

Goal 2: Zero Hunger

Objective Alignment: This research directly contributes to SDG Goal 2 by developing an IoT-enabled system to optimize the growth conditions of crops in polyhouses. By monitoring and adjusting environmental factors such as temperature, humidity, and soil moisture, the project ensures optimal crop growth, leading to improved agricultural productivity. This, in turn, helps in enhancing food security and availability by increasing yields and reducing wastage, particularly in controlled environments where space and resources are limited.

Impact: The system aims to provide sustainable and efficient farming practices that can increase food production while reducing resource wastage. By empowering small-scale farmers to use technology for better crop management, the project contributes to a more sustainable and resilient agricultural system, ultimately helping achieve global food security goals.

Goal 12: Responsible Consumption and Production

Objective Alignment: This project also aligns with SDG Goal 12 by promoting responsible resource usage in agriculture. Through IoT-enabled monitoring and data analytics, the project enables more efficient use of water, energy, and other resources in polyhouses. It helps minimize waste by ensuring that only necessary inputs (e.g., water, light) are used, contributing to more sustainable farming practices.

Impact: The research leads to reduced resource consumption and waste, aligning with the principles of responsible production and consumption. By optimizing agricultural inputs, the project supports sustainable production processes and contributes to environmental conservation, helping ensure that farming practices do not deplete resources and remain eco-friendly in the long run.

Conclusion

In conclusion, this research demonstrates that while all three multimodal deep learning models-GSIFN, MDFCL, and Perceiver IO-achieve strong baseline performance for soil potassium classification, each excels under different deployment scenarios. GSIFN offers rapid convergence and reliability, making it ideal for applications requiring frequent retraining or online adaptation. MDFCL stands out for its robustness in environments with sensor faults, ensuring stable predictions even with compromised data. Perceiver IO provides a practical balance between accuracy and computational efficiency, supporting scalable deployment where resources are limited. These findings underscore that model selection should be guided not only by accuracy but also by real-world constraints such as data quality, modality reliability, and available computing resources, ultimately enabling more informed and effective precision agriculture solutions.